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.
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]. SVN 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 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.
Results
To understand the patterns of somatic mutations in ATL, we characterized genome aberrations from 71 patients with different subtypes of ATL. We used the bioinformatics tool SciClone to quantify ITH from exome-sequencing data of tumor and matched normal samples.

Landscape of Genetic Alterations in ATL
We identified an average of 697 SNVs in coding regions from each sample using our SNV analysis pipeline (Figure S1). A comprehensive list of detected mutations with information including the type of mutation, VAF, and corresponding annotations for each sample is provided in Table S2. The relative abundance of the detected mutation categories including nonsynonymous, synonymous, stop gain, stop loss, frameshift deletion, and frameshift insertion in each sample and across all samples is presented in Figure 1. The total point mutation content in coding regions in genomic DNA from individual tumors ranged from 313 to 2365. The distribution of the somatic mutation content in coding regions in genomic DNA from individual samples from smoldering, chronic, acute, and lymphoma subtypes of ATL are shown in Figure 4. The visualized outputs across the remaining samples including SNVs both in neutral and altered copy number regions are presented in Figure S6.

Detecting Intratumor Genetic Heterogeneity by SciClone
Clustering of mutations revealed the subclonal structure of ATL in patients harboring multiple mutations. We used the detected SNV and CNA profiles for each sample to analyze the corresponding clonal/subclonal architecture using SciClone. To minimize the effect of CNAs and regions with loss of heterozygosity, we considered SNVs in neutral copy number regions. The results of four representative samples from smoldering, chronic, acute, and lymphoma subtypes of ATL are shown in Figure 4. The visualized outputs across the remaining samples including SNVs both in neutral and altered copy number regions are presented in Figure S6.

Subclonal Mutations Identified Within Each Sample
To provide further information on subclonal architecture based on mutational profiles, we used a list of recently identified highly with a neutral copy number (Figure 2B and Table S1). We used these regions to analyze the ITH in each sample. The distributions of the somatic mutations in these regions across samples and within each subtype are presented in Figures 2B and S2. Samples with aggressive ATL had more SNVs located in copy number altered regions ($P = .049$ (Figure S3)).

We also analyzed significantly altered regions of amplification and deletion across the samples (Figure 3). Information regarding cytoband, q-value, wide-peak boundaries, and genes in wide peaks is presented in Table S3.

Regions with significant changes to the copy number were detected in most tumors. Significant loss of 1p13.1, in which CD58 is located, was observed $(q = 0.000012641)$. Significant loss of 9p21.3, in which CDKN2A and CDKN2B are located, was observed $(q = 0.000030436)$. Significant loss of 10p14, in which GATA3 is located, was observed $(q = 0.0055013)$. Significant loss of 17p13.1, in which CD68 and TP53 are located, was observed $(q = 0.012124)$. Loss in the 6p21.33 region, the chromosomal location of $HLA-B$, was also observed $(q = 0.0386)$. Significant gain was observed in 9p24.1, the chromosomal location of $CD274$, which is commonly referred to as $PDLI$ $(q = 0.0000024788)$. Significant gain was observed in 6p25.3, the chromosomal location of $IRF4$ $(q = 0.034469)$. Significant gain was observed in 14q32.2, the chromosomal location of SETD3 $(q = 0.00046686$; Figure 3).

Figure 1. Somatic SNV profiles. (a) Distribution of different types of mutations within coding regions among samples. The proportion of nonsynonymous, synonymous, stop gain, stop loss, frameshift insertion, and frameshift deletion mutations is indicated. (b) Relative abundance of the detected mutation categories across samples. The color legend corresponds to both 1a and 1b.
Figure 2. Mutations in regions of copy number neutral, amplification and deletion among all samples. a) Somatic copy number analysis across samples. Numbers of detected amplification, neutral, and deletion events detected from each sample following copy number analysis are presented. b) Total detected somatic SNVs, SNVs located in regions with a neutral copy number, and SNVs located in regions with an altered copy number. The order of samples is identical to that shown in Figure 1A.

Figure 3. CNA in specific regions detected by GISTIC 2.0. Significantly observed CNAs across samples are presented. The genome is oriented vertically from top to bottom along the y axis. Numbers in the middle refer to the chromosome number. GISTIC q-values indicating the false discovery rate at each locus are presented on a log scale (x axis). Annotated peaks have residual q < 0.25. Red bars (right side), blue bars (left side) indicate amplifications and deletions, respectively. For each plot, known or interesting (i.e., with possible biological impact in ATL) candidate genes are written in black.
Mutated genes in ATL [40] and analyzed the clonal/subclonal status of these genes in each sample (Figures 4, S6 and Table S4). Subclones are numbered based on their descending order of VAF. The last subclone is the cluster with the lowest VAF.

To further explain the observed results and their interpretation, we present four main samples. ATL56 is derived from a patient with lymphoma type ATL with 132.7% proviral load (PVL), a measure of the number of provirus copies among 100 peripheral blood mononuclear cells. We detected 722 total coding SNVs, with 458 SNVs in neutral copy number regions and 264 SNVs in altered copy number regions. SciClone detected nine clusters in this case. Mutations in \textit{FAS}, \textit{NOTCH1}, and \textit{TP53} were located in a region with an altered copy number (i.e., one copy). \textit{DNMT3A} and \textit{PIK3CD} mutations were clustered together in subclone-7. Mutation of \textit{IRF2BP2} was clustered in subclone-6 (Figure 4).

ATL15 is derived from an acute patient with 84.9% PVL. We detected 850 total coding SNVs, with 813 SNVs in neutral copy number regions and 37 SNVs in regions with an altered copy number. SciClone detected seven clusters in this case. We also examined the overlap of mutated genes in this sample with the list of genes that are significantly mutated in ATL. The clonal status of the overlapping genes is indicated in Figure 4. Mutation of \textit{TP53} occurred in a region with an altered copy number (Figure S6). An \textit{IRF2BP2} mutation was clustered in subclone-7, which had the lowest VAF. Of note, \textit{CD58}, \textit{EP300}, \textit{NOTCH1}, \textit{POT1}, \textit{S1PR1}, and \textit{ZNF638} mutations were clustered together in subclone-6 (Figure 4).

ATL36 is derived from a chronic patient with 78.6% PVL. We detected 609 total coding SNVs, with 591 SNVs in neutral copy number regions and 18 SNVs in regions with an altered copy number. SciClone detected six clusters in this case. \textit{HLA-B}, \textit{NOXA1}, \textit{CD58}, \textit{EP300}, \textit{NOTCH1}, \textit{POT1}, \textit{S1PR1}, and \textit{ZNF638} mutations were clustered together in subclone-6 (Figure 4).
and RELA were the mutated genes that overlapped with the list of significantly mutated genes in ATL. Mutations in HLA-B were in three different locations. Nonsynonymous mutations of HLA-B with p.K292E in exon 4 and p.R45H in exon 2 were clustered together in subclone-6, which had the lowest VAF. HLA-B with p.A65T in exon 2 and the RELA mutation were clustered in subclone-4. NOXA1 showed a stop gain mutation and was clustered in subclone-5 (Figure 4).

Nag6 is derived from a smoldering patient with 34.88% PVL. We detected 441 total coding SNVs, with 439 SNVs in neutral copy number regions and 2 SNVs in regions with an altered copy number. SciClone detected five clusters in this case. Mutations in PIK3CD, FAS, and POT1 were clustered in subclone-3. The GATA3 mutation was clustered in subclone-4. Mutations in PRKCB, CSNK2A1, CCR4, and CARD11 were clustered in subclone-5, which had the lowest VAF (Figure 4).

We describe below the findings from five aggressive samples (lymphoma and acute ATL) that harbored mutations and clusters of interest. Additional samples are included in Figure S6.

Sas1 is derived from an acute patient with 84.6% PVL. We detected 1273 total coding SNVs, with 1163 SNVs in neutral copy number regions and 110 SNVs in regions with an altered copy number. SciClone detected three clusters in this case. S1PR1 (p.M80I), an IRF2BP2 mutation, and HLA-B (p.V272M) were clustered in subclone-3, which had the lowest VAF, whereas S1PR1:stop gain, PRKCB, TP53, CCR4, RHOA, HLA-B (p.G338R), HLA-B (p.Y140F), and NOTCH1 were clustered in subclone-2 (Figure S6).
ATL50 is derived from a patient with indolent ATL with 107.1% PVL. We detected 451 total coding SNVs, with 433 SNVs in neutral copy number regions and 18 SNVs in regions with an altered copy number. SciClone detected seven clusters in this case. Mutation of HLA-B (p.D98Y) was clustered in the main clone with the highest VAF, indicating that it is shared among almost all cells of the population and evolved during a very early stage of clonal evolution. Mutations in IRF2BP2, IF4, PLCG1, and HLA-B (p.E69G) were clustered in subclone-5. Mutations in HLA-B (p.A93T) and HLA-B (p.R45H) were clustered in subclone-6. Mutations in HLA-B (p.R243W) and HLA-B (p.D201E) were clustered in subclone-7, which had the lowest VAF (Figure S6).

ATL51 is derived from a patient with acute type ATL with 34.6% PVL. We detected 1214 total coding SNVs, with 59 SNVs in neutral copy number regions and 1155 SNVs (a particularly high number) in altered copy number regions. SciClone detected only one cluster in this case. A mutation in HLA-B was highly abundant and was detected in more than nine regions. Mutations in HLA-B, ATXN1, STAT3, and IKKββ were detected in regions with an altered copy number (three copies; Figure S6).

Kyo5 is derived from a patient with acute type ATL with 87.38% PVL. We detected 397 total coding SNVs, with 381 SNVs in neutral copy number regions and 16 SNVs in altered copy number regions. SciClone detected six clusters in this case. Multiple mutations in HLA-B (at five positions) were also detected in this patient. All HLA-B mutations were clustered together in subclone-6, which had the lowest VAF. Two mutations in CARD11 at p.S622Y and p.S622P were clustered in subclone-5. A mutation in GATA3 was also clustered in subclone-5. A mutation in TP53 was clustered in subclone-4. A mutation in CRLB was clustered in subclone-3. A mutation of PRKCB was clustered in subclone-2 (Figure S6).

ATL01 is derived from a patient with lymphoma type ATL with 25.8% PVL. We detected 1074 total coding SNVs, with 756 SNVs in neutral copy number regions and 318 SNVs in regions with an altered copy number. SciClone detected eight clusters in this case. Mutations in HLA-B (p.V9L), NOTCH1 (p.T349P), and TP53 were clustered in subclone-6. Mutations in ZFP36L2 (p.G471D), ZFP36L2 (p.R188H), and NOTCH1 (stop gain) were clustered in subclone-7. Mutations in HLA-B (stop gain), HLA-B (p.G338S), and HLA-B (p.Y140F) were clustered in subclone-8, which had the lowest VAF. (Figure S6).

The distribution of isolated mutations and detected clusters across all samples and among different subtypes is summarized in Figure 5A. The number of clusters ranged from one to nine, with an average of six clusters across the samples (the median estimated number of clusters was seven). Figure 5, B and C illustrate the distribution of clusters across all samples. Although the number of clonal/subclonal frequency modes tended to increase slightly with the number of mutations, the relationship was not linear (R^2 = 0.1718; Figure S4). Two or more clusters were observed to coexist in >94% of the samples. Therefore, we concluded that genetic ITH occurs in the majority of ATL samples in this study. To determine whether indolent and aggressive ATL differ in their clonality, we compared the distribution of clusters, in both a case-by-case and an overall distribution manner, and noted that aggressive ATL was likely to result in slightly more clonal clusters than did indolent types, although this difference was not significant (P = .41; Figure 5D). These distributions suggest that during the early stages of disease in indolent ATL, patients already display a wide variety of clonal clusters, which thus indicates the presence of mutational evolution. The distribution of clonal clusters among different subtypes of ATL is shown by boxplots and kernel density plots in Figure 5, D and E. Analysis of the mutation and clinical data among different subtypes of ATL showed that patients with the acute form had averages of 96% PVL, 4.6 clusters, and 734 coding mutations and were an average of 66 years old. These averages for lymphoma patients were 97% PVL, 6.2 clusters, 638 coding mutations, and 66 years old; for chronic patients, they were 80% PVL, 5.2 clusters, 644 coding mutations, and 65 years old; and for smoldering patients, they were 27% PVL, 4.5 clusters, 831 coding mutations, and 58 years old.

HLA-B was one of the highly mutated genes, with an average VAF of 25.96 (range, 7–100; median, 19). In most samples in this study, HLA-B harbored non-synonymous mutations in more than one region. HLA-B mutations were typically clustered in subclones with low VAF. CARD11 was one of the frequently mutated genes, with an average VAF of 25.65 (range, 10–74; median, 24). CARD11 mutations were typically observed in subclones with low VAF. ATL13, Kyo5, and ATL55 had multiple CARD11 mutations. In ATL13, the three mutations p.S433R, p.E626K, and p.S621F were clustered separately in subclones-5, -6, and -7, respectively. In Kyo5, mutations p.S622Y and p.S622P were clustered in subclone-5. In ATL55, both mutations were clustered in subclone-8, which had the lowest VAF. NOTCH1 was one of the frequently mutated genes, with an average VAF of 20.9 (range, 2–38; median, 21). This mutation was typically observed in subclones with low VAF. Sas1 and Sas2 had multiple NOTCH1 mutations. In Sas1, two mutations (p.T349P and p.G200 V) were clustered in subclone-2. In Sas2, three mutations (p.V1216 L, p.T349P, and p.A348P) were clustered in subclone-6, which had the lowest VAF. CCR4 was one of the frequently mutated genes, with an average VAF of 40.76 (range, 14–81; median, 44). This mutation was almost always observed in subclones with moderate VAF. Most of the samples contained only a single CCR4 mutation; however, ATL10 had two CCR4 mutations in p.Y331X (stop gain) and p.G33D (non-synonymous), both of which were clustered in the clone with the highest VAF. Nag6, ATL55, ATL33, and ATL13 each had a CCR4 mutation in the subclone with the lowest VAF. TP53 was one of the frequently mutated genes, with an average VAF of 42.4 (range, 14–86; median, 37). This mutation was typically observed in subclones with moderate VAF. Kyo9 was the only sample that had two mutations of TP53 at p.S246P and p.T245P, both of which were clustered in subclone-6, which had the lowest VAF. In other samples with a TP53 mutation, this mutation was found in subclones with moderate VAF. IRF4 was one of the frequently mutated genes, with an average VAF of 36.33 (range, 11–50; median, 42). This mutation was almost exclusively observed in subclones with moderate VAF. PLCG1 was one of the frequently mutated genes, with an average VAF of 32.94 (range, 10–72; median, 42). This mutation was almost always observed in subclones with moderate VAF. ATL27 had three mutations of PLCG1 at p.R48W, p.S866R, and p.V850A, which were found in subclone-1, subclone-2, and subclone-4, respectively (Figure S5).

Discussion

Clonal heterogeneity plays a central role in cancer, and quantification of ITH is an essential measure of tumor evolution [13]. Clonal heterogeneity complicates our understanding of tumor biology; in-depth characterization of ITH is needed for improvement of therapeutic responses and personalized medicine [52]. Genomic
instability and ITH have the potential to be useful prognostic and predictive measures in malignancies including ATL. Clonal genetic diversity in tumors negatively affects the clinical outcomes. Thus, recent studies have emphasized the significance of quantifying ITH and have detected and monitored the dynamics of subclonal events within tumors [20,53]. In this study, we tried to improve our understanding of tumor heterogeneity and analyze its correlation with different stages of disease in ATL to open new paths for its contribution to personalized medicine.

Combining ITH data with clinical information in a simple, quantitative, and practical manner is useful for clinical decision-making. However, it is difficult to overcome the technical hurdles in analyzing SNVs and CNAs and to use the appropriate bioinformatic pipeline for clustering the detected mutations. To overcome these difficulties, we performed an analysis of ITH in 71 ATL patients using 142 WES data for tumors and matched normal samples. Remarkably, the results showed a wide range of modes over clonal frequencies, with only one case showing a single cluster, and 94% of cases having more than two clusters. These findings indicate the presence of an extensive intratumor subclonal diversity in ATL.

During the past several years, the presence of considerable ITH in hematological malignancies and cancers, including cancers of the ovary, breast, lung, prostate, pancreas, bladder, and kidney, has been reported [21,54–57]. These studies have contributed substantially to a greater understanding of ITH and have suggested the potential application of ITH itself as a biomarker for prognosis [57,58]. More extensive studies are, however, needed to draw definitive conclusions about the role of ITH in the initiation, maintenance, and progression of cancers, as well as its clinical implications. Compared with other malignancies, ATL has been relatively neglected, and even the landscape of genomic events in ATL has been revealed only recently [41]. The current study is the first attempt to investigate the role of ITH in ATL. Clearly, more works needs to be conducted on ITH in ATL using various multiregional and/or longitudinal datasets.

As in other types of cancers, ITH is likely to be of potential biological and clinical significance in ATL. Considering the significant association between a high clone number and the poor outcome reported in several malignancies, the detection of a large number of clusters in these ATL samples is consistent with the fact that most ATL patients manifest very poor prognoses and therapeutic outcomes [31,33,59]. Overall, we observed a relatively large number of mutations in these subclones. We also noted the presence of a large number of small clones among these samples. Small clone sizes indicate their late evolution during tumor progression. The coexistence of multiple subclones indicates their similar fitness advantage, which prevents them from outgrowing each other. Therefore, we can hypothesize that mutational evolution is recently active and continuing in these ATL cases.

Recently, a high-throughput study of the mutation profiles of ATL patients proposed a list of mutated genes of particular relevance [41]; however, the clinical relevance of these mutated genes has not yet been clearly demonstrated. They also comprehensively analyzed viral gene expression (including HBZ and Tax). They reported that HBZ are generally detectable in all ATL cases, but Tax expression is rare [41]. Our data presented here are consistent with those from the previous publication by [41]. We also analyzed the clonal distribution of the top 50 mutated genes in ATL for each sample in this study. Alteration of cell cycle–related genes such as TP53 and CDKN2A has been suggested as a determining factor for acute transformation in ATL cases [40]. TP53 has been reported as a candidate gene that plays important roles in acute transformation [60]. Mutation of TP53 was considered a driving force for clonal expansion. In our analysis, we confirmed the presence of TP53 in subclones with relatively moderate or high VAF. CDKN2A has been suggested as a probable candidate tumor suppressor gene in ATL [40]. Loss of CDKN2A in acute ATL may lead to cell cycle deregulation [61]. We also confirmed the loss of CDKN2A in our cases. Moreover, genomic alteration of CD58, which plays a role in escape from immune-surveillance, was reported in acute ATL [40]. It has been reported that using immunosuppressive therapy for HTLV-1 carriers leads to early development of ATL [62–64]. Thus, the presence of genomic alterations related to immune escape, such as CD58 mutation, might be one of the significant determining factors that need to be considered in developing immunotherapeutic approaches for ATL. In addition, mutation of CCR4 is an important determinant of the clinical course of ATL, and a frameshift mutation in this gene has been linked to poor prognosis [65]. CCR4 is expressed on the surface of 90% of ATL cells [66], and a monoclonal antibody against CCR4 (mogamulizumab) demonstrated very strong cytotoxicity against ATL cells [67,68]. CCR4 mutations in our study were clustered in subclones with medium and low VAFs. The effect of harboring a CCR4 mutation and its clonal status on the response to mogamulizumab treatment is a topic that should be investigated in future studies.

Longitudinal monitoring of clonality among infected cells using HTLV-1 integration sites indicates that clonal expansion of infected cells occurs from the early stages of infection, even when individuals are still healthy and are diagnosed as asymptomatic carriers. Individuals with a largely expanded infected clone have a high risk of ATL development [37,38]. However, it is not clear whether the expanded cell populations are homogeneous or heterogeneous at a mutation level. In this study, analyzing smoldering, chronic, acute, and lymphoma subtypes of ATL, suggested that expanded cells are not homogeneous. A high degree of ITH was evident in the analyzed cases of this study, even among patients with smoldering and chronic subtypes, which represent early stages of the disease. Therefore, we suggest that, like other malignancies, the presence of heterogeneity among clones (ITH) is likely to be an essential factor for ATL development. Also, ITH is likely to occur during early stages of ATL development. Indeed, the presence of more than one clone in many tumor types raises important—and thus far unanswered—questions regarding the biological mechanisms behind ITH. Why is heterogeneity present among various tumor types, and why does the fittest clone not take over within a tumor? In pursuit of revealing the functional relevance of ITH, a highly interesting concept of clonal cooperation has been suggested. In the case of breast cancer, an analysis of the tumorigenicity of different clones using in vivo models showed the coexistence of two types of clones that are essential for tumorigenicity [69,70]. It has been suggested that different clones, including those that are most fit, have a paracrine interdependence with other subclones, and hence clonal cooperation is essential for maintaining tumorigenicity [69,70]. Further in-depth functional analyses of ITH in various malignancies are required to clarify the significant role of ITH in cancer development.

**Conclusion**

Our findings suggest that clonal diversity might provide additional prognostic information for ATL. Measurement of ITH might thus
eventually become an aid to clinical decision-making with respect to treatment of patients with ATL. Although these observations require confirmation by longitudinal analyses of ATL cases pre- and post-chemotherapy as well as molecular target therapy, clonal heterogeneity might hold promise for a deeper understanding of ATL. As we move toward an era of personalized medicine for ATL and other cancers, we need to build an understanding of the clinical impact of subclone admixtures containing specific mutations. In the case of ATL, the presence of multiple subclones may increase the clonal competition, and thus the probability of the presence of resistant clones among a complex population of subclones is very high.

Declarations

Ethics Approval

Access to the raw sequencing data was approved based on the agreement between Kyoto University and the University of Tokyo.

Availability of Data and Material

The raw sequencing data have been deposited in European Genome-phenome Archive (EGA) under accession number of EGAD00001001410.

Disclosure

The authors have declared no conflicts of interest.

Funding

This work was supported by a “Kakenhi” grant from the Japan Society for the Promotion of Science [grant number 17 K15044]; and The Tokyo Biochemical Research Foundation [TBRF-RF 17–104]; and Japan Agency for Medical Research and Development [k0106526h0001].

Authors’ Contributions

AF and SF contributed to the conception and design of the study, development of the methodology, analysis and interpretation of the data (e.g., statistical analysis and computational analysis), and the writing of the first draft of the manuscript, as well as designing the figures. RK and WM contributed in bioinformatics and methodological writing of the first draft of the manuscript, as well as designing the figures. RK and WM supported with critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to professor Seishi Ogawa from Kyoto University for providing raw sequencing data. AF expresses deep respect and gratitude to the Otsuka Toshimi scholarship foundation for supporting his graduate studies and to U. Firouzi for help in the design of figures.

Computational analyses were provided by the Supercomputer System at Human Genome Center, The Institute of Medical Science, The University of Tokyo. We appreciate the technical assistance of H. Nishijima.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neo.2018.07.001.

References

[1] Greaves M (2015). Evolutionary determinants of cancer. Cancer Discov 5(8), 806–820.

[2] Sproulliske K, Merlo LM, Gerrish PJ, Maley CC, and Sniegoski PW (2012). Cancer in light of experimental evolution. Curr Biol 22(17), R672–771.

[3] Merlo LM, Pepper JW, Reid BJ, and Maley CC (2006). Cancer as an evolutionary and ecological process. Nat Rev Cancer 6(12), 924–935.

[4] Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, and Ramakrishna M, et al (2012). The life history of 21 breast cancers. Cell 149(5), 994–1007.

[5] Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Varathath S, and Liang H, et al (2014). Clonal evolution in breast cancer revealed by single nucleotide genome sequencing. Nature 512(7513), 155–160.

[6] Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radiyovitevich T, Suzuki H, Przychodzen B, Nagata Y, Meggendorfer M, and Sanada M, et al (2017). Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet 49(2), 204–212.

[7] Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, and Haffari G, et al (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486(7403), 395–399.

[8] Roth A, McPherson A, Laks E, Bieke J, Yap D, Wan A, Smith MA, Nielsen CB, McAlpine JN, and Aparicio S, et al (2016). Clonal genotype and population structure inference from single-cell tumor sequencing. Nat Methods 13(7), 573–576.

[9] Andor N, Graham TA, Jansen M, Xia LC, Akopis CA, Petrisch C, Ji HP, and Maley CC (2016). Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med 22(1), 105–113.

[10] Hardiman KM, Ulitz RJ, Kucik RD, Hovelson DH, Gates CM, Bhasi A, Rodrigues Grant A, Liu J, Cani AK, and Greenkorn J, et al (2016). Intra-tumor genetic heterogeneity in rectal cancer. Lab Invest 96(4), 4–15.

[11] Fischer A, Vazquez-Garcia I, Illingworth CJR, and Maley CC, et al (2014). High-definition reconstruction of clonal composition in cancer. Cell 175(5), 1740–1752.

[12] McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, and Swanton C (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med 7(283), 1–11. doi:10.1126/scitranslmed.aaf1400 283r324.

[13] Randolph C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xia M, Zhang Q, McMichael JF, and Wyczalkowski MA, et al (2013). Mutational landscape and significance across 12 major cancer types. Nature 502(7471), 333–339.

[14] Hong MK, Macintyre G, Wedge DC, Van Loo P, Patel K, Lunke S, Alexandre LB, Sloggett C, Cmero M, and Marazz F, et al (2015). Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nat Commun 6, 1–12. doi:10.1038/ncomms7605 6605.

[15] Bolli N, Aver-Liisaj H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, Dawson KJ, Iorio F, Nik-Zainal S, and Bignell GR, et al (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun 5, 1–3. doi:10.1038/ncomms997 2997.

[16] Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA, Paulson TG, Blount PL, Risques RA, and Rabinovitch PS, et al (2006). Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nat Genet 38(4), 468–473.

[17] Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Siva Chenko A, and Wang L, et al (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152(4), 714–726.

[18] Kim H, Zheng S, Amiri SS, Vilk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, and Hu J, et al (2015). Whole-genome and multisite exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res 25(3), 316–327.
mutation identifies a distinct group of adult T cell leukaemia/lymphoma with poor prognosis. *J Pathol* 238(5), 621–626.

[66] Yoshie O, Fujisawa R, Nakayama T, Harasawa H, Tago H, Izawa D, Hieshima K, Tatsumi Y, Matsushima K, and Hasegawa H, et al (2002). Frequent expression of CCR4 in adult T-cell leukemia and human T-cell leukemia virus type 1-transformed T cells. *Blood* 99(5), 1505–1511.

[67] Ishida T, Joh T, Uike N, Yamamoto K, Utsunomiya A, Yoshida S, Saburi Y, Miyamoto T, Takekuto S, and Suzuki H, et al (2012). Defucosylated Anti-CCR4 Monoclonal Antibody (KW-0761) for Relapsed Adult T-Cell Leukemia-Lymphoma: A Multicenter Phase II Study. *J Clin Oncol* 30(8), 837–842.

[68] Yonekura K, Kanzaki T, Gunshin K, Kawakami N, Takatsuka Y, Nakano N, Tokunaga M, Kubota A, Takeuchi S, and Kanekura T, et al (2014). Effect of anti-CCR4 monoclonal antibody (mogamulizumab) on adult T-cell leukemia-lymphoma: Cutaneous adverse reactions may predict the prognosis. *J Dermatol* 41(3), 239–244.

[69] Cleary AS, Leonard TL, Gestl SA, and Gunther EJ (2014). Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* 508(7494), 113–117.

[70] Polyak K and Marusyk A (2014). Cancer: Clonal cooperation. *Nature* 508(7494), 52–53.