Monitoring chronic lymphocytic leukaemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns

**Short Title:** longitudinal whole genome sequencing in CLL

Anna Schuh¹, Jennifer Becq², Sean Humphray², Adrian Alexa², Adam Burns¹, Ruth Clifford¹, Stephan M. Feller³, Russell Grocock², Shirley Henderson¹, Irina Khrebtukova⁴, Zoya Kingsbury², Shujun Luo⁴, David McBride², Lisa Murray², Toshi Menju³,⁵, Adele Timbs¹, Mark Ross², Jenny Taylor¹ and David Bentley²

¹Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford UK

²Illumina Cambridge Ltd., Chesterford Research Park, Saffron Walden, UK

³Biological Systems Architecture Group, Department of Oncology, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

⁴Illumina Inc., 25861 Industrial Boulevard, Hayward CA 94545, USA

⁵Department of Thoracic Surgery, Kyoto University, Graduate School of Medicine, Kyoto 606-8507, Japan

**corresponding author:** Anna Schuh, Oxford NIHR Biomedical Research Centre, Molecular Diagnostic Laboratory, level 4 John Radcliffe Site, University of Oxford, anna.schuh@ndcls.ox.ac.uk; Oxford UK; +44-1865-572766
Abstract

Chronic lymphocytic leukaemia (CLL) is characterized by relapse after treatment and chemotherapy resistance. Like in other malignancies, leukaemia cells accumulate mutations during growth, forming heterogeneous cell populations that are subject to Darwinian selection and may respond differentially to treatment. There is therefore a clinical need to monitor changes in the subclonal composition of cancers during disease progression.

Here, we use whole-genome sequencing (WGS) to track subclonal heterogeneity in three CLL patients subjected to repeated cycles of therapy. We reveal different somatic mutation profiles in each patient and use these to establish likely hierarchical patterns of subclonal evolution, to identify subclones that decline or expand over time, and to detect founder mutations. We show that clonal evolution patterns are heterogeneous in individual patients. We conclude that genome sequencing is a powerful and sensitive approach to monitor disease progression repeatedly at the molecular level. If applied to future clinical trials- this approach might eventually influence treatment strategies as a tool to individualise and direct cancer treatment.

Introduction

Despite significant progress in the management of lymphomas and leukaemias, relapse remains the major cause of death. Increased use of expensive targeted therapies and toxic chemotherapies (especially in the elderly) confronts us with an urgent need to improve response prediction for all cancer patients in order to reduce side-effects and costs from ineffective treatment. Current diagnostic approaches to treatment selection, response monitoring and relapse prediction are limited to single genes and apply only to a minority of haematological cancers. This is at odds with modern concepts of tumour propagation and maintenance, which propose that every cell in an individual cancer is characterised by a combination of mutation events that comprise tumorigenic (driver) mutations, passive (passenger) mutations, and possibly predisposing germline risk variants. Cancer cells propagate and diversify during tumour growth, resulting in a heterogeneous population of genotypically and phenotypically distinct subclones that are related in a hierarchical lineage. As the composition of the local environment changes, for example as a consequence of drug treatment, tumour cell populations adapt and evolve by Darwinian selection1-3.
Whole genome sequencing (WGS) of a single tumour sample can be used to generate a comprehensive catalogue of variants that provides a snapshot of the cell population en masse at a particular time-point. However, over time and with continued evolution of the cancer, this snapshot becomes progressively less representative of the disease. Recent reports have described whole tumour genomes from single patients or cohorts of individuals mostly at single time-points and irrespective of treatment. This approach has enabled identification of mutations representative and in some cases highly predictive of histological cancer type, outcome and/or treatment response. Comparison of sequence data from primary and metastatic tumour samples, or from multiple locations within a tumour, reveals major differences in the somatic mutation profiles within an individual, illustrating the dynamic nature of tumour evolution. Recently, two time-point analysis of relapsed and secondary acute myeloid leukaemia have also demonstrated clonal evolution at a molecular level.

We elected to study subclonal evolution in B-cell chronic lymphocytic leukaemia (CLL). CLL is characterised by immunodeficiency, autoimmunity, a chronically relapsing course and the development of chemotherapy resistance, making it an ideal model to study tumour progression. Using WGS analysis, we tracked molecular changes in pre-treatment, post-treatment and relapse samples in three patients. We defined cellular subpopulations on the basis of somatic mutation profiles and revealed changes within the tumour clonal architecture over time as patients were subjected to multiple rounds of treatment. We describe for the first time the heterogeneous patterns of clonal evolution in patients with IgHV unmutated CLL throughout the lifetime of their disease. This proof-of-principle study enabled us to evaluate how large-scale sequence information might be used in future clinical trials to evaluate response and to target therapies more effectively for patients suffering from CLL and perhaps other cancers.
Material and Methods

Samples. Informed consent from CLL patients was obtained in line with the Declaration of Helsinki and in line with the Oxford IRB ethics 09/H0606/5. DNA and RNA were extracted from peripheral blood CLL lymphocytes and control DNA from buccal smear samples.

WGS. Application of Sequencing by Synthesis (SBS) to human WGS has been described previously21. Cluster formation and sequencing to an average depth of 40x using were carried out the Illumina TruSeq v3 Cluster and SBS kits, respectively. 100 bp paired-end sequence reads were generated using HiSeq 2000.

Somatic mutation calling and analysis. The Illumina CASAVA v1.8 pipeline was used for quality control, alignment to human GRCh37.1 reference and variant calling for single genome analysis. Identification of somatic mutations was performed with prediction software that uses a joint Bayesian model combining analysis of the tumour and normal genomes. Annotation of mutation consequence was performed using Ensembl Variant Effect Predictor on Ensembl database release e6222. Mutant allele frequencies (AF) and consecutive AF differences were calculated at each tumour stage for base substitutions. Mutations were clustered into groups showing similar mutation AF profiles using a k-means algorithm.

Targeted Deep Amplicon Sequencing. Selected somatic substitution sites in protein-coding genes were amplified from genomic DNA by two-step PCR. Partial adapter sequences were added as 5’ extensions to target-specific primers and utilised as priming sites in the second PCR step to complete the adapter sequences required for cluster generation and SBS. Amplicons were sequenced using the Genome Analyzer IIx to an average depth of 100000x.

See Supplementary Methods for a detailed Methods section and for a list of all primer sequences.
Results

Somatic mutation detection at five time-points in each patient’s disease progression

We selected three patients with CLL who received multiple different treatments sequentially over a period of up to seven years (Supplementary Table 1 and Supplementary Information). We took peripheral blood samples at five specific time-points during disease progression together with one matched buccal swab per patient (Supplementary Table 2), and performed WGS and mutation analysis (Supplementary Tables 3-5). Genome-wide somatic mutations were in the range of 1,744-2,829 substitutions and 204-385 insertions/deletions per sample (Supplementary Tables 4 and 5). There was a clear bias towards C>T / G>A substitutions (Supplementary Fig. 1) as seen previously in other cancers\(^2\). C>T substitutions have previously been linked to UV light exposure in skin cancers\(^23\) and recently to specific sequence signatures (eg: TpCpX) in breast cancer\(^24\).

Between 14 and 22 mutations per sample are predicted to alter protein-coding sequences (Supplementary Tables 4-8). WGS analysis confirmed copy number aberrations (CNAs) seen in a previous array-based analysis and revealed additional CNAs (Supplementary Fig. 2 and Supplementary Table 9). CLL003 had three large CNAs: del11q23.2 and del13q14.1 remained unchanged over time and were detected at all time-points; loss/gain of 8p/8q was first seen at first relapse (time-point b) in a small subclone that subsequently expanded. CLL077 developed a deletion of chromosome 6q first observed before Ofatumumab treatment (time-point d). CLL006 had trisomy 12 at all time-points (data not shown).

Mutation frequency profiles differ between patients and change over time

We determined allele frequencies of all somatic single nucleotide variants (SNVs) at each disease stage, established a profile for each mutation during disease progression and grouped similar mutation profiles together. This revealed changes in mutation profiles over time and clear differences between patients (Fig. 1a-c). The most dynamic profiles were seen in CLL003. Mutation profiles of CLL077 were relatively stable initially and then underwent a change at later stages. Finally, CLL006 mutation profiles remained relatively stable throughout.

To extend the sensitivity of the study we selected specific somatic mutations from each profile, focusing on those predicted to alter protein structure. We performed targeted deep sequencing to an average depth of 100000x to quantify the mutation
frequency at each stage to high accuracy and to observe low levels of somatic mutations (down to ~0.5%) previously undetected by WGS (Fig. 1d, Supplementary Fig. 3 and Supplementary Tables 5-7). All mutations selected were confirmed by deep sequencing (see Methods). The quantitative analysis revealed a striking similarity in frequency profiles for different mutations in the same group (Fig. 1d and Supplementary Fig. 3). Considering the deep sequence and WGS data we defined five mutation profiles: (1) high (H) frequency at initial diagnosis and later (HH); (2) high at diagnosis then low (L) or disappearing after treatment (HL); (3) initially at low frequency but then increasing (LH); (4) undetectable by deep sequencing at diagnosis (0H); (5) present at low frequency throughout (LL). All five profiles are evident in CLL003, but only HH, 0H and LL profiles are present in CLL077, while CLL006 is characterised exclusively by HH and LL profiles (Fig. 1, Supplementary Fig. 3).

**Defining leukaemia architecture and founder mutations**

We used the deep sequencing data to define tumour subclones and to infer an evolving and branching cellular hierarchy of tumour cells for the three patients (Fig. 2). This analysis enabled us to define a founder subclone in each patient that was genetically characterized by mutations present in all tumour cells at all time points (mutation profile HH in Fig. 1d, Fig. 2 and Supplementary Fig.3). Mutations of this type should include the initial drivers of tumourigenesis, as well as passenger mutations that were fixed in the originating tumour cell. Additional subclone diversity was due to other mutations (HL, LH, 0H, HH profiles in Fig. 1d, Fig. 2 and Supplementary Fig. 3) that arose on the background of the founder mutations.

The list of mutated genes was unique to each patient (Supplementary Tables 6-8). However, each carried one or more candidate driver mutations based on recurrence in CLL or other cancers (Supplementary Table 10). Importantly, a single somatic founder mutation in each patient affected a gene recurrently mutated in CLL (CLL003: **SF3B1**; CLL077: **SAMHD1**; CLL006: **MED12**). A further 5-10 non-recurrent mutations were fixed within the founder clone and could include both driver and passenger events. By contrast, **ATM, PLEKHG5** and **IRF4** mutations, although recurrent in CLL, were clearly secondary events, as they were not observed in all tumour cells and because their allele frequency reduced during treatment.
Patterns of clonal evolution are heterogenous in CLL

Next, we explored how the patterns of subclonal evolution differed between the three patients. Using conventional prognostic markers, all three patients belonged to an intermediate risk group (IgVH unmutated, no TP53 abnormalities, no genomic complexity $^{32-36}$ (Supplementary Table 1) and were treated with similar combinations of alkylating agents, purine analogues and immunotherapy.

CLL003 showed dramatic shifts in subclonal composition over time. At diagnosis, 82% of cells from CLL003 carried a nonsense mutation in $ATM$ and had lost the other copy of the gene as the result of an 11q deletion (“subclone 2” in Fig. 2a-c). Subclone 2 expanded at first relapse and accounted for more than 90% of the tumour cells. The patient subsequently achieved a minimal residual disease (MRD) positive complete remission following Fludarabine/Cyclophosphamide/Rituximab (FCR) (time point c) (Supplementary Fig. 4). This coincided with a dramatic contraction of subclone 2. However, another subclone defined by mutations in genes known to be mutated in malignancies ($FAT3$, $NPY$, $NRG3$, $ASXL1$, $MUSK$, $SEMA3E$)$^{31}$ was detected in a large fraction of this remission sample and became dominant at later stages (“subclone 4” in Fig. 2a-c). $ASXL1$, $MUSK$ and $SEMA3E$ mutations were detected at low level by WGS in a sample two years previously and before the patient ever received treatment.

By contrast to patient CLL003, the major subclones of CLL077 remained initially largely unchanged (timepoints a-c), consistent with the clinical picture of refractory but stable disease. As with CLL003, clinical disease progression in CLL077 coincided with expansion of a subclone (“subclone 4” in Fig. 2d-e) containing a mutation in a cancer gene ($MAP2K1c.171G>T$, Supplementary Figure 5). $MAP2K1$ mutations are rare events in lung cancer$^{37}$, and $MAP2K1$ (Mek-1) inhibitors have entered Phase III clinical trials for solid tumours. However, we did not find any $MAP2K1$ mutations by targeted sequencing of 90 patients as a follow-up in this study (data not shown). The $MAP2K1c.171G>T$ mutation causes increased phosphorylation of ERK1/2 in transfection experiments using heterologous cells$^{37}$. We showed that phosphorylated ERK1/2 increased over time in lymphocytes of patient CLL077 and that this mirrored the expansion of the subclone containing the mutation (Supplementary Figure 6). Although this does not prove the pathogenic mechanism, it implies that the mutated $MAP2K1$ gene product was active in this patient at these time-points. At relapse before death subclone 4 was pre-dominant in
CLL077 (time-point e). Importantly, the presence of the MAP2K1 change was detected at low levels by WGS in the pre-Ofatumumab sample taken 9 months earlier (time-point d).

Clonal evolution in CLL006 was characterized by the absence of expanding or emerging subclones. With every treatment, absolute lymphocyte cell numbers decreased dramatically. The same subclones re-emerged at relapse albeit in different proportions. At later stages (timepoints d-e), subclones containing IRF4 mutations (4 and 5) outcompeted all other cells.

The findings for patient CLL006 demonstrate that relapse does not always coincide with expansion of subclones containing new or rare mutations. Instead, it could be due to cell-extrinsic factors, such as the pharmacokinetic properties of monoclonal antibodies resulting in incomplete penetration of lymphoid organs and subsequent redistribution of residual leukaemic cells into the periphery.

Discussion

Cancers initiate from a single cell with one or more founder mutations and acquire additional mutations, some of which may give rise to resistance or confer sensitivity to treatment. We demonstrate that WGS provides an abundance of mutated sites whose allele frequency profiles can be grouped to reveal the likely evolving subclonal hierarchy of leukaemia. Additionally, WGS provides information on somatic mutations in non-coding regions, whose significance for cancer currently remains to be determined.

In all three patients, we identified founder events that could be future targets for curative therapy in CLL. Further, in all patients WGS defined the genetic composition of subclones that later became dominant even before initiation of relapse treatment. In some cases these changes to the molecular phenotype of the tumour became apparent months or years ahead of an obvious clinical phenotype, thus offering the possibility for earlier targeted and sequential treatment selection directed against these subclones, which are characterized by the presence of non-recurrent or low-recurrence mutations. In practice however, this approach might only shift the balance of the different subclones but not affect the ultimate outcome for the patient.

We show that genome-wide tracking of somatic mutation profiles over time reveals heterogeneous patterns of clonal evolution in CLL. All three patients had multiple
subclones before treatment. With the exception of subclone 2 in CLL003, which became undetectable and never recurred, these subclones persisted through later disease stages. This is different from ALL\(^1\) and AML models of relapse where either the dominant (model 1) or a single minor subclone (model 2) give rise to relapse\(^{19}\), and is more similar to the pattern observed in MDS/AML progression\(^{20}\). Relapse in two of the CLL patients is also characterized by emergence and expansion of subclones that were not present at diagnosis.

We identify both dynamic/rapid and stable/gradual shifts in the interclonal balance. The clinical and prognostic significance of these subclonal shifts remains to be established. For example, we have seen that a relatively stable molecular phenotype in patient CLL006 correlates with slow disease progression and a good response to repeated treatment, whereas emerging or increasing subclones in the other two patients correlate with resistance to different treatments and death. It could be that in patients CLL003 and CLL077 the chemotherapy itself selected for chemo-resistant subclones and/or induced new mutations causing resistance through DNA damage. These conferred a survival advantage and led to expansion of resistant subclones. By contrast, this did not occur in CLL006 who was treated almost exclusively with antibodies. Going forward, the potential clinical utility of our analysis approach will need to be evaluated systematically within clinical trials in larger cohorts of patients. Depending on the outcome of these studies, longitudinal WGS studies may eventually provide a means to individualise treatment\(^{38}\).

Challenges remain to genome-wide sequencing being applied within clinical trials. These include simplifying and standardising the currently complex analysis methods, and improving turnaround time, costs and interpretation of clinically actionable information. Clinical implementation also depends on the availability of sequential biopsies, access to integrated phenotype-genotype databases and effective therapeutics.

Given progress in all these areas, we anticipate that genome-wide sequencing will become an effective approach to monitor disease progression systematically and also prospectively, and that it will direct future clinical trials and therapeutic decisions. Its successful implementation could fundamentally change our strategy for treatment selection and monitoring and provide the tool for delivering more successful and cost-effective healthcare with better outcomes for individual patients.
Acknowledgements

We thank Helen Northen for performing experimental work.

"This work was supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health."

Authorship Contributions and Disclosure of Conflicts of Interest

AS, SMF, MR, and DB designed, analysed and interpreted experiments. JB performed bioinformatics. SH, AA, AB, RC, RG, SH, IK, ZK, SL, DMB, LM, TM, AT performed experiments. AS wrote the first draft of the manuscript. AS, JT, MR, DB wrote the final draft.

All authors at Illumina (see affiliations) are employees of Illumina Inc., a public company that develops and markets systems for genetic analysis.
References


Figure Legends

**Figure 1 Genome-wide clustering reveals changes in mutation profiles**

a-c: Grouping of somatic mutation profiles for all single nucleotide variants (SNVs). Absolute white blood cell (WBC) and lymphocyte (LY) counts are shown at the top of each figure. The lower panels show genome-wide SNV frequencies plotted against the five time points. Mutation profiles for coding genes are shown as black lines.

a, CLL003 a= before chlorambucil; b= before fludarabine, cyclophosphamide, rituximab; c= immediately after 6 cycles of fludarabine, cyclophosphamide, rituximab; d= before ofatumumab; e= after ofatumumab.

Coding mutations: red plot SLC9A11, NLRP3, SF3B1, ADAD1, IL11RA, TRIM58, HERC2, RPRGIP1, MUC16, SHROOM1; Green plot ATM, PLEKHG5, NFATC1, FCGBP, BPI2, AMTN, MTUS1, SPTAN1; Purple plot SLITRK4, SEMA3E, ASXL1, MUSK, NPY, CHRNB2, ZNF534, FAT3. blue plot: non-coding mutations only

b, CLL077 a= before chlorambucil; b= before fludarabine, cyclophosphamide; c= immediately after 4 cycles of fludarabine, cyclophosphamide; d= before ofatumumab; e= relapse 9 months after ofatumumab.

Coding mutations: green plot: OCA2, SLC12A1, PLA2G16, DAZAP1, EXOC6B, LRRC16A; orange plot: NAMPTL, BCL2L13, GHDC; red plot: SAMHD1, IRF2BP2, GPR158; blue plot: MAP2K1, ZFHX4, HMCN1, DDX1, KHLDC2, NOD1, ZNF566, COL24A1; purple plot: non-coding mutations only.

c, CLL006 a= before fludarabine, cyclophosphamide; b=before Rituximab; c=before Ofatumumab; d= immediately after Ofatumumab; e= relapse 12 months after Ofatumumab.

Coding mutations: red plot: MED12, KLHL4, CNOT7, SLK1, U2AF1, C3orf43, PILRB, ARHGAP29, KIAA0182, MAP4, TMPRSS9; blue plot: PCLO, IRF4, LRRRC7B, KIAA0319L; green plot: RBPJ.

b: Mutation profiles based on deep sequencing in patient CLL003. Coloured boxes to the right of each plot indicate mutation profile type (HHF: red box, HL: yellow box, LH: green box, 0H: blue box). See Supplementary Table 5.
Figure 2 Schematic presentation of the changes in subclonal architecture over time

**a, d, f:** Schematic representation of the subclonal hierarchy for patients CLL003 (a), CLL077 (d) and CLL006 (f). Tumour subclones (red circles) are mapped to each stage and extrapolated back to the origin. The number beside each circle shows the percentage of cells calculated using mutant allele frequencies. Coloured boxes denote mutation profile groups (see Fig.1d and Supplementary Fig. 3).

**b, c, e, g, h:** Graphic illustration of absolute cell numbers for each subclone at all stages for patients CLL003 (b-c), CLL077 (e) and CLL006 (g-h). Plots are expanded for stage c of patient CLL003 (c) and stage c of patient CLL006 (h).
Figure 1

Figure 1 shows the cell count and somatic SNV allele frequency across different samples. The plots display the frequency of SNVs across various alleles (HH, HL, LH/0H, LL, 0H) for different samples (a, b, c, d, e). The graphs illustrate the variation in cell count and allele frequency, highlighting the genetic diversity and abundance of specific SNVs in each sample.

Key figures:
- **Figure 1a**: Cell count per 10^6 cells/litre across samples.
- **Figure 1b**: Somatic SNV allele frequency for HH allele.
- **Figure 1c**: Somatic SNV allele frequency for HL allele.
- **Figure 1d**: Somatic SNV allele frequency for LH/0H allele.
- **Figure 1e**: Somatic SNV allele frequency for LL allele.
- **Figure 1f**: Somatic SNV allele frequency for 0H allele.

The figures are color-coded to differentiate between different SNVs, with labels indicating the specific SNVs (e.g., ADAM1, SHROOM1, SF3B1, HNRNPA2, CHRNA8, IL11RA). The graphs provide a clear visual representation of the genetic landscape across the samples.
Figure 2

For personal use only. on April 15, 2017. by guest www.bloodjournal.org
Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns

Anna Schuh, Jennifer Becq, Sean Humphray, Adrian Alexa, Adam Burns, Ruth Clifford, Stephan M. Feller, Russell Grocock, Shirley Henderson, Irina Khrebtukova, Zoya Kingsbury, Shujun Luo, David McBride, Lisa Murray, Toshi Menju, Adele Timbs, Mark Ross, Jenny Taylor and David Bentley