

https://dspace.org/

CloneRetriever An Automated Algorithm to Identify Clonal B and T Cell Gene Rearrangements by NextGeneration Sequencing for the Diagnosis of Lymphoid Malignancies

Eitan Halper-Stromberg; Chad M McCall; L Haley; Ming-Tseh Lin; Wendy Stevens ...

https://hdl.handle.net/10539/35318
WOS

Downloaded from a DSpace repository. https://dspace.org/

Clinical Chemistry 67:11
1524–1533 (2021)

CloneRetriever: An Automated Algorithm to Identify Clonal B and T Cell Gene Rearrangements by Next-Generation Sequencing for the Diagnosis of Lymphoid Malignancies

Eitan Halper-Stromberg [a], Chad M. McCall,[b] Lisa M. Haley [a], Ming-Tseh Lin,[a] Samantha Vogt,[c] Christopher D. Gocke,[a,d] James R. Eshleman,[a,d] Wendy Stevens,[e] Neil A. Martinson,[c,f] Marta Epeldegui,[g] Matthias Holdhoff,[d] Chetan Bettegowda,[h,i] Michael J. Glantz,[j] Richard F. Ambinder,[d] and Rena R. Xian [a,d,*]

BACKGROUND: Clonal immunoglobulin and T-cell receptor rearrangements serve as tumor-specific markers that have become mainstays of the diagnosis and monitoring of lymphoid malignancy. Next-generation sequencing (NGS) techniques targeting these loci have been successfully applied to lymphoblastic leukemia and multiple myeloma for minimal residual disease detection. However, adoption of NGS for primary diagnosis remains limited.

METHODS: We addressed the bioinformatics challenges associated with immune cell sequencing and clone detection by designing a novel web tool, CloneRetriever (CR), which uses machine-learning principles to generate clone classification schemes that are customizable, and can be applied to large datasets. CR has 2 applications—a "validation" mode to derive a clonality classifier, and a "live" mode to screen for clones by applying a validated and/or customized classifier. In this study, CR-generated multiple classifiers using 2 datasets comprising 106 annotated patient samples. A custom classifier was then applied to 36 unannotated samples.

RESULTS: The optimal classifier for clonality required clonal dominance $\geq$4.5$\times$ above background, read representation $\geq$8% of all reads, and technical replicate agreement. Depending on the dataset and analysis step, the optimal algorithm yielded sensitivities of 81%–90%, specificities of 97%–100%, areas under the curve of 91%–94%, positive predictive values of 92–100%, and negative predictive values of 88%–98%. Customization of the algorithms yielded 95%–100% concordance with gold-standard clonality determination, including rescue of indeterminate samples. Application to a set of unknowns showed concordance rates of 83%–96%.

CONCLUSIONS: CR is an out-of-the-box ready and user-friendly software designed to identify clonal rearrangements in large NGS datasets for the diagnosis of lymphoid malignancies.

Introduction

Lymphoid malignancies harbor uniquely rearranged immunoglobulin (Ig), or T-cell receptor (TCR), genes that serve as clonal markers of disease. Identification of clonal rearrangements is critical to the diagnosis and monitoring of lymphoblastic leukemia, lymphoma, and multiple myeloma. Given the clinical importance, Ig and TCR clonality assessment by Southern blot and PCR has been performed for over 20 years (1, 2). With the advancements in massively parallel sequencing, several next-generation sequencing (NGS)-based technologies (3–6) are now available to study lymphoid malignancies, as well as interrogate the immune repertoire (7). NGS technologies can improve sensitivity and specificity of primary diagnosis (5, 8, 9), and can also monitor treatment response (10–12). Such techniques have enabled the study of tumor-derived cell-free DNA

[a]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD; [b]Department of Pathology, Duke University School of Medicine, Durham, NC; [c]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD; [d]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD; [e]Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa; [f]Perinatal HIV Research Unit (PHRU), University of the Witwatersrand, Johannesburg, South Africa; [g]Department of Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, CA; [h]Department of Neurosurgery, Johns Hopkins School of Medicine, Baltimore, MD; [i]Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins School of Medicine, Baltimore, MD; [j]Department of Neurosurgery, Medicine, and Neurology, Penn State Milton S. Hershey Medical Center, Hershey, PA.
*Address correspondence to this author at: Molecular Diagnostics Laboratory, Johns Hopkins Genomics, 1812 Ashland Ave., Suite 200, Baltimore, MD 21205, USA. E-mail rxian1@jhmi.edu.
Received April 6, 2021; accepted July 10, 2021.
DOI: 10.1093/clinchem/hvab141

©American Association for Clinical Chemistry 2021. All rights reserved. For permissions, please email: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

in patients with lymphoma, even longer V(D)J rearrangements exceeding 150 bases (13), supporting a role for Ig clonality in liquid biopsies (13–15).

To date, PCR remains the gold-standard method for Ig and TCR clone detection in clinical practice. Although this technique was standardized in 2003 (1, 16, 17), and early attempts were made to address interpretation challenges (18), consensus interpretation guidelines remained elusive for nearly a decade thereafter (19). Collaborative efforts from the EuroClonality-NGS working group (20) are beginning to standardize immune cell sequencing for minimal residual disease detection, but there are no available recommendations for the interpretation of NGS data for the diagnosis of lymphoid malignancies, despite good concordance between PCR and NGS methods (5, 9, 20). The lack of standardization of the technique and/or interpretation guidelines contributes to the limited adoption of Ig and TCR NGS for the diagnosis of lymphoid malignancies.

A major problem encountered with Ig and TCR sequencing is that the large datasets generated by NGS require several steps of data processing before human-readable outputs can be reviewed. This often requires customized bioinformatics pipelines bespoke to specific sequencing chemistries, assays, and research and clinical questions. To address this problem, several publicly available tools have been developed and customized for the Ig and TCR loci to perform immune repertoire profiling and clonality detection (21–24). Some of these tools are able to process raw FASTQ files to generate aligned and filtered clonotypes (24), while other tools streamline workflows, incorporate data quality analyses, as well as statistical and plotting functions that can summarize large cohorts (23, 25). However, very few groups have addressed clonality interpretation in the NGS era. Those that have, have used a variety of criteria, thresholds, and interpretation caveats without clear consensus (5, 9, 11, 12, 20). Different groups have used absolute read depth, relative read depth, and relative clone dominance, as well as other features to determine clonality. This small number of variables makes an attempt at recapitulating the pattern recognition that an experienced reviewer applies to the interpretation of PCR-based clonality assays. Nevertheless, subtle variations in interpretation, such as oligoclonal and gray-zone calls, are lost this way.

With an eye toward future standardization of NGS clone interpretation, we have developed an automated algorithm that simplifies the identification of clonal Ig and TCR gene rearrangements. This tool addresses obstacles related to clonal criteria determination and clone detection on a large scale, and returns some of the visual cues of PCR-based techniques back in control of the user. The present study summarizes the design, development, and validation of CR using distinct patient cohorts, and demonstrates the accuracy and flexibility of this intuitive and novel tool.

Methods

SAMPLE SELECTION AND SEQUENCING

A total of 142 deidentified patient samples (92 from site 1, Johns Hopkins University and 50 from site 2, Duke University) were studied after appropriate institution review. DNA was extracted from 58 plasma, 31 formalin fixed paraffin embedded (FFPE) tissue, and 3 cerebral spinal fluid from site 1, and 23 FFPE, 16 whole blood, 7 bone marrow aspirate, and 4 fresh tissue from site 2 (summarized in Tables S1 and S2 in the online Data Supplement). Sequencing libraries from site 1 were prepared using 10–50 ng of input DNA per reaction using the LymphoTrack® immunoglobulin heavy chain (*IGH* FR1, FR2, FR3) MiSeq panel (Invivoscribe Inc.) per vendor protocols with minor adjustments. Sequencing libraries from site 2 were prepared using 50 ng of input DNA per reaction using either the LymphoTrack *TRG*-PGM, *IGH* FR1-PGM, *IGH* FR3-PGM or *IGK*-PGM assays (Invivoscribe) following vendor protocols. See online Supplemental Table S1 and the Supplemental Methods.

GOLD-STANDARD DATA INTERPRETATION

As a reference standard for comparison against CR-derived interpretations, the following criteria were used (Table 1). In brief, at site 1, visual assessment of clonality was performed by 2 molecular pathologists. The top 200 unique reads from each sequencing reaction were used to generate histograms using GraphPad Prism (v.8.4.3, GraphPad Software). A sample was considered clonal if $\geq$1 primer set(s) was clonal, either mono- or bi-allelic, with the same sequence(s) in both replicates. When clonal dominance was deemed borderline by visual review, the top 10 merged reads were reviewed to confirm that the candidate clone was $\geq$4× the fourth ranked sequence. This threshold was adopted based on previously published studies (5, 12). Dominant sequences that were found in both replicates but were on average <4× the fourth ranked sequence were deemed indeterminate (IND). In brief, at site 2, clonality interpretation was performed by one laboratory technologist followed by one pathologist according to previously validated clinical laboratory criteria. A sample was deemed clonal if $\geq$1 primer set(s) were clonal in singlicate. Primer sets showing candidate clones that met 1, but not both, criteria were deemed IND. Seventy-five percent (106/142) of the samples were annotated before CR analyses (online Supplemental Table S3).

CLONERETRIEVER DATA PROCESSING

CloneRetriever (CR) is a python-based automated pipeline used to analyze Ig and TCR gene rearrangement

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

Table 1. Gold-standard clonal criteria.

	Site 1	Site 2
Clonal criteria		
Clone rank	1st or 2nd	1st or 2nd
Repeatability	Found in duplicate as 1st or 2nd	NA
Clonal dominance	≥4× over 4th	≥5× over 5th (B cell) ≥6.1× over 4th (T cell)
Read representation	Not applicable	AND ≥ 5% of all reads
Indeterminate criteria		
Clone rank	1st or 2nd	1st or 2nd
Repeatability	Found in duplicate as 1st or 2nd	Not applicable
Clonal dominance	<4× over 4th	≥5× over 5th (B cell) ≥6.1× over 4th (T cell)
Read representation	Not applicable	OR ≥ 5% of all reads
Clonal sample (B cell)	≥1 clonal framework for *IGH*	≥1 clonal framework for *IGH* or clonal *IGK*
Features for optimization		
Repeatability	Yes vs. no	Not applicable
Clonal dominance	Fold change (×)	Fold change (×)
Read representation	%	%

sequences from NGS. CR is deployed as a web application (https://clone-retriever.herokuapp.com) using the Heroku cloud platform (Salesforce) and the Streamlit python library (Streamlit Inc.). Two applications, "validation" and "live," are available. To assess the functionality of validation mode, NGS data files (top 10 merged read files) from 106 annotated samples were processed through CR to train and validation clonal classifiers (Fig. 1). Each row/sequence within a sequencing file was annotated by gold-standard interpretation as clonal (1), nonclonal (0), or IND (0). In this portion of the study, IND sequences were grouped as nonclonal. Seventy-three percent (41/56) of the annotated samples from site 1, all of the plasma samples, served to train a classifier using 3 features: repeatability, clonal dominance (fold change above fourth ranked sequence), and read representation (merged count of sequence divided by total count). Fragment length and V, or J gene, usage were not considered for this analysis. See online Supplemental Methods.

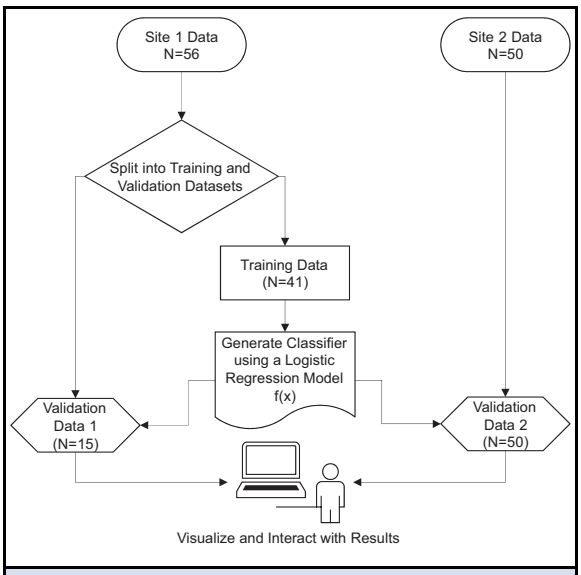

Fig. 1. Experimental design. B- and T-cell receptor gene rearrangement sequencing data from a total of 106 mixed-type samples were used to train and validate a clonality classifier using CloneRetriever.

CLONERETRIEVER DATA ANALYSIS TO TRAIN A CLASSIFIER

Using these inputs, a logistic regression classification was performed using the 3 aforementioned features and the sequence-level annotation of clonality (0 or 1) as the outcome. The classifier mimics visual review considerations by using "AND" among the features, as opposed to "OR". To input this logic into the classifier, and be able to subsequently return discrete values for each variable, a single predictor term that encapsulates all 3 features was derived. This term was defined as $f(\chi)$ = clonal dominance × repeatability × read representation (online Supplemental Methods). After calculating this term for each sequence, one half of the training data

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

was used to train a classifier. To maximize sensitivity as a screening tool, the best classifier was defined as the one yielding the highest sensitivity in the first half of the training data with ≥0.8 specificity in the remaining half. Confidence intervals were then generated for each classifier using a bootstrapping approach.

CLONERETRIEVER DATA ANALYSIS TO VALIDATE THE CLASSIFIER

To determine whether 1 or more features yielded higher sensitivity/specificity, 3 possible classifiers were generated from the training data. These classifiers were applied to the remaining annotated samples from site 1 (validation 1) and all samples from site 2 (validation 2). A fourth classifier was created to illustrate the role of repeatability in validation 1. CR returns sequence-, primer-set-, and sample-level performance and graphs, as CR was designed to collapse replicate results, and multiple Ig, or TCR, targets for a sample to generate aggregate primer-set-level and sample-level calls. Interactive controls in CR allow further customization of a classifier to best fit a given dataset. The custom classifier generated for site 1 was then applied to the remaining unannotated samples (39%, 36/92) from site 1 to assess CR "live" mode. The CR calls on the sequence, primer-set and sample levels were exported from CR, and provided to 2 pathologists who performed consensus data review. The pathologists compared the CR results against visual interpretations using gold-standard criteria (Table 1).

Results

SEQUENCING DATA REPRESENTED DIVERSE SAMPLE TYPES, TARGETS, AND CLONAL OUTCOMES

To assess generalizability of CR, data generated from 2 different sites using different targets (online Supplemental Tables S1 and S2) and different gold-standard interpretation criteria (Table 1) were analyzed. A total of 142 samples (92 from site 1 and 50 from 2; 106 annotated and 36 unannotated) were studied comprising 62% fresh tissue samples and 38% formalin fixed samples (online Supplemental Table S2). Samples from site 1 comprised 55 nonhodgkin lymphoma, 16 classical Hodgkin lymphoma, and 21 controls without lymphoma. Samples from site 2 were blinded to diagnoses, but comprised a mixture of lymphoid malignancies and nonlymphoid processes. Seventy-one percent of samples underwent Ig sequencing (*IGH* and/or *IGK*), and 29% underwent TCR sequencing (*TRG*) generating 572 sequencing files (504 files from site 1 totaling 4861 rows of sequences in duplicate; 68 files from site 2 totaling 683 rows of sequences in singlicate). For the training and validation studies, clonally annotated samples ($n = 96$) were selected to mimic real-world conditions with 31% defined as clonal, 59% as nonclonal, and 9% as IND for clonality using gold-standard criteria (online Supplemental Table S3). Representative histograms that were used in the visual assessment of clonality at site 1 are shown in Fig. 2. These 96 samples yielded 356 individual sequencing files (288 files from site 1 totaling 2756 rows of sequences generated in duplicate; 68 files from site 2 totaling 683 rows of sequences generated in singlicate) that were submitted to CR to train and validate a classifier.

A THREE-FEATURE CLASSIFIER PERFORMED BETTER THAN 2- OR 1-FEATURE CLASSIFIERS

Seventy-three percent (41/56) of the annotated samples from site 1, all plasma samples, totaling 1875 sequences in replicate, served to train a classifier in a divided manner based on sequence-level sensitivity and specificity described in the Methods. Training data yielded 3 different classifiers. Since data for validation 2 were only available in singlicate, repeatability was always presumed to be true, or 1, for those data, and the impact of repeatability would not be directly assessed here (see later). After training, the best classifier (Table 2) that used both parameters (and assumed repeatability) corresponded to a clonal dominance of ≥4.5× [95% CI = (3.8×, 7.2×)] and a read representation ≥0.08 [95% CI = (0.02,0.08)]. Training using only one parameter showed classifiers that corresponded to higher thresholds, either 16.9× [95% CI = (8.9×, 18.1×)] clonal dominance, or 0.14 [95% CI = (0.12,0.20)] read representation, respectively, and wider confidence intervals. The sensitivities and specificities of all 3 classifiers applied to the training data were similarly high (95–100% sensitivity, 99–100% specificity). These classifiers were then applied to the remaining annotated samples (site 1: 27%, 15/56, 881 rows of sequences; site 2: 100%, 50/50, 683 rows of sequences) as validation data 1 and 2, respectively. Once these classifiers were applied to the validation data, performance characteristics changed (Table 2 and online Supplemental Table S4). The three-feature classifier (classifier 1) showed superior performance in both datasets. Whereas, classifier 3 outperformed classifier 2 in validation 1, and classifier 2 outperformed classifier 3 in validation 2. Since the impact of repeatability could not be formally assessed above, we performed a subanalysis only using the 56 samples from site 1 to generate classifier 4 that only used repeatability. Repeatability resulted in 100% sensitivity across all levels of analysis, although specificity reduced to 92%, 91%, and 71% respectively. On the sequence level, this classifier showed the lowest specificity (92%), as compared to the other classifiers.

Since these classifiers were trained exclusively using plasma samples, we next evaluated the effect of sample

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

Table 2. Performance characteristics using classifier-generated thresholds.

		Sequence level					Primer level					Sample level				
		%Sensitivity	%Specificity	%AUC	%PPV	%NPV	%Sensitivity	%Specificity	%AUC	%PPV	%NPV	%Sensitivity	%Specificity	%AUC	%PPV	%NPV
Classifier 1																
Clonal dominance: 4.5x 95% CI = [3.8×, 7.2×] Read representation: 0.08 95% CI = [0.02, 0.08] Repeatability: Required*	Validation 1	81	99	91	93	98	85	100	92	100	94	88	100	94	100	88
	Validation 2	86	99	92	96	98	90	98	94	94	96	86	97	92	92	95
Classifier 2																
Clonal dominance: 16.9× 95% CI = [8.9×, 18.1×] Read representation: NA Repeatability: Required*	Validation 1	75	95	96	65	97	77	88	93	71	90	75	57	79	67	67
	Validation 2	86	100	100	100	97	84	100	100	100	94	86	100	100	100	95
Classifier 3																
Clonal dominance: NA Read representation: 0.14 95% CI = [0.12,0.20] Repeatability: Required*	Validation 1	81	93	91	62	97	85	75	89	58	92	88	57	88	70	80
	Validation 2	71	98	98	87	95	79	98	98	94	92	79	100	98	100	92
Classifier 4																
Clonal dominance: NA Read representation: NA Repeatability: Required	Validation 1	100	92	96	64	100	100	91	95	81	100	100	71	86	80	100
	Validation 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

*Repeatability was presumed true for all clonal sequences present in validation 2; PPV: positive predictive value; NPV: negative predictive value.

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

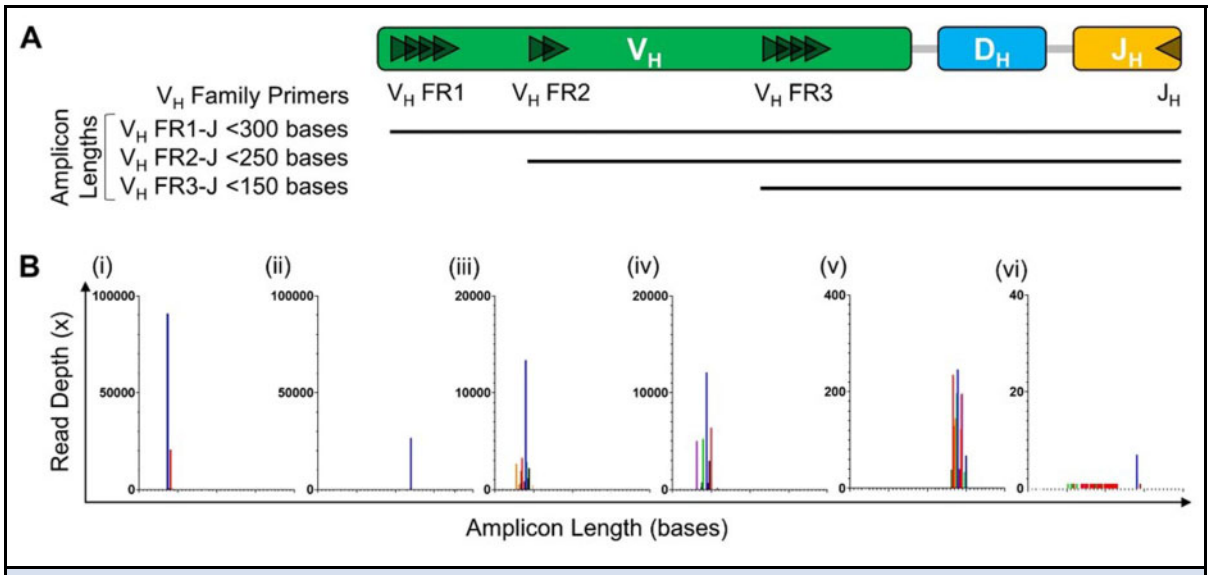

Fig. 2. IGH sequencing and gold-standard clonality determination (Site 1). (A), Primers used to sequence *IGH*. (B), Representative histograms used for clonality determination. Unique VDJ sequences are separated by amplicon length (0–400 bases) versus read depth. (i) Clonal, bi-allelic. (ii) Clonal, mono-allelic, (iii). Clonal, mono-allelic. (iv). Indeterminate. (v). Nonclonal, polyclonal. (vi). Nonclonal, pseudoclonal. All profiles are reproduced in replicate (data not shown).

type on classification schemes by repeating training and validation using a more similar admixture of samples. The new training cohort included 29 plasma, 10 FFPE, and 2 cerebral spinal fluid samples from site 1. Again, 4 different classifiers (5–8) were generated, and applied to the remaining annotated validation data (online Supplemental Table S5). The new classifiers showed similar improvements in performance using a 3-feature classifier, as opposed to a 2- or 1-feature classifier. Overall, these new classifiers demonstrated superior performance in the validation data, particularly on the sample level, and showed tighter 95% CI compared to the original classifiers 1–4.

CUSTOMIZATION OF THE CLASSIFIER IDENTIFIED INDETERMINATE (GRAY-ZONE) SAMPLES

Since CR classifiers were only trained to differentiate clonal vs nonclonal results, IND clones were "lost" in the nonclonal category in the previous analysis. In practice, borderline findings are occasionally encountered. As it would be inappropriate to classify these samples as either clonal or nonclonal (5, 19), CR-generated classifiers can be modified to rescue these gray-zone samples (Fig. 3). To capture the IND samples (2 from site 1, 8 from site 2), manual manipulation (Fig. 4, A) of features and thresholds was performed in CR to customize a classifier that best fit the respective data. This exercise showed that both datasets performed best when a 2-feature classifier was applied (Fig. 3 and online Supplemental Fig. S1). Clonal dominance and repeatability best delineated site 1 data, and read representation showed no added benefit; whereas clonal dominance and read representation best delineated site 2 data, as repeatability was not considered. Data from site 1 showed that despite substantial overlap in the $\log_2$ ratios of fold change among the 3 clonality outcomes, the custom classifier showed 95% concordance with the annotated sample-level interpretations. Data from site 2 showed a more distinct $\log_2$ ratio of fold change breakpoint between clonal and nonclonal samples, although indeterminate samples overlapped with nonclonal samples. The site 2 custom classifier showed 100% concordance in clonality assignment.

APPLICATION OF THE CUSTOM CLASSIFIER TO UNKNOWN SAMPLES SHOWED HIGH CONCORDANCE

The site 1 custom classifier was applied to the remaining samples ($n = 36$) from site 1, which were annotated (online Supplemental Fig. S3). From a total of 647 candidate clones (rank 1–3), 96% (619/647) were correctly assigned as clonal, nonclonal, or IND. On the primer set and sample levels, CR showed 92% and 83% concordance with visual interpretations, respectively. Discrepancies were related to differing thresholds for clonal dominance (4.7× by CR vs 4× by visual review), differences in clonotype read counts based on merged (CR) and unmerged (visual review) reads, requirements for threshold agreement between replicates (visual review), as opposed to single/best-replicate (CR), and last

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

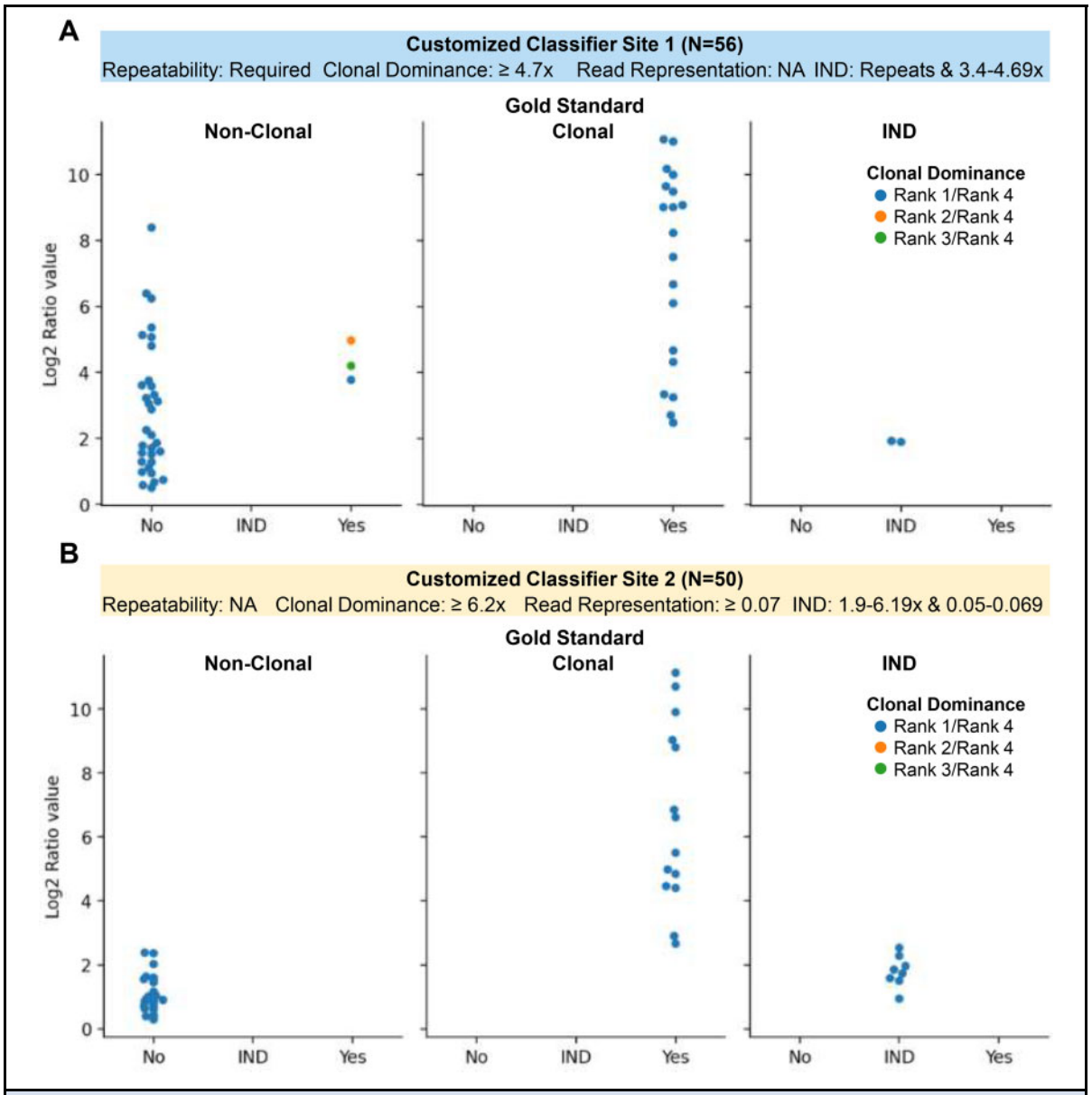

Fig. 3. Custom classifiers in CloneRetriever. Sample-level results from all available data after applying site-specific custom classifiers. (A), Custom classifier for site 1. (B), Custom classifier for site 2. Panels represents gold-standard clone evaluation. The *x* axis shows either nonclonal (No), indeterminate (IND), or clonal (Yes) results based on the custom classifier, which are plotted against clonal dominance of the top sequence (log_2 ratio).

the lack of size range or V–J mapping parameters in CR. For example, during visual review, the presence of a potential clone outside of the expected primer-set size range, or V-only or J-only sequence alignment, would exclude a sequence from being clonal irrespective of clonal dominance and/or repeatability. Besides exporting data, "live" mode of CR has additional web-based functions, including a simple control panel (Fig. 4, A) and plotting functions, such as a "Merge Count vs Sample" graph, which allow users to rapidly identify potentially clonal sequences and samples (Fig. 4B), and likely nonclonal samples (online Supplemental Fig. S2).

Discussion

To address the analysis and interpretation needs presented by Ig and TCR rearrangement NGS, CR was designed as an out-of-the-box ready and user-oriented software useful for clonality interpretation. This study summarized the design, training and validation of

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

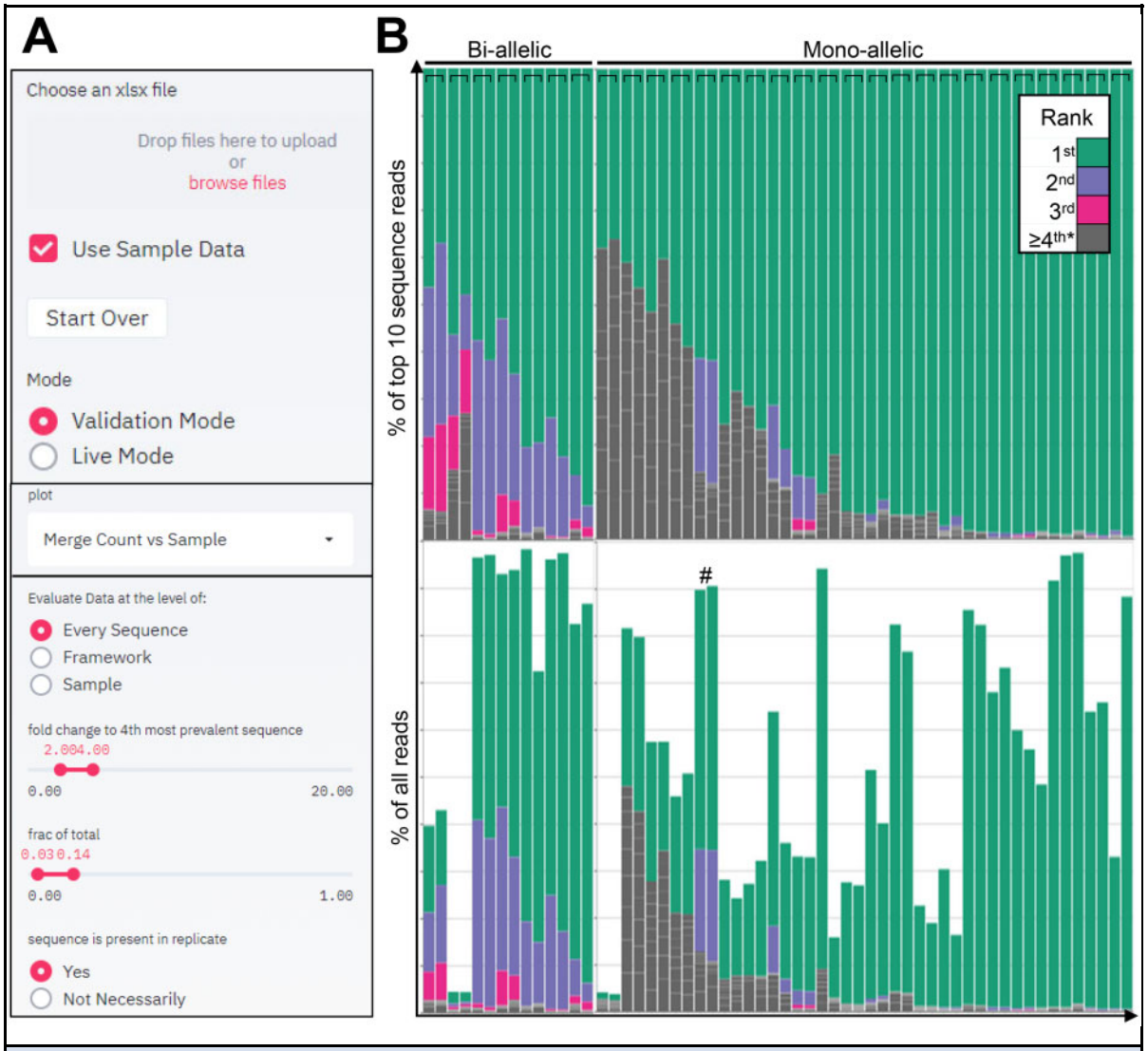

Fig. 4. CloneRetriever control panel and plotting. (A), Control panel. (B), Representative plots of clonal patterns (bi- or mono-allelic) when replicate data are available. See online Supplemental Materials for detailed description. *Nonrepeatable sequences and rank ≥4th sequences are colored gray. # Mono-allelic clone with subclone.

clonality classifiers, and additional functionality of CR, as applied to 2 distinct datasets. Despite fundamental technical and analytical differences, including gold-standard clonal criteria, a single "best" classifier was generated that could be applied to both datasets. Yet, different custom classifiers could also be created that best fit, and most accurately assessed clones from, the respective datasets. By mirroring gold-standard interpretation criteria, CR was able to identify key clonal features inherent to particular datasets, and return these features as discrete and meaningful values to the reviewer. Returning these metrics on the sequence-, primer-set-, and sample-level allows users to easily review the validity of the classifier, and modify it accordingly. To our knowledge, CR is the only bioinformatics tool available that simplifies clonality criteria determination and clone detection for the diagnosis of lymphoid malignancies.

Strengths of this study are multiple. The datasets varied by sample type, sequencing targets, sequencing platform, clonal composition, and clonality criteria. This afforded ample diversity in the datasets, and showcased the flexibility of CR's algorithm. Despite the fact a single sample type (plasma cell-free DNA) and target (*IGH*) was used for training, the resultant classifiers could be applied to other sample types and targets with good accuracy. Matching the sample types of the

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

training and validation datasets changed thresholds and confidence intervals suggesting sample type-related differences in clonal criteria. Since CR can train on any data admixture, users can determine how best to design their training and validation cohorts, including target-specific analyses. The ability to customize CR-generated classifiers enabled rescue of IND samples, and use in "live" mode in a set of unknown samples. The high concordance seen in clonal calls highlighted the possibility of using CR as a screening tool.

Additional strengths of this study include the use of highly salient features needed for clonality, which were selected based on previously published data (5, 9, 11, 12, 20) and expert opinion. The annotation of sequences was of high-quality owing to independent analyses by experienced reviewers. Although the algorithm for clone classification is machine-learning-driven, the algorithm design was purpose-built to retain the original meaning of the features thereby avoiding the impenetrable "black-box" criticism associated with certain machine-learning algorithms (26). Returning results based on decreasing levels of granularity (sequence > primer set > sample) enables rapid assessment of classifier performance by an analyst. The end-result is simplification and streamlining of a workflow that integrates existing accumulated knowledge of clone calling in the context of NGS into a single web tool. Unlike previous studies examining clonality interpretation and diagnostic performance of Ig and TCR NGS, this work does not give recommendations on optimal thresholds, nor specific features of clonality. Rather, CR provides a fixed algorithm and framework to manipulate large datasets and extract meaningful clone-calling results.

Weaknesses of this study could include limitations on generalizability. Although the samples studied were diverse, a single vendor manufactured all sequencing primers, and a comparable raw data processing pipeline was used. Unlike software that performs end-to-end clone calling from FASTQ files (23, 24), CR only uses processed data files. The performance of CR as a clone-calling tool using processed data generated by other methods remains unknown, which can be addressed in future studies. Other limitations relate to the hard-coded CR algorithm. The combination of features selected by CR may be considered simplistic, as additional features, such as productivity and degree of somatic hypermutation, can also factor into clone determination (27–29). Currently, the CR algorithm only accepts and returns clonal and nonclonal results forgoing subtleties, such as oligoclonality and indeterminate clones. While indeterminate clones are readily rescued by the user by customizing a CR-generated classifier, oligoclonal samples may currently be masked by CR as nonclonal. CR calculates clonal dominance based on the fourth ranked sequence. In oligoclonal samples, the fourth sequence may actually represent the fourth clone, and CR may erroneously label such a sample as nonclonal due to the absence of dominant clones. Lack of fragment length parameters and requirements for complete V(D)J alignment could result in over- and under-calling by CR. Last, lack of any quality metric cut-offs, such as total read count, may assign clonality values to samples that should receive a no-result classification (5).

In this initial description of this tool, we focused on a complete software that accepts postprocessed data to identify important features of clonality, and apply these features to screen unannotated datasets for clones. The fallibility of an automated pipeline, irrespective of downstream user-control, needs to be considered. As such, CR should only be used as a screening tool, until such time as extensive clinical validation can be completed. Future versions of the tool may incorporate data quality thresholds, secondary statistical analyses, such as confidence intervals, additional clonality features such as fragment length, V(D)J alignment and productivity, and exploratory features that refine the determination of indeterminate and oligoclonal samples. As it stands, CR is a ready-to-use solution that can ease the bioinformatics burden of assessing Ig and TCR clonality by NGS. Integrating CR into practice broadly will require further demonstration of ease of use, generalizability, flexibility, accuracy, and efficiency, all of which will be explored in future studies. Should tools like CR become widely implemented, this may create a shared framework in which consensus NGS interpretation guidelines can be developed in the future for the diagnosis of lymphoid malignancies.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: CR, CloneRetriever; NGS, next-generation sequencing; IND, indeterminate; FFPE, formalin fixed paraffin embedded;

Human Genes: *IGH*, Immunoglobulin heavy chain; *IGK*, Immunoglobulin kappa locus; *TRG*, T-cell receptor gamma locus.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: R.R. Xian, Invivoscribe Inc.

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021

Research Funding: U.S. Department of Health & Human Services | NIH | National Cancer Institute (NCI) - R01CA250069, R21CA232891, P30CA006973, P30AI094189, UM1CA121947, R21CA220475, U01AI035040, and R01CA228157. N.A. Martinson has received institutional research funding from Pfizer Inc.
Expert Testimony: C.M. McCall has served as an expert witness for private organizations.
Patents: None declared.
Other Remuneration: R.R. Xian, Invivoscribe Inc.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors would like to thank the patients for participating in our study, including patients from Johns Hopkins hospital, Duke University hospital, the Chris Hani Baragwanath Hospital and Milton S. Hershey Medical Center. The authors would like to thank the Duke University Clinical Molecular Diagnostics laboratory for performing the sequencing assays, and the Duke University Molecular Laboratory Directors and Pathologists, Drs. Siby Sebastian, Michael Datto, Jadee Neff, and Anand Shreeram Lagoo, for performing the gold-standard interpretation of the Duke sequencing results. The authors would also like to thank Dr. Roger Detels and Dr. Otoniel Martinez-Maza for providing access to appropriate archival specimens from the UCLA MACS repository.

References

1. Van Dongen J, Langerak A, Brüggemann M, Evans P, Hummel M, Lavender F, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;17:2257–317.
2. Nikiforova MN, Hsi ED, Braziel RM, Gulley ML, Leonard DGB, Nowak JA, et al.; Molecular Pathology Resource Committee, College of American Pathologists. Detection of clonal IGH gene rearrangements–summary of molecular oncology surveys of the college of American pathologists. Arch Pathol Lab Med 2007;131:185–9.
3. Faham M, Willis T, inventors. Monitoring health and disease status using clonotype profiles. US Patent Application US20140235454A1 2014 Aug 21.
4. Scheijen B, Meijers RWJ, Rijntjes J, van der Klift MY, Mobs M, Steinhilber J, et al.; EuroClonality-NGS Working Group. Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: a technical feasibility study by EuroClonality-NGS. Leukemia 2019;33:2227–40.
5. Arcila M, Yu W, Syed M, Kim H, Maciag L, Yao J, et al. Establishment of Immunoglobulin Heavy (IGH) chain clonality testing by next-generation sequencing for routine characterization of B-cell and plasma cell neoplasms. J Mol Diagn 2019;21:330–42.
6. Cavagna R, Guinea Montalvo ML, Tosi M, Paris M, Pavoni C, Intermesoli T, et al. Capture-based next-generation sequencing improves the identification of immunoglobulin/t-cell receptor clonal markers and gene mutations in adult acute lymphoblastic leukemia patients lacking molecular probes. Cancers 2020;12:1505–23.
7. Johansson G, Kaltak M, Rîmniceanu C, Singh AK, Lycke J, Malmeström C, et al. Ultrasensitive DNA immune repertoire sequencing using unique molecular identifiers. Clin Chem 2020;66:1228–37.
8. Sufficool KE, Lockwood CM, Abel HJ, Hagemann IS, Schumacher JA, Kelley TW, et al. T-cell clonality assessment by next-generation sequencing improves detection sensitivity in mycosis fungoides. J Am Acad Dermatol 2015;73:228–36. e2.
9. Schumacher JA, Duncavage EJ, Mosbruger TL, Szankasi PM, Kelley TW. A comparison of deep sequencing of TCRG rearrangements vs traditional capillary electrophoresis for assessment of clonality in T-cell lymphoproliferative disorders. Am J Clin Pathol 2014;141:348–59.
10. Faham M, Zheng J, Moorhead M, Carlton VE, Stow P, Coustan-Smith E, et al. Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. Blood 2012;120:5173–80.
11. Bruggemann M, Kotrova M, Knecht H, Bartram J, Boudjogrha M, Bystry V, et al.; EuroClonality-NGS working group. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. Leukemia 2019;33:2241–53.
12. Rustad EH, Hultcrantz M, Yellapantula VD, Akhlaghi T, Ho C, Arcila ME, et al. Baseline identification of clonal V(D)J sequences for DNA-based minimal residual disease detection in multiple myeloma. PLoS ONE 2019;14:e0211600.
13. Vogt SL, Patel M, Lakha A, Philip V, Omar T, Ashmore P, et al. Feasibility of cell-free DNA collection and clonal immunoglobulin sequencing in South African patients with HIV-associated lymphoma. JCO Glob Oncol 2021;7:611–21.
14. Kurtz DM, Green MR, Bratman SV, Scherer F, Liu CL, Kunder CA, et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Blood 2015;125:3679–87.
15. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, et al. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. Lancet Oncol 2015;16:541–9.
16. Evans PA, Pott C, Groenen PJ, Salles G, Davi F, Berger F, et al. Significantly improved PCR-based clonality testing in B-cell malignancies by use of multiple immunoglobulin gene targets. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 2007;21:207–14.
17. van Krieken JH, Langerak AW, Macintyre EA, Kneba M, Hodges E, Sanz RG, et al. Improved reliability of lymphoma diagnostics via PCR-based clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 2007;21:201–6.
18. Lee S-C, Berg KD, Racke FK, Griffin CA, Eshleman JR. Pseudo-spikes are common in histologically benign lymphoid tissues. J Mol Diagn 2000;2:145–52.
19. Langerak AW, Groenen PJ, Bruggemann M, Beldjord K, Bellan C, Bonello L, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. Leukemia 2012;26:2159–71.
20. Lay L, Stroup B, Payton JE. Validation and interpretation of IGH and TCR clonality testing by Ion Torrent S5 NGS for diagnosis and disease monitoring in B and T cell cancers. Pract Lab Med 2020;22:e00191.
21. Chaudhary N, Wesemann DR. Analyzing immunoglobulin repertoires. Front iImunol 2018;9: 462.
22. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'Connor KC, Hafler DA, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 2014;30:1930–2.
23. Bystry V, Reigl T, Krejci A, Demko M, Hanakova B, Grioni A, et al.; EuroClonality-NGS. ARResT/Interrogate: an interactive immunoprofiler for IG/TR NGS data. Bioinformatics 2017;33:435–7.
24. IJspeert H, Van Schouwenburg PA, van Zessen D, Pico-Knijnenburg I, Stubbs AP, van der Burg M. Antigen receptor galaxy: a user-friendly, web-based tool for analysis and visualization of T and B cell receptor repertoire data. J Immunol 2017;198:4156–65.
25. Knecht H, Reigl T, Kotrova M, Appelt F, Stewart P, Bystry V, et al.; EuroClonality-NGS Working Group. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. Leukemia 2019;33:2254–65.
26. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15.
27. Brázdilová K, Plevová K, Francova HS, Kockova H, Borsky M, Bikos V, et al. Multiple productive IGH rearrangements denote oligoclonality even in immunophenotypically monoclonal CLL. Leukemia 2018;32:234–6.
28. Davi F, Langerak AW, de Septenville AL, Kolijn PM, Hengeveld PJ, Chatzidimitriou A, et al.; ERIC, the European Research Initiative on CLL, and the EuroClonality-NGS Working Group. Immunoglobulin gene analysis in chronic lymphocytic leukemia in the era of next generation sequencing. Leukemia 2020;34:2545–51.
29. Katsibardi K, Braoudaki M, Papathanasiou C, Karamolegou K, Tzortzatou-Stathopoulou F. Clinical significance of productive immunoglobulin heavy chain gene rearrangements in childhood acute lymphoblastic leukemia. Leuk Lymphoma 2011;52:1751–7.

Downloaded from https://academic.oup.com/clinchem/article/67/11/1524/6365845 by UNIV OF WITWATERSRAND user on 08 December 2021