

LC-FACSeq is a method for detecting rare clones in leukemia

Eileen Y. Hu,^{1,2} James S. Blachly,^{1,3} Caner Saygin,¹ Hatice G. Ozer,³ Stephanie E. Workman,¹ Arletta Lozanski,¹ Tzyy-Jye Doong,¹ Chi-Ling Chiang,¹ Seema Bhat,¹ Kerry A. Rogers,¹ Jennifer A. Woyach,¹ Kevin R. Coombes,³ Daniel Jones,⁴ Natarajan Muthusamy,¹ Gerard Lozanski,⁴ and John C. Byrd¹

¹Division of Hematology, Department of Internal Medicine and Comprehensive Cancer Center, ²Medical Scientist Training Program, ³Department of Biomedical Informatics, and ⁴Department of Pathology, Ohio State University, Columbus, Ohio, USA.

Detecting, characterizing, and monitoring rare populations of cells can increase testing sensitivity, give insight into disease mechanism, and inform clinical decision making. One area that can benefit from increased resolution is management of cancers in clinical remission but with measurable residual disease (MRD) by multicolor FACS. Detecting and monitoring genomic clonal resistance to treatment in the setting of MRD is technically difficult and resource intensive due to the limited amounts of disease cells. Here, we describe limited-cell FACS sequencing (LC-FACSeq), a reproducible, highly sensitive method of characterizing clonal evolution in rare cells relevant to different types of acute and chronic leukemias. We demonstrate the utility of LC-FACSeq for broad multigene gene panels and its application for monitoring sequential acquisition of mutations conferring therapy resistance and clonal evolution in long-term ibrutinib treatment of patients with chronic lymphocytic leukemia. This technique is generalizable for monitoring of other blood and marrow infiltrating cancers.

Introduction

In many types of cancer, disease relapse manifests with distinct genomic patterns of resistance and evolution. Relapse prediction and clinical decisions hinge on the ability to monitor clonal evolution in the setting of measurable residual disease (MRD) — small and persistent populations of leukemia cells whose presence directly correlates with disease progression and increased mortality. However, clinical detection of relapsed disease with small clones is challenging due to difficulties in obtaining enough cells to analyze for druggable targets. Available clinical DNA sequencing and MRD monitoring techniques lack the ability to easily characterize emergence of clonal evolution in rare tumor cells. To address this problem, we developed a highly sensitive, fast, and cost-effective technique to characterize rare residual tumor cells in the blood or marrow.

The most common approaches for monitoring clonal evolution in the setting of MRD in leukemia are surface phenotyping by multiparameter flow cytometry (MPFC), allele-specific oligonucleotide quantitative PCR (ASO-PCR), droplet digital PCR (ddPCR), and next-generation sequencing (NGS). Although fast, quantitative, and easily accessible, MPFC requires the analysis of millions of events to detect disease burden of 1×10^{-6} and is unable to characterize emerging clonal evolution of leukemia cells (1). ASO-PCR on the other hand requires high-quality DNA input and is restricted by patient-specific primer designs based on genetic mutations found in pretreatment baselines such as in the case of IGHV typing in acute lymphocytic leukemia and chronic lymphocytic leukemia (CLL) (2). Although ddPCR is both quantitative and qualitative, it is less applicable for examining multiple genomic targets (3). Last, conventional NGS approaches such as error-corrected NGS are highly sensitive but cost prohibitive and require a high degree of bioinformatics expertise for data analysis (4). Here, we present an assay that enables the use of limited cells acquired by FACS in tandem with targeted amplicon-based sequencing (LC-FACSeq), permitting the fast and reliable detection of more than 40 variants from as few as 50–300 sorted tumor cells by exclusion of a commonly applied step of nucleic acid extraction and purification that typically results in loss of genetic material.

Authorship note: NM, GL, and JCB are co-senior authors.

Conflict of interest: The authors have declared that no conflict of interest exists.

Copyright: © 2020, American Society for Clinical Investigation.

Submitted: November 15, 2019

Accepted: May 6, 2020

Published: June 18, 2020.

Reference information: *JCI Insight*. 2020;5(12):e134973.
<https://doi.org/10.1172/jci.insight.134973>.

Results

Assessing the reproducibility of LC-FACSeq in low numbers of cells. LC-FACSeq uses FACS to isolate pure populations of rare tumor cells, after which in situ lysis and amplicon-based library generation are performed. This allows us to amplify multiple target signals from rare subclones normally undetectable in DNA extracted from whole blood (Figure 1). We test the reproducibility of LC-FACSeq in the context of 50–500 sorted cells and demonstrate the applicability of broad, 29- and 43-gene panels for monitoring disease-related variants in CLL and acute myeloid leukemia (AML) cells, respectively (Supplemental Tables 1 and 2; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.134973DS1>).

CLL cells are phenotypically mature malignant B lymphocytes and can be easily identified as CD45⁺CD19⁺CD5⁺CD3[−] cells by flow cytometry. Despite having a low overall mutation burden, over 40 driver genes have been characterized in CLL, on which we based our target panel (5–8). Although the immature myeloid blasts of AML are more difficult to identify by flow cytometry, we define them as high side scatter cells that are both CD45⁺ and lineage marker-negative cells to test the performance of LC-FACSeq on different cell and panel constitutions.

To validate the reproducibility of this method with low cell numbers, we generate high-quality and high-depth libraries using DNA isolated from FACS purified bulk (number of cells >20,000) versus cell lysates from 50, 100, 200, 300, and 500 tumor cells isolated from patients with CLL and AML. All samples analyzed have an average read depth of 1212 (SEM = 56) per gene and an average coverage uniformity of 88.24% (SEM = 0.01%). To assess the theoretically expected deviation in variant allele frequency (VAF) when using limited cell numbers from a true gene variant frequency in a cell population, we created a Poisson/ β -binomial model with inputs of cell number (N), total number of sequenced reads for the variant of interest (M), the observed number of variant reads (U), and finally the gene variant zygosity (ς). As expected, our model shows that within a range of small cell numbers (50–500 cells), VAFs of rare (VAF <10%) homozygous variant from 300 and 500 cells show the smallest deviation from the true VAF (Figure 2A). Overall, unpurified cell lysates used in LC-FACSeq preserved both sequencing quality and VAF even when using as few as 50 cells, when compared with phenol/chloroform-extracted tumor cell DNA, which represents our true VAF (Figure 2B and Supplemental Tables 3 and 4). Violin plots show the median deviation, interquartile range (25th percentile, 75th percentile), for each cell titration as follows: CLL, 500 cells (−0.6%, [−3.9, 0.08]); CLL, 300 cells (0%, [−0.8, 0.1]); CLL, 100 cells (0%, [−1.5, 0.1]); CLL, 50 cells (0%, [−0.8, 0.1]); AML, 500 cells (0%, [−2.6, 0.1]); AML, 300 cells (−0.2%, [−3.5, 0.1]); AML, 200 cells (0%, [−2.0, 0]). Similar to the theoretical model, for our CLL panel, we observe that 300 cells showed little deviation in VAF when compared with bulk purified tumor cells. Because different panel sizes and constitution may have different optimal numbers of input cells, we select 300 cells for subsequent experiments with the 43-gene CLL panel. These data suggest that simple exclusion of the DNA isolation and purification step facilitates analysis of genomic heterogeneity inferred by bulk sequencing on a very small number of tumor cells.

Assessing performance of LC-FACSeq in detecting BTK p.C481S clones in ibrutinib-treated patients with CLL. Cytotoxic chemotherapy has been supplanted by targeted treatment in many types of cancer, which manifests with distinct genomic patterns of resistance. For reversible inhibitor targeted therapy, often a variety of mutations develop in both the target gene and the genes that either bypass or activate alternative pathways. With irreversible kinase inhibitors, this same pattern of resistance can occur in some diseases, whereas in others, a high frequency of mutation at the site of drug binding can occur (9, 10). Such is the case with CLL, in which the irreversible Bruton's tyrosine kinase (BTK) inhibitor ibrutinib is now widely used (11–13). Despite achieving durable responses in 90% of patients, only 10.2% of those patients achieve MRD-negative status (14–16). MRD-positive patients have limited (1–2/mm³), but persistent, leukemic cells. Patients with CLL and others with cancers who relapse after targeted therapy exhibit clonal, genetic evolution of residual cancer cells (17). Approximately 33% of patients who need to discontinue ibrutinib therapy do so due to disease progression or transformation (8, 18). A majority of patients who relapse while being treated with ibrutinib carry acquired mutations within BTK and/or phospholipase C γ 2 that gradually arise from small treatment-resistant subclones (8, 19–21). To test the sensitivity, specificity, and theoretical detection limit of LC-FACSeq in detecting rare resistant tumor clones, we analyze tumor samples from patients with previously characterized BTK p.C481S mutation. Sensitivity is 100% in 11 separately sorted and generated sets of 300-cell libraries from a single patient with a BTK mutation frequency of 32% (Supplemental Table 5). Specificity is 100% in 300-cell libraries generated from 9 separate patients without BTK mutations (Supplemental Table 6). To estimate the theoretical detection limits of LC-FACSeq, we sort known numbers

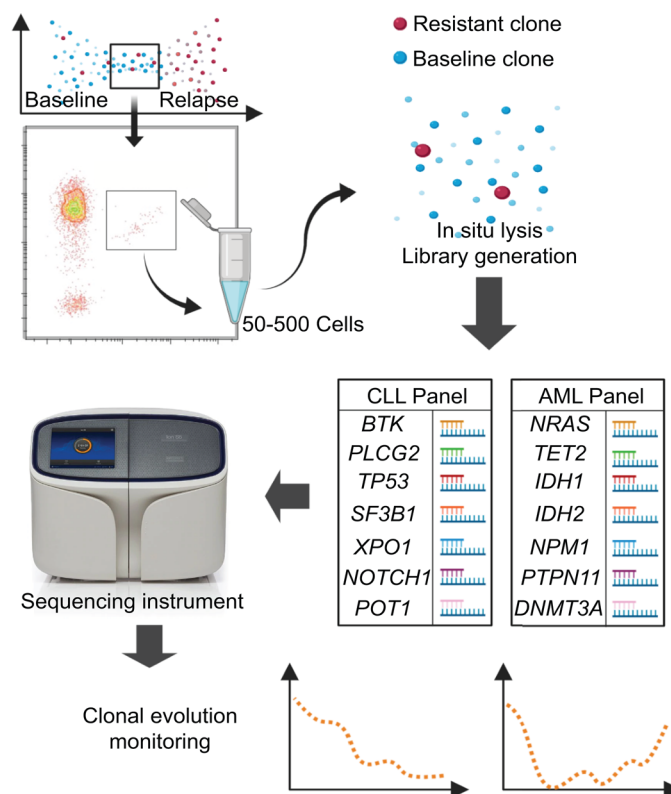
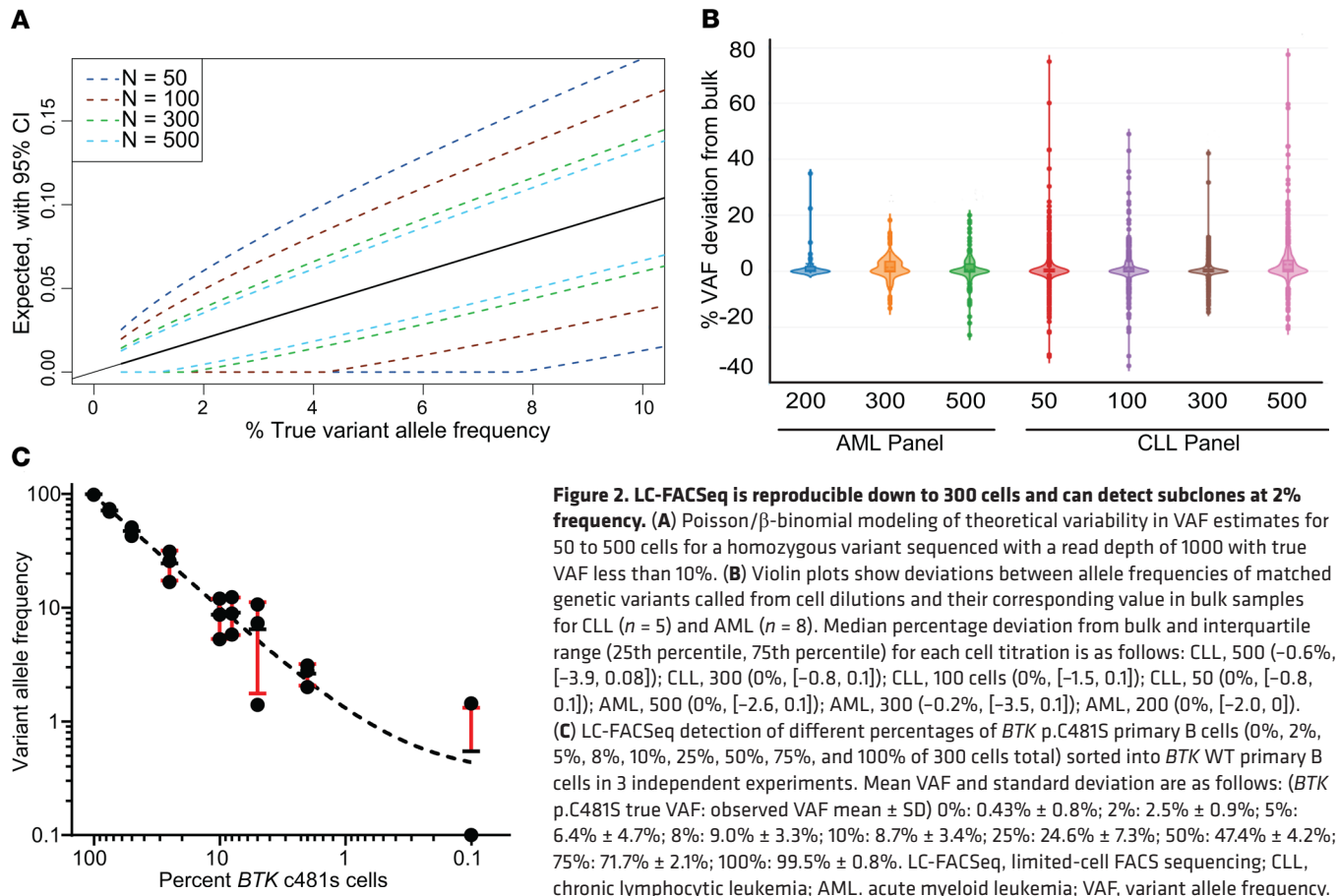


Figure 1. Graphic representation of LC-FACSeq workflow. Tumor cells are sorted from blood into lysis buffer, where targeted amplicon primer panels are added for subsequent sequencing. LC-FACSeq, limited-cell FACS sequencing.

of cells from an ibrutinib-refractory patient with *BTK* p.C481S (VAF = 100%) into WT *BTK* cells (VAF = 0%) to a total of 300 cells. For a sample with 10% *BTK*-mutated cells, we sort 30 mutated cells into 270 WT cells. We sort and sequence 9 different dilutions of mutant to WT cells, giving the following results: (*BTK* p.C481S true VAF: LC-FACSeq detected VAF mean \pm SD) 0% *BTK* p.C481S: 0.43% \pm 0.8%; 2% *BTK* p.C481S: 2.5% \pm 0.9%; 5% *BTK* p.C481S: 6.4% \pm 4.7%; 8% *BTK* p.C481S: 9.0% \pm 3.3%; 10% *BTK* p.C481S: 8.7% \pm 3.4%; 25% *BTK* p.C481S: 24.6% \pm 7.3%; 50% *BTK* p.C481S: 47.4% \pm 4.2%; 75% *BTK* p.C481S: 71.7% \pm 2.1%; 100% *BTK* p.C481S: 99.5% \pm 0.8%. We show that LC-FACSeq reliably amplifies signals from 2% mutated cells (Figure 2C and Supplemental Table 7) among 300 total cells. From these data, we can estimate that for a patient with an MRD burden of 1×10^{-4} (1 leukemic cell in 10,000 total WBC), if 2% of the successfully sorted cells are resistant subclones, then LC-FACSeq can potentially detect 2×10^{-6} (2 treatment-resistant cells in 1,000,000 total WBC), which is comparable to highly sophisticated techniques such as error-corrected NGS.

Using LC-FACSeq to characterize clonal evolution in ibrutinib-treated patients with CLL. We next use LC-FACSeq to detect ibrutinib-resistant tumor cells from baseline and treatment blood samples from 20 MRD-positive patients with CLL on ibrutinib therapy (Supplemental Table 8). MRD measurements for leukemic cells range from 4 to 25,874 CD19⁺CD5⁺ cells/mm³ blood, as determined by clinical multiparameter flow cytometry. High VAF *BTK* p.C481S mutations were clinically observed in the latest available on-treatment samples of 4 patients. Clonal evolution is defined as genes that were changed by greater than 5%, had PolyPhen-2 scores >0.5, and were predicted to be pathogenic in the ClinVar database (22, 23). Using LC-FACSeq we observe canonical CLL-associated clonal mutations similar to those observed in previous studies (Supplemental Table 9) (5–8, 24). Of the 20 MRD-positive patients, 7 show subclonal changes in *TP53*, *NOTCH1*, *POT1*, *SF3B1*, and *MYD88* over the course of ibrutinib treatment, although we found no correlation or consensus in these clonal shifts, suggesting continual clonal suppression by ibrutinib (Figure 3). We retrospectively analyzed cells from 4 patients that had either relapsed on or developed tumor subclones resistant to ibrutinib. Leukemic counts for these 4 patients range from 4 cells/mm³ blood during remission to over 80,000 cells/mm³ blood at treatment baseline, as determined by standard clinical multiparameter flow cytometry. Analysis of CLL 01 show the emergence of subclones bearing *BTK* mutations as early as 10 months before clinical relapse was identified (indicated by the red arrow). Notably, resistant subclones are detectable by LC-FACSeq in all 4 patients despite clinically stable WBC counts and CD19⁺ B cell percentages (Figure 4 and Supplemental Tables 10 and 11).



In 2 patients (CLL 02 and CLL 03), we saw an expansion of multiple subclonal *BTK* mutations resulting in *BTK* p.C481S. These results are compared with those generated by the clinical standard of care *BTK* mutation screen performed at our institution that differs from less sensitive techniques used commercially, where CD19 magnetic bead selection of B cells occurs followed by a mutation-specific quantitative PCR assay (ddPCR) (Supplemental Table 11). Compared with this ultrasensitive, Ohio State University institution-specific mutational assay, LC-FACSeq is equally effective at detecting the presence of mutant *BTK* p.C481S. These data demonstrate that LC-FACSeq can be leveraged to detect important treatment-resistance variants with low cell input well in advance of signs of clinical relapse.

Discussion

It is critical to detect emerging therapy resistance and disease evolution to inform clinical course. LC-FACSeq is a pipeline that leverages both FAC-sorting and amplicon-based genome amplification in the detection of more than 40 possible variants from as few as 50–300 sorted tumor cells derived from blood or bone marrow of patients with CLL or AML. The ability to accomplish this analysis for DNA-specific mutations comes from eliminating the processing step of isolating DNA with nucleic acid extraction and purification before generation of the analysis library. Although quite simple, description of eliminating this step in DNA genomic analysis has not been described. Using this approach and introducing it as part of LC-FACSeq enables clinical decisions based on the detection of multiple targetable mutations within very small tumor populations comparable to highly sophisticated techniques such as error-corrected NGS.

Most methods of fast and cost-effective clonal tracking in the setting of MRD are limited to the detection of a few genes of interest from purified genomic DNA. However, LC-FACSeq enables the use of broad gene panels and can reliably amplify signals from as few as 50. Using 300 cells for our 43-gene CLL panel, LC-FACSeq was able to amplify variant signals if just 6 of those cells carried the target variant. From these data, we estimated that if we could successfully sort out 300 cells from an MRD-positive patient, then a 2% variant allele could be extrapolated to the existence of a much rarer subclone. However, this estimate

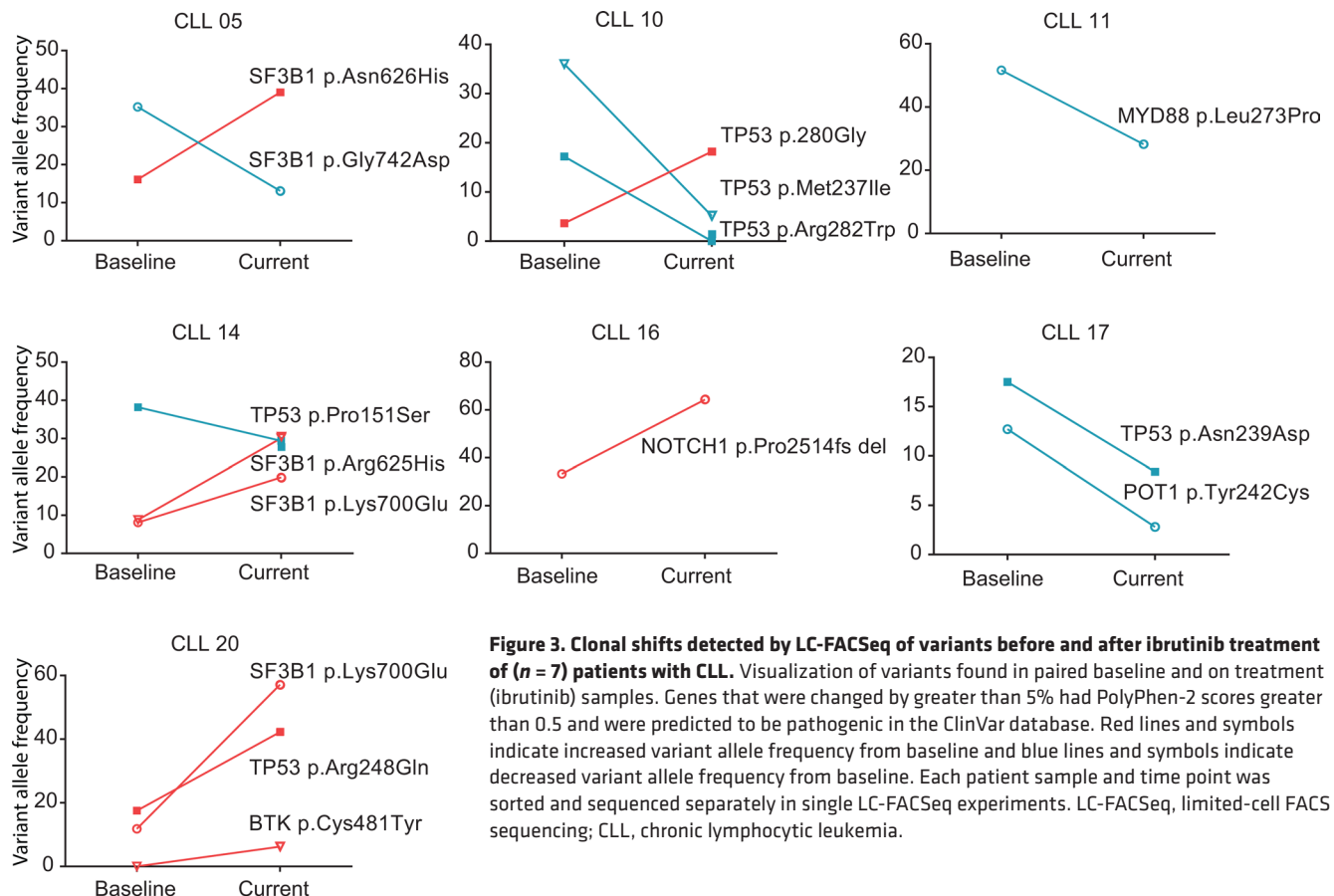


Figure 3. Clonal shifts detected by LC-FACSeq of variants before and after ibrutinib treatment of ($n = 7$) patients with CLL. Visualization of variants found in paired baseline and on treatment (ibrutinib) samples. Genes that were changed by greater than 5% had PolyPhen-2 scores greater than 0.5 and were predicted to be pathogenic in the ClinVar database. Red lines and symbols indicate increased variant allele frequency from baseline and blue lines and symbols indicate decreased variant allele frequency from baseline. Each patient sample and time point was sorted and sequenced separately in single LC-FACSeq experiments. LC-FACSeq, limited-cell FACS sequencing; CLL, chronic lymphocytic leukemia.

is depends on a number base assumptions. Rare subclones must be present in circulation and within the patient sample. These rare cells must also be efficiently sorted and evenly sampled with FACS. Sampling limitations may necessitate the use of higher numbers of cells to sufficiently sample rare tumor subclones. Furthermore, as the frequency of the subclone decreases, the quantitative nature of LC-FACSeq decreases. Of our cell dilutions, samples with less than 10% *BTK* p.C481S cells showed an increase in VAF variability. Not surprisingly, in one of our 100% WT *BTK* samples, we see a VAF of approximately 1.4% for the mutant allele, signifying a base level of sequencing noise. These observations suggest that low VAF detection is more qualitative than quantitative and precludes the ability to accurately quantify exact subclone frequency. Going forward, it will be critical to rigorously determine real-world limits of detection on large patient cohorts with low MRD levels. Additionally, future studies would need to measure and establish thresholds to differentiate true variant signals from signaling noise.

In our study, we tested the applicability and performance of LC-FACSeq in MRD-positive, ibrutinib-treated patients with CLL. Ibrutinib therapy has revolutionized the treatment of CLL, but for reasons unknown, most patients treated with ibrutinib have persistent, low levels of MRD for years after therapy initiation. Whole-exome sequencing in patients with CLL within the first 6 months of ibrutinib treatment initiation found that, for the most part, clonal dynamics of CLL cells are stable after the first month of therapy (8). Clonal dynamics years after treatment initiation have never been examined due to difficulties in obtaining enough cells for analysis. Herein in a small number of patient samples, we have found that the majority of patients did not exhibit clonal evolution up to 6 years after treatment initiation. However, the amplicon-based sequencing technique of LC-FACSeq limited our analysis to previously described CLL driver mutations (5–8). Compared with whole-genome sequencing, targeted sequencing approaches such as the amplicon-based sequencing leveraged by LC-FACSeq has increased sensitivity and are capable of detecting signals from limited genomic materials. Conversely, the nature of targeted sequencing limits variant detection to only known mutations and genome regions included within sequencing primer pools and not true novel discoveries. Our target panel for CLL was based on previous whole-exome sequencing studies characterizing the genomic landscape of ibrutinib-treated patients with CLL. Within our patient cohort,

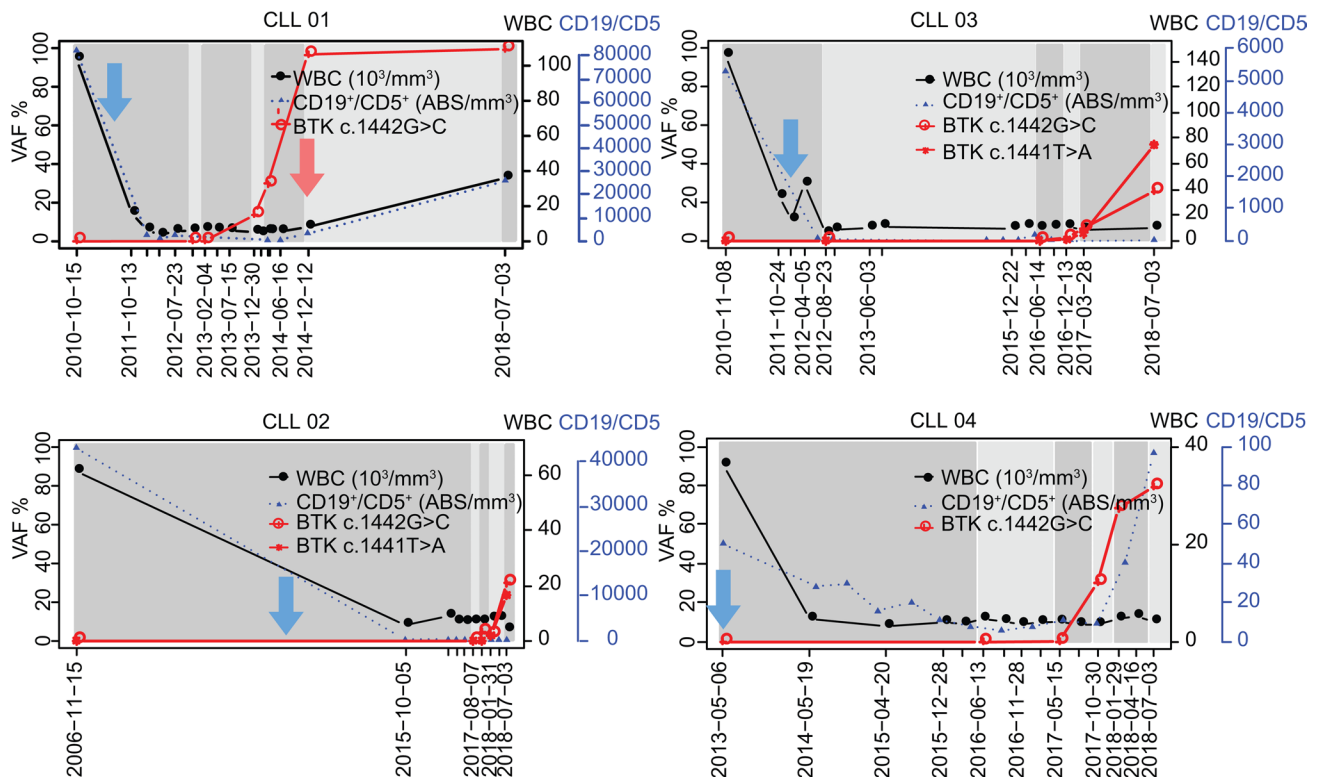


Figure 4. LC-FACSeq can detect early presence of ibrutinib-resistant subclones despite clinically stable disease. Visualization of *BTK* p.C481S variants found in serial samples of ($n = 4$) patients previously clinically determined to have developed ibrutinib resistance. Time points are depicted on the x axis as year/month/date. Left y axis and red line/points denote the average measures for *BTK* p.C481S VAF percentage at each time point from 2 independent experiments. Different single nucleotide polymorphisms resulting in *BTK* p.C481S mutations are denoted with a red square or red asterisk. The black right y axis and black line/points denote the absolute values of clinically determined WBCs (1×10^3 cells/mm³ blood) for each time point. Right y axis and blue dashed line/points denote the absolute count of leukemic CD19⁺CD5⁺ B cells as determined by clinical multiparameter flow cytometry. Blue arrows show time of ibrutinib therapy initiation and red arrows show time of clinical relapse. Each sequenced time point is the average of the variant allele frequency detected in 2 independent experiments. LC-FACSeq, limited-cell FACS sequencing; VAF, variant allele frequency.

it is entirely possible that the evolution of variants was occurring outside of the regions contained in our panel. Indeed, there are multiple noncoding variants associated with CLL that will have to be considered for future studies (24). Future clinical utilization of LC-FACSeq will likely initially be more suitable for tracking well-characterized and actionable targets.

One of the actionable targets we focused on was the single nucleotide substitution that results in a cysteine to serine amino acid change in the ibrutinib-binding domain of *BTK*. This is a well-characterized mutation hotspot in ibrutinib-treated CLL samples that has been shown in multiple studies to affect therapy resistance and disease progression (19). Its immediate clinical relevance and increased prevalence in ibrutinib-refractory samples led us to use it as a standard variant for testing the sensitivity, specificity, and detection limit of LC-FACSeq. However, these parameters can vary among different genomic loci (due to primer quality, unbiased PCR amplification) and variant structure (small indels). Further assessment with different targets and genes would have to be rigorously assessed for each genomic locus used on future panels.

LC-FACSeq analysis of CLL demonstrates its application in detecting and tracking therapy resistance-related clonal evolution in the setting of MRD that could potentially alter treatment strategies. In addition to its applications in monitoring therapy-resistant clones in CLL, the LC-FACSeq approach can potentially be extended to monitor clonal evolution in small but disease-driving populations including Lin[−]CD34⁺ cells in AML to identify mutations that drive tumor relapse such as those present after allogeneic stem cell transplant. Similarly, using a phenotype such as Ki67 (a marker of proliferation) and determining if mutations are enriched in such a small population of cells are possible (25). LC-FACSeq is low cost, fast, and accessible for clinical labs with little bioinformatics support. Compared with current techniques, it is low input, highly sensitive, easily standardized, and customizable to accommodate our evolving knowledge of therapy resistance and disease drivers in multiple blood and bone marrow-derived malignancies.

Methods

Human samples. PBMCs were obtained from patients with CLL and AML in accordance with the Declaration of Helsinki. Blood from patients with CLL and AML were collected at The Ohio State University Comprehensive Cancer Center. PBMCs were isolated using Ficoll density-gradient centrifugation, Ficoll-Paque Plus (GE Healthcare), and cryopreserved. Cryopreserved patient primary PBMCs were thawed and briefly cultured in 37°C and 5% CO₂ in RPMI 1640 and supplemented with 10% heat-inactivated FBS, 2 mM L-glutamine (Invitrogen), and 56 U/mL penicillin and 56 µg/mL streptomycin (Invitrogen) after isolation and before sorting. Clinical flow cytometry used methodology to measure malignant B cells has been previously described (26). Similarly, the clinical laboratory methods used to detect *BTK* and other related mutations have been previously described (26, 27).

FACS. Cell sorting was performed using a BD Biosciences BD FACSARIA Fusion flow cytometer with a blue, red, yellow-green, violet, and UV laser. PBMCs from patients with CLL and AML were selected and sorted with the following antibodies: Brilliant Violet 510 anti-human CD19 (clone HIB19, 302242, BioLegend); Alexa Fluor 700 anti-human CD3 (clone UCHT1, 344822, BioLegend); PE/Cy7 anti-human CD5 (clone UCHT2, 364008, BioLegend); PE anti-human CD45 (clone HI301, 555483, BD Biosciences); APC anti-human CD45 (clone HI30, 561864, BD Biosciences); Pacific Blue anti-human Lineage Cocktail (348805, BioLegend); and LIVE/DEAD Fixable Near-IR Dead Cell Stain (L34975, Invitrogen). CLL cells were identified as live, CD3⁻, CD45⁺, CD19⁺, and CD5⁺ cells. AML leukemic blasts were defined as CD45⁺ and Lin⁻ cells. Briefly, cryopreserved or fresh mononuclear cells from patients with CLL or AML were stained with antibody cocktail on ice for 30 minutes before being washed once and resuspended in DPBS (Gibco) at a concentration of 5 million cells/mL. Cells were sorted through an 85-µm nozzle at a sheath pressure of 45 psi and a drop drive frequency of 60 to 70 kHz. A highly pure sorting modality (single-cell sorting for FACSARIA Fusion) was used for cell numbers less than 1000 into PCR strip tubes. Four-way purity sorting was used for bulk samples (defined as >20,000 cells) into 1.5-mL Eppendorf tubes. A representative sorting scheme for a CLL sample can be seen in Supplemental Figure 1. Because treatment samples have extremely low numbers of CLL cells, the sort purity and cytometer function was frequently determined by sorting patient autologous CD45⁺CD3⁺CD5⁺CD19⁻ T cells (Supplemental Figure 2). Sorted purity achieved for all samples was greater than 98.0%.

DNA extraction and preparation. For bulk tumor samples (defined as >20,000 sorted cells), DNA was extracted from isolated CLL and AML cells using the Gentra Puregene Tissue Kit (158667, QIAGEN) according to the manufacturer's recommendations for less than 1 million cells (Gentra Puregene Handbook). DNA was quantified using the NanoDrop 2000 (Thermo Fisher Scientific) using standard 260/280 OD ratios. For first-step amplification, 10 ng of DNA from bulk samples was used. For samples with low cell numbers (defined as <1000 sorted cells), DNA was prepared from sorted cells using the Single Cell Lysis Kit (4458235, Invitrogen). Briefly, 9 µL of Single Cell Lysis solution (no DNase I) was directly added to sorted cell droplets. After incubating for 5 minutes at room temperature, 1 µL of Stop Solution was added to the reaction for 2 minutes at room temperature. For first-step amplification, 5 µL of the total reaction was used.

ddPCR. For *BTK* C481S mutation detection, the referent ddPCR method was performed on genomic DNA following B cell purification from blood using anti-CD19–positive selection (RoboSep, StemCell Technologies). ddPCR was performed on the Raindance Sense and Source instruments (Bio-Rad) using 2-color custom-designed primers with normalization of the mutant events to total mutant plus nonmutant events. The lower reportable range for the assay was 0.1% mutant ratio with at least 10 mutant events, with linearity established with ongoing 4-point standard curves.

Ion Torrent sequencing and analysis. Sequencing and analysis of the “CLL panel” and “AML panel” were performed using the NGS Ion Torrent platform and reagents from Life Technologies. Libraries were prepared with the Ion AmpliSeq Library Kit 2.0 (4475345, Ion Torrent) with a custom-designed panel of AmpliSeq primers (CLL: IAD14429, 304 amplicons, 2 pools, 31.66-kb panel size, 99.63% coverage; AML: IAD145356, 129 amplicons, 2 pools, 15.32-kb panel size, coverage 100%; Supplemental Tables 1 and 2). To run multiple samples on the same chip, we used IonExpress barcode adapters (4471250 and 4474009). Libraries were constructed by first amplifying genomic DNA preparations with individual primer pools. Amplification was completed on the GeneAmp PCR System 9700 Dual 96-well thermal cycler (Applied Biosystems), using the following protocol: initial hold at 99°C for 2 minutes followed by 21 cycles for the CLL panel (99°C for 15 seconds, 60°C for 4 minutes), 20 cycles for the AML panel (99°C for 15 seconds, 60°C for 4 minutes), and a final hold at 10°C. Corresponding PCR product pools were then combined and

barcoded with Ion Xpress Barcode Adapters (Ion Torrent). Barcoded PCR products were purified with an Agencourt AMPure XP kit (A63881, Beckman Coulter). While bulk-extracted DNA libraries were eluted with low TE at this point and directly quantified, libraries constructed with less than 1000 cells underwent an additional amplification cycle to equalize library amounts using the following protocol: initial hold at 98°C for 2 minutes followed by 5 cycles (99°C for 15 seconds, 64°C for 1 minute) and a final hold at 10°C. All libraries were quantified using real-time PCR with the Ion Library TaqMan Quantitation Kit (44688022, Ion Torrent) on the Applied Biosystems ViiA7 Real-Time PCR System instrument, to allow for optimal final dilution of library for template preparation. Diluted libraries were loaded onto the OneTouch OT2 version instrument with either the Ion PGM Template OT2 200 Kit (4480974) or the Ion 530/540 Kit OT2 (A27751/A27753). ISP enrichment and purification were performed on Ion One Touch 2 ES. Purified ISPs were analyzed on the Ion Torrent personal Genome Machine using the Ion PGM Hi-Q Sequencing kit (A25591) with 318 Chip V2 (4484354) and 316 Chip V2 (4483188). ISPs were also analyzed with the Ion Torrent S5 instrument using either the 530/540 Kit OT2 (A27751/A27753) or the Ion 530/540 Chip Kit (A27764/A27766). Data were collected and analyzed using the Torrent Server (4462616) with Torrent Suite Software (versions 5.0 and 5.6.0). Final analysis of sequence data was performed using a combination of the following software: Variant Caller v5.0.28 and v5.6.8-1, IGV v5.01, Ion Reporter v4.6, Samtools v1.6, Bcftools v1.31, and Ensembl Variant Effect Predictor v92.03. The hg19 reference sequence was used for analysis. The entire length of sequences was reviewed manually using these programs to assess for deviation from reference sequence and to evaluate the quality of sequence and the depth of coverage.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE149545 (28).

Poisson/ β -binomial model. The observed data for a single variant of interest consists of M , the total number of observed reads that cover the variant, and U , the observed number of variant reads. The parameter we are interested in is ψ , the true but unknown VAF. We typically estimate ψ by computing the expected value $E[U/M]$, but we would like to understand the precision and accuracy of this estimate. For this purpose, we take $U \mid M, \nu, \omega \sim \text{BetaBinomial}(M, \nu, \omega)$. Because M itself varies depending on the actual read depth achieved, we take $M \sim \text{Poisson}(\lambda)$, where λ is the targeted read depth, often on the order of 1000 or more. To proceed, we let N denote the number of cells selected by design (typically between 50 and 500). We let ς denote the zygosity of the variant (so $\varsigma = 1$ for a heterozygous variant and $\varsigma = 2$ for a homozygous variant). For convenience, we also define $\phi = \nu + \omega$. Using the method of moments, we can force the beta distribution to optimally approximate the (unobserved) binomial experiment that describes the number of variant cells among the N selected cells. Then we can show that $\nu = \phi\psi$, that $\omega = \phi(1 - \psi)$, and that $\phi = -1 + N(1 - \psi)/(\varsigma/2 - \psi)$. We can show that $E[U] = \lambda\psi$ by using the Law of Total Expectation and that $V[U] = \lambda\psi + \lambda^2\psi(1 - \psi)/(1 + \phi)$ by using the Law of Total Variance. By applying Taylor series expansions to the ratio, we can show that $E[U/M] \approx \psi + \psi/\lambda - \psi/\lambda\varsigma$, and that $V[U/M] \approx V[U]/\lambda^2 + \psi^2/\lambda(1 - 2/\varsigma)$. In the homozygous case ($\varsigma = 2$), the bias is on the order of $\psi/2\lambda < 0.001$. The second equation provides a tight estimate on the variance of the estimator.

Study approval. All human subjects gave written informed consent for the blood products to be used for research under an institutional review board–approved protocol at The Ohio State University (Columbus).

Statistics. Data are represented as individual data points, means \pm SD, or median and interquartile range where appropriate. Figure 1 was created with Biorender.com. Figure 2C and Figure 3 were created with GraphPad Prism 8. R statistical software was used to aggregate data and prepare Figure 2, A and B, and Figure 4.

Author contributions

EYH designed, conducted and analyzed experiments, and wrote the manuscript. CS designed, conducted, analyzed experiments, and edited the manuscript. HGO analyzed experiments and edited the manuscript. AL and TD conducted and analyzed experiments, and edited the manuscript. CC analyzed experiments and edited the manuscript. SB, KAR, JAW provided clinical samples and edited the manuscript. KRC created the Poisson/ β -binomial statistical model. SW provided clinical data points for each sample sequenced. DJ ran clinical flow cytometry and ddPCR sequencing on clinical samples. JSB, NM, GL, and JCB conceptualized and designed the experiments, and edited the manuscript. All authors analyzed the data, reviewed the manuscript, and agree with submission for publication.

Acknowledgments

We thank Jihad Skaf for his help in conceptualizing LC-FACSeq. Funding support was provided by R35-CA197734, R01-CA177292, F30-CA225070, UG1-CA233338, Four Winds Foundation, Connie Brown CLL Fund, Leukemia and Lymphoma Society, and the D Warren Brown Foundation. KAR is a scholar in clinical research of the Leukemia & Lymphoma Society.

Address correspondence to: John C. Byrd, 455B Wiseman Hall//CCC, 400 West 12th Avenue, Columbus, Ohio 43210, USA. Phone: 614.293.8330; Email: John.Byrd@osumc.edu.

1. Hallek M, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008;111(12):5446–5456.
2. van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJ. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia*. 2003;17(6):1013–1034.
3. Drandi D, et al. Minimal residual disease detection by Droplet Digital PCR in multiple myeloma, mantle cell lymphoma, and follicular lymphoma: a comparison with real-time PCR. *J Mol Diagn*. 2015;17(6):652–660.
4. Duncavage EJ, et al. Mutation clearance after transplantation for myelodysplastic syndrome. *N Engl J Med*. 2018;379(11):1028–1041.
5. Puente XS, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101–105.
6. Landau DA, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714–726.
7. Landau DA, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525–530.
8. Landau DA, et al. The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nat Commun*. 2017;8(1):2185.
9. Westover D, Zugazagoitia J, Cho BC, Lovly CM, Paz-Ares L. Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors. *Ann Oncol*. 2018;29(suppl_1):i10–i19.
10. Schwartz PA, et al. Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance. *Proc Natl Acad Sci U S A*. 2014;111(1):173–178.
11. Woyach JA, et al. Bruton's tyrosine kinase (BTK) function is important to the development and expansion of chronic lymphocytic leukemia (CLL). *Blood*. 2014;123(8):1207–1213.
12. Herman SE, et al. Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood*. 2011;117(23):6287–6296.
13. Woyach JA, et al. Ibrutinib regimens versus chemoimmunotherapy in older patients with untreated CLL. *N Engl J Med*. 2018;379(26):2517–2528.
14. Byrd JC, et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N Engl J Med*. 2013;369(1):32–42.
15. Ahn IE, et al. Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase 2 study. *Blood*. 2018;131(21):2357–2366.
16. O'Brien S, et al. Single-agent ibrutinib in treatment-naïve and relapsed/refractory chronic lymphocytic leukemia: a 5-year experience. *Blood*. 2018;131(17):1910–1919.
17. Burger JA, et al. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat Commun*. 2016;7:11589.
18. Jain P, et al. Outcomes of patients with chronic lymphocytic leukemia after discontinuing ibrutinib. *Blood*. 2015;125(13):2062–2067.
19. Woyach JA, et al. Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. *N Engl J Med*. 2014;370(24):2286–2294.
20. Maddocks KJ, et al. Etiology of ibrutinib therapy discontinuation and outcomes in patients with chronic lymphocytic leukemia. *JAMA Oncol*. 2015;1(1):80–87.
21. Kadri S, et al. Clonal evolution underlying leukemia progression and Richter transformation in patients with ibrutinib-relapsed CLL. *Blood Adv*. 2017;1(12):715–727.
22. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–249.
23. Landrum MJ, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(database issue):D980–D985.
24. Puente XS, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519–524.
25. Giné E, et al. Expanded and highly active proliferation centers identify a histological subtype of chronic lymphocytic leukemia ("accelerated" chronic lymphocytic leukemia) with aggressive clinical behavior. *Haematologica*. 2010;95(9):1526–1533.
26. Rogers KA, et al. Phase 1b study of obinutuzumab, ibrutinib, and venetoclax in relapsed and refractory chronic lymphocytic leukemia. *Blood*. 2018;132(15):1568–1572.
27. Jones D, et al. PLCG2 C2 domain mutations co-occur with BTK and PLCG2 resistance mutations in chronic lymphocytic leukemia undergoing ibrutinib treatment. *Leukemia*. 2017;31(7):1645–1647.
28. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–210.