Perspective

sensitive detection of residual lymphoproliferative disease by NGS and clonal rearrangements-how low can you go?

Hansen, Marcus H; Cédile, Oriane; Larsen, Thomas S; Abildgaard, Niels; Nyvold, Charlotte G

Published in:
Experimental Hematology

DOI:
10.1016/j.exphem.2021.03.005

Publication date:
2021

Document version:
Final published version

Experimental Hematology

Experimental Hematology 2021;98:14–24

PERSPECTIVE

Perspective: sensitive detection of residual lymphoproliferative disease by NGS and clonal rearrangements—how low can you go?

Marcus H. Hansen[a,b], Oriane Cédile[a,b], Thomas S. Larsen[b], Niels Abildgaard[a,b], and Charlotte G. Nyvold[a,b]

[a]*Hematology–Pathology Research Laboratory, Research Unit for Hematology and Research Unit for Pathology, University of Southern Denmark and Odense University Hospital, Odense, Denmark;* [b]*Department of Hematology, Odense University Hospital, Odense, Denmark*

(Received 2 February 2021; revised 22 March 2021; accepted 30 March 2021)

Malignant lymphoproliferative disorders collectively constitute a large fraction of the hematological cancers, ranging from indolent to highly aggressive neoplasms. Being a diagnostically important hallmark, clonal gene rearrangements of the immunoglobulins enable the detection of residual disease in the clinical course of patients down to a minute fraction of malignant cells. The introduction of next-generation sequencing (NGS) has provided unprecedented assay specificity, with a sensitivity matching that of polymerase chain reaction-based measurable residual disease (MRD) detection down to the 10^{-6} level. Although reaching 10^{-6} to 10^{-7} is theoretically feasible, employing a sufficient amount of DNA and sequencing coverage is placed in the perspective of the practical challenges when relying on clinical samples in contrast to controlled serial dilutions. As we discuss, the randomness of subsampling must be taken into account to accommodate the sensitivity threshold—in terms of both the required number of cells and sequencing coverage. As a substantial part of the reviewed studies do not state the depth of coverage or even amount of DNA in some cases, we call for increased transparency to enable critical assessment of the MRD assays for clinical implementation and feasibility.

The minimal/measurable residual disease (MRD) concept and its use in hematology have been reinforced in the last couple of years as next-generation sequencing (NGS) has become more accessible to clinical laboratories after a continuous increase in sequencing capacity and lowered costs. The introduction of NGS has provided improved resolution of the nucleotide composition of the gene rearrangements in both B and T cells, making it highly effective for longitudinal surveillance of a single but also multiple coexisting clones. Much work has already focused on superseding the high sensitivity provided by quantitative polymerase chain reaction (qPCR) while increasing assay specificity and circumventing the need for patient-specific PCR primers/probes. Because clonal gene rearrangements provide specific signatures of both B-cell receptor (BCR) immunoglobulins and T-cell receptor (TCR) chains, these target sequences are of immense importance in the diagnostic workup, and sometimes prognostically, because they can serve as a target for MRD quantification in the setting of malignant lymphoproliferative disorders. A frequent notion on assay sensitivity relates to the capability of detecting residual malignant cells among normal counterparts down to one malignant cell out of a million healthy cells and potentially beyond.

Constituting a key component of the adaptive immune system, the B lymphocytes are highly specialized in antigen presentation, BCR antigen binding, and

Offprint requests to: Marcus Høy Hansen, Department of Hematology, Odense University Hospital, J.B. Winsløws Vej 15, 3rd floor, Odense 5000, Denmark; E-mail: marcus.hoy.hansen@rsyd.dk

https://doi.org/10.1016/j.exphem.2021.03.005

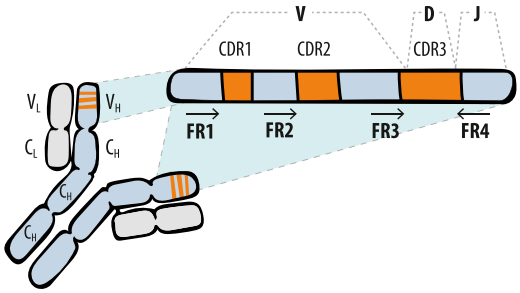

Figure 1. The complementarity-determining regions (CDR1–3) of the immunoglobulin variable domain (VH or VL) provide the diversity for antigen binding. The hypervariability directly stems from the DNA rearrangement of the variable (V), diversity (D), and joining (J) genes in the heavy chain locus and the V and J genes in the immunoglobulin light chain (not shown here), whereas the constant regions of the immunoglobulin (CL or CH) do not provide specificity toward an antigen. To amplify parts of the rearranged locus, the PCR primers (→ or ←) target conserved framework regions in the leader sequence present in the beginning of VH (not shown) or in the VH genes (FR1–3) and JH genes (FR4).

mounting a humoral response by the secretion of antibodies. The flexibility and extensive diversity of the immunoglobulins are generated primarily by gene rearrangements of the light and heavy chain loci. The profound implication is that the genomes of differentiating lymphocytes are not static in these regions, introducing irreversible genomic changes. It is this highly specific signature of the immunoglobulins that is exploited in diagnostics and MRD laboratory assays (Figure 1). After the naive B cell encounters an antigen, it may undergo additional somatic hypermutations. Although clonal proliferation also represents a normal counteraction inherent to the immune response, acute or chronic excessive expansion can be the result of a malignant clonal transformation, with a distinct clonotype constituting the lymphoproliferative disorder. The motivation for performing NGS analyses of clonal gene rearrangements is both to identify the presence of clonal lymphoproliferation by a specific unique gene rearrangement at diagnosis and to detect a minute amount of residual disease of this clonotype during clinical remission or as an early sign of relapse at follow-up.

In this perspective, we critically approach the topic of sensitivity in the setting of MRD detection by sequencing of clonal immunoglobulin heavy chain (IGH) gene rearrangements (Figure 1) in lymphoid neoplasms. We evaluate how sensitive measurable disease detection is being performed by ultrahigh sequencing coverage, potential technical problems to consider, and biological pitfalls to establish a more nuanced discourse of the sensitivities. Foremost, we address the theoretical amount of DNA and sequencing reads needed to confidently detect MRD at a given sensitivity level and factors that may influence the resolution. Although the topics are narrowed to IGH gene rearrangements, the general discussion also pertains to light chain and T-cell receptor gene rearrangements. For general reviews on the topic of MRD detection by means of clonal gene rearrangements, we instead refer to previously published general overviews [1–6].

Detection of residual lymphoproliferative disease

It has long been recognized that MRD during remission is of prognostic value, hence there is a strong incentive for pushing the limit of detection further to detect even smaller amounts of clonal cells than possible with current qPCR techniques. The topic has been explored extensively throughout the last two decades, most recently by NGS owing to its scalability, specificity, sensitivity, and, equally important, the avoidance of patient-specific assays. The undetectability of MRD after intervention in acute lymphoblastic leukemia (ALL) provides an improved measurement of favorable prognosis when evaluated by NGS in comparison to flow cytometry [7,8]. Although ALL, and lately multiple myeloma (MM), has received much of the attention, the use is widening to other malignancies, such as chronic lymphocytic leukemia (CLL) [9]. Thompson et al. [9] reported that the majority of patients found MRD negative by flow cytometry had measurable disease when evaluated with NGS down to the 10^{-6} level, but they also noted that the recent developments in flow cytometry may provide a resolution comparable to that of NGS. Patients with undetectable MRD using NGS were found to have superior progression-free survival [9]. Moreover, MRD was evaluated to be a determining factor in the risk of relapse after stem cell transplantation [10], using TCR and immunoglobulin gene rearrangements as MRD targets using PCR in patients with relapsed ALL. Patients with less than 1 in 10,000 malignant cells had a higher probability of event-free survival and a lower cumulative incidence of relapse. One interesting finding of these results related to a "dose-dependent" relation between the level of residual malignant cells and event-free survival (Figure 1).

Before the implementation of NGS, qPCR provided a leap in the detection of clonal gene rearrangement, normally reported within the range of 10^{-4} to 10^{-6} [11], compared with DNA fragment analyses. Before qPCR, Southern blots had been the gold standard. In a thorough introduction to the topic of clonal gene recombinations concerning the standardization of PCR protocols (BIOMED-2), van Dongen et al. [12] assessed the sensitivity of Southern blots to be in the range 5%–10% but cumbersome. That article provided recommendations for PCR that are still highly relevant in the age of NGS. Although both immunoglobulins (IGH, IGK, and IGL) and T-cell receptor genes

(TCRB, TCRD, and TCRG; TCRA was not included in PCR) were addressed in the study, one important fact was that not all expanded clonal gene rearrangements arose from an underlying malignancy. This argument is of course central to the strategy of mapping diagnostic clonotypes before reliable MRD measurements can be achieved, regardless of the assay type. Gene rearrangements are not directly lineage specific, as aptly exemplified by immature acute B- and T-cell leukemias [13–15], and not all lymphocytic neoplasms necessarily have a detectable rearrangement with the applied technologies. In approximately 80%–90% of the studies referred to here, a gene rearrangement is identified [16–18]. As such, a central part of the NGS MRD assay is to identify the primary clonal gene rearrangement, that is, the clonotype, at diagnosis to implement in follow-up quantification. Failure to detect any diagnostic clone may arise from primer issues, biologically incomplete gene rearrangement [19], or possibly immature malignancies devoid of rearranged immunoglobulin.

Although the concept of MRD has existed and been used extensively in hematology for several decades, the underlying foundation, methods, and calculations have changed and lack standardization when relying on NGS modalities. Currently, it is assumed that NGS has a clinical capability of detection down to the 10^{-7} sensitivity level [6], although such achievements have not fully been demonstrated in practice. This level of sensitivity may eventually be demonstrated empirically, without the use of extrapolating standard curves from serial dilutions, which is statistically feasible when larger cohorts are analyzed. However, its general applicability remains somewhat speculative, especially because of the increased cost of reaching such a high sensitivity.

Recently, Yao et al. [20] reported concordance between allele-specific oligonucleotide qPCR (ASO-qPCR) and NGS, similar to that described in earlier studies [21], in terms of MRD negativity and positivity detection but also comparable dynamic ranges. The authors provided information on the sequencing depth and the rationale for obtaining 10^{-5} by employing triplicates of 1 μg each [20] and commented critically on articles pushing the sensitivity two orders of magnitude down to the 10^{-7} level. Considering the notion that "sequencing assay sensitivity is limited only by the number of input cells and thus can detect residual disease at levels well below 1 in 1 million leukocytes" [22], one may question how sequencing coverage is being taken into account. The number of reads was in this particular article stated as 10^6, and thus, *Poisson sampling* may have a detrimental effect on the assay. It cannot be assumed that the DNA of the cells will be converted to sequencing reads in a one-to-one manner. However, if the sensitivity is based on a total number of, for example, nucleated cells, then the identification of a clonal sequence among background immunoglobulin gene rearrangements may suffice without reaching the aforementioned coverage threshold.

A proper sensitivity is not just detecting one in a million cells

On one hand, it may important to ask why the sensitivity cannot go lower than currently reported, and on the other hand, how such a threshold is feasible in practical terms and how it is evaluated [6]. Although sensitivities of 10^{-6} are reported in the literature (Table 1), *stable* detection of one clonal cell in a million is, empirically, another issue. Neither does sporadic MRD negativity at 10^{-6} correspond to an established assay sensitivity at this level, nor do serial dilution assays, for example, demonstrating linearity down one in a million, provide applicable or robust assays for the clinical laboratory. Another issue relates to the biological aspects. As Thompson et al. [9] briefly note, MRD values below 10^{-6} do not necessarily indicate a cure [9]. It is impossible, logically, to establish a limit at which long-term event-free survival can be guaranteed.

Several criteria must be addressed before achieving high sensitivity. MRD assessment based on detection of somatic single or short nucleotide variants is limited by the error rate of the sequencing platform [36] and, thus, is potentially detrimental to the specificity of the assay. This poses less of a problem in the detection of B- or T-cell gene rearrangements, which rely on a long sequence as the clonal fingerprint. However, it does still relate to the number of base mismatches to accept when including reads in the cumulative clonal frequency. It is known that the effective error rate is not constant across the length of the DNA sequence. Currently, there is no consensus on when to exclude reads from the detected clonotype. As an example, the LymphoTrack assay allows up to two mismatches (Instructions for Use, Nos. 280364 and 280473, Invivoscribe, San Diego, CA, USA).

Role of PCR, stochastic subsampling, and the impact on sensitivities

The process of drawing a liquid or tissue biopsy, through DNA extraction, PCR, and NGS to the final computational output of clonal sequences comprises discrete steps that contribute to the final resolution. Much effort has been exerted into understanding the kinetics of polymerase chain reactions, in which the amplification of DNA in most simple terms can be described as an exponential growth rate, depending on the starting amount, amplification efficiency, and number of thermocycles (Figure 2A). One oversimplification of this model relates to the PCR amplification of immunoglobulin stretches in the diagnostic sample, which involves multiple primers targeting consensus

Table 1. Reported sensitivities across selected studies

Study	Disease	Platform	Amount of material/reads	Estimated sensitivity	Cohort[a]
[23]	ALL	PCR	1 μg, ~6.5×10^5 cells	10^{-4}–10^{-6}	251 samples/168 patients
[10]	ALL	PCR (BIOMED-1)	—	10^{-3}–10^{-4}	91 patients
[22]	B-ALL	NGS (LymphoSIGHT), ASO-qPCR, FC	6.0×10^5 (0.9–17×10^5) cells	10^{-6} (10^{-7})	106 patients
[24]	AML	NGS (*FLT3*/*NPM1*)	7,758 and 15,278 (393–24,997) reads	5×10^{-4}	80 samples
[21]	ALL/MCL/MM	NGS, PCR (Multiplex)	1.5 μg (3×500 ng)	10^{-5}	378 samples/55 patients
[25]	ALL	NGS (LymphoSIGHT)	10 mL PB/5 mL BM (MNCs)	10^{-4}–10^{-6}	237 samples/29 patients
[26]	MM	NGS (LymphoSIGHT)	<300,000 cells	10^{-3}–10^{-5}	133 patients
[27]	B-ALL	FC	>750,000 cells	10^{-4}	2,479 patients
[7]	B-ALL	NGS (ImmunoSeq IgH)	3 μg (sequences)	10^{-7}	41 patients
[28]	MM	FC	—	10^{-4}	700 patients
[29]	MM	FC	~107 cells	10^{-5}–10^{-6}	385 samples/patients
[30]	B-ALL	NGS (BIOMED-2)	500 ng	10^{-4} (-10^{-5})	228 samples/30 patients
[31]	MM	NGS (LymphoSIGHT)	103 (59–288) μg/ 19 (6–36) μg	10^{-6}–10^{-7}	125 patients
[32]	B-ALL	NGS (LymphoTrack)	0.5–5 μg (105 reads)	10^{-6}	122 samples/30 patients
[33]	AML	NGS (*FLT3*-ITD)	0.7 μg, ~100,000 cells	10^{-5}	80 patients
[34]	MM	NGS (LymphoSIGHT)	—	10^{-6}	700 patients in overall study
[8]	B-ALL	NGS (ImmunoSEQ), FC	0.4–8 μg	10^{-6}	619 patients
[20]	MM	NGS (Lymphotrack)	3×1 μg	10^{-5}	4 patients
[9]	CLL	NGS (ClonoSEQ)	N/A (>1.9×10^6 cells?)	10^{-6}	62 patients
[17]	T-ALL	NGS (capture-based panel)	0.6–1 μg	10^{-4}–10^{-5}	23 patients
[35]	MCL	PCR (digital droplet)	—	10^{-4}–10^{-5}	416 samples/166 patients
[16]	ALL	NGS (RNA)	0.4–1 μg	10^{-4}–10^{-5}	258 patients

B-ALL=B-cell acute lymphoblastic leukemia; *FC*=flow cytometry; *MCL*=mantle cell lymphoma; *T-ALL*=T-cell acute lymphoblastic leukemia.
[a]Cohort is defined as estimated number samples and/or estimated number of patients included.

sequences in the leader region and FR1, FR2, FR3, and FR4 regions (Figure 1). Such multiplex reactions with multiple templates may entail both different starting concentrations and varying efficiencies. Thus, the PCR end product can be considered as a sum of both identical, clonal sequences and nonidentical, background sequences of amplified DNA as exemplified in Figure 2A (Eq. 1). Although the ratio of amplified IGH gene rearrangement from the malignant clone to that from the background is expected to be constant before and after PCR amplification (Figure 2A, Eq. 2), one potential consequence of analyzing minute amounts of target DNA from a patient in remission or molecular MRD is that the sample or replicate may contain zero copies. Also, uneven amplification may occur when the clonotype amplification efficiency differs from that of the immunoglobulin background. With each subsequent step that makes up an MRD assay expected to follow a Poisson distribution (Figure 2A, Eq. 3), the overall risk of failing to demonstrate MRD increases dramatically with a small amount of input and subsequent subsampling (Figure 2B). Basically, and most importantly, the Poisson distribution helps to predict the probability of failing to detect MRD given the known mean number of targets. Because this is actually not known, it is fixed as the intended sensitivity level, and then this probability distribution is implemented to theoretically predict the amount of DNA needed. Undoubtedly, the amplification efficiency and error rate will also affect the sensitivity to some degree. In circumstances when the amplification efficiency is lower for the clonal sequence, the chain reaction may be rapidly skewed during the PCR cycles (Figure 2A, Eq. 4).

One practical feature of the clonotype assays pertains to the high specificity; because these assays rely on unique clonal sequences, they are very unlikely to produce false-positive results, in contrast to MRD using single-nucleotide variants, which is more frequently employed in myeloid disorders. Instead, the specificity is comparable to that of *FLT3* internal tandem duplications (ITDs) or other large indels—apart from the fact that the ITD can be difficult to sequence and analyze because of its variable size [33,36,37]. As Thol et al. [24] emphasized, the "amount of DNA used in the PCR amplification and the error rate of the sequencing system are additional factors that may influence MRD sensitivity" [24]. Finally, false-positive results may arise from cross-contamination between samples [38–40], which is a known potential risk in NGS [41]. This issue is relevant in research studies involving longitudinal assessment of patient samples.

Reaching adequate amounts of DNA and sequencing depths

One may expect that DNA from a million cells, crudely equivalent to 6.5 μg, is sufficient to reach a sensitivity

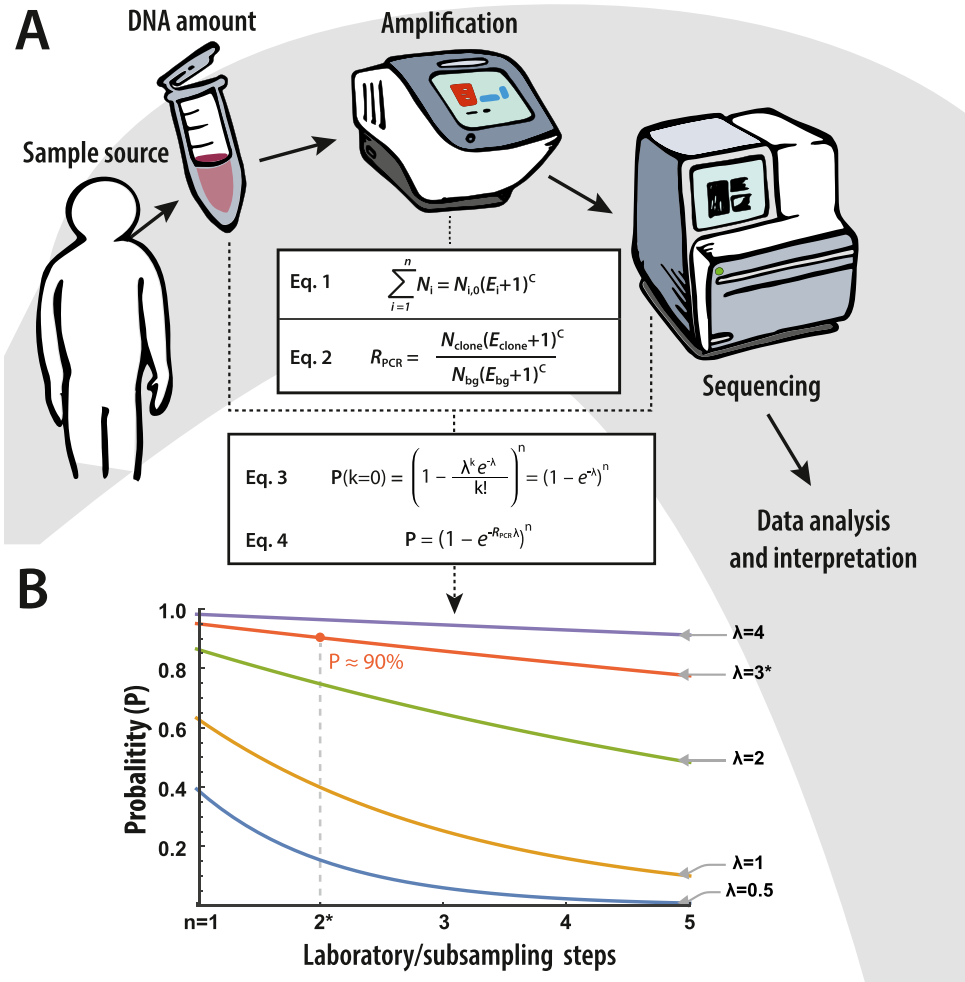

Figure 2. The individual steps of the workflow of measurable disease detection affect the final sensitivity of the assay. In its most simple form, the PCR may be described as a power function, where the DNA molecule is doubled for each cycle (C): 2^C. However, the PCR amplification of gene rearrangements may be considered as a multiplex reaction, each with potentially different sequences and affinities between primers and target and, consequently, chain reaction efficiencies ($E \leq 1$). The PCR product is thus a sum of the individual reactions (Eq. 1, N_i). In some circumstances, the efficiency of the clonotype sequence amplification may be lower than that of the background resulting in a skewed ratio (Eq. 2, $R_{PCR} \neq 1$), which is augmented by the number of cycles (C). The biological sample drawn from the patient and the subsequent laboratory steps **(A)** (n) may be interpreted as a stochastic process of subsampling, each following a Poisson distribution (Eq. 3). It simply describes the probability of observing the specified number of independent events, that is, rearrangements, when considering the amount of cells/DNA relative to the sensitivity level (denoted by λ). Note that the probability of a false-negative observation is here given as $1 - P(k=0)$. If the PCR efficiencies are not comparable, that is, $R_{PCR} \neq 1$, then this may heavily affect each step of the sampling and, hence, the final resolution (Eq. 4). The simplified model shown here explains that at least three times ($\lambda = 3\ \times$) the number of cells/DNA, compared with the sensitivity level, must be analyzed to attain confidence of approximately 90% when DNA sampling and sequencing are considered as 2 × individual steps (95% each) **(B)**. The DNA may be distributed in a single reaction or in replicates. Here, it also suggests that for each successive subsampling in the workflow, the probability (p) of detecting a clone roughly decreases to the power of individual steps if the amount of material or reads if kept constant. Consequently, the amount of material, sequencing depth, and so forth must match the desired sensitivity.

of 10^{-6}. However, calculations reveal that this is not the case, or as van Dongen et al. [2] explain, "the sensitivity of HTS [high-throughput sequencing] is dependent on the number of analyzed cells and the corresponding amount of DNA" [2]. It is concluded that a small amount, well below a million cells, for example, 2–4 μg, will not suffice. Even with several million cells obtained at the starting point, 50% of the total DNA may be lost during purification. A practical option is to operate with a cell equivalent derived from a known quantity of purified DNA, which is then piped directly into the PCR amplification step.

Currently, there are several commercial assays such as LymphoTrack (Invivoscribe) and clonoSEQ (Adaptive Biotechnologies, Seattle, WA, USA). As is exemplified in the clonoSEQ Assay Technical Information (PNL-10027-02), a linear correlation between the empirical and expected levels of MRD can be obtained

at a wide dynamic range. However, a large variance or coefficient of variation in the lowest detection range is also observed, as expected from such assays. Multiple research studies have used serial dilutions to reveal dynamic ranges of detection down to 10^{-6} or 10^{-7} [22,32,42]. As with other biomedical laboratory assays, the dynamic range found by serial dilutions in a controlled setup may not be entirely indicative of the performance in a clinical application. Although common and generally accepted methods, such assays have limitations that may lead to false indications on the level of sensitivity.

The random fluctuation observed in the lower-range measurements from serial dilutions is partly attributed to variance, being inversely proportional to a decreasing number of DNA molecules or cells. As Yao et al. [20] warned, despite detecting residual disease at low frequencies (10^{-5}), the results "were not reproducible among replicates" [20], consistent with the expected variation described by Poisson statistics. Another consequence of this stochastic nature is that if one wants to determine that a certain assay can reach sensitivities down to the 10^{-6}–10^{-7} level, it crudely becomes a matter of running enough samples.

From an analytical stance, one problem is evident in the current body of literature: establishing a sensitivity level set by the amount of input DNA. By providing 103 and 19 μg of DNA to assess MRD from autograft and bone marrow (BM) cells [31], respectively, Takamatsu et al. [31] concluded that NGS provides a one to two orders of magnitude increase in sensitivity compared with ASO-qPCR (10^{-4}–10^{-5}). Because 0.6 μg of DNA was used for the ASO-qPCRs [31], corresponding to approximately 100,000 cells, the results are not directly comparable. Furthermore, sensitivity is not only limited by the amount of DNA used, as we will argue. Converting the reads to MRD is another matter. Yao et al. [20] provided useful guidelines in the attempt to standardize the NGS MRD assay using LymphoTrack and MiSeq, but they noted that there is currently no consensus on the implementation of spike-in controls in NGS [20]. In comparison, the clonoSEQ assay additionally implements the total number of nucleated cells for internal normalization.

Kotrova et al. [30] argued that although NGS is commonly presented as the most sensitive modality, with the number of reads being the main adjustable factor, it is primarily a matter of sample input amount and PCR efficiency [30]. Presumably, it is all three. When MRD quantification is performed based on NGS, it is important to aim at a sufficient number of reads to obtain the desired sensitivity. Aiming at a number of reads lower than that required by the desired sensitivity [32] or not reporting the number of reads [31] poses a problem when evaluating the assay. Reviewing NGS, Takamatsu argued that the PCR products are sequenced at least 10^6 times using high-throughput NGS to achieve MRD detection of one in a million [43]. Contrary to the reasoning in an earlier article on monitoring *FLT3*-ITD and *NPM1* mutations [24], 10,000 sequencing reads do not equal a sensitivity of 10^{-4}, nor do a million reads confidently reveal one in a million cells (see also [22,24,44]). One of the strategies has been presented as "Typically, the number of reads exceeded the number of starting molecules, allowing every starting molecule to be sampled" [22], while being a statistical problem. As follows, a million reads are not sufficient in most cases to reach a 10^{-6} detection level, nor to sample every starting molecule. If each clonal gene rearrangement is successfully converted to sequencing reads in an assumed one-to-one manner, theoretically, it follows from the Poisson distribution that to confidently detect a single or more clonal reads in 95% of the cases (Figure 2B), three times higher coverage of the region and approximately 20 μg of DNA must be used. This amount is three times as much as may be expected if the same initial reasoning is used for DNA requirements. As Rustad and Boyle phrased it, calling the phenomenon the "rule of three," reaching 1 in 10 million is more than a technical challenge [45] and requires a minimum of 30 million cells. One can easily envision an exhaust of biological material, while reaching a low 10^{-7} detection level. As exemplified statistically, as crude as it may be, this corresponds to just around 200 μg of DNA, when estimating the amount of DNA in each cell to be approximately 6.5 pg. A quarter more is needed to lower the theoretical false-negative rate from 5% to 2%, that is, corresponding to four times the effective sensitivity level (Figure 2B). On the basis of these calculations, even the impressive assessment of MRD in autografts using 103 μg (59–288 μg) of DNA [31] may fall short of the material needed to reach the desired threshold with sufficient confidence. It is realized that the "rule of three," derived from the Poisson distribution, may be insufficient in practice [45] because one would expect a loss of material, carrying DNA from several laboratory steps or if the conclusion requires more than a single observation to confirm MRD clinically. This theoretical basis also justifies the use of replicates when requirements are not met for each individual analysis. It has previously been assessed that replicates mitigate experimental errors and increase the reliability and statistical power of the NGS assay [46,47] in comparison to the increased depth of coverage. Although several factors influence this effect according to whether a technical or biological replicate is implemented, it is trivial to see that the effective probability of an assay failure decreases with the number of replicates. In its most simple form, this

can be described as the combined probability p(false negative)^n, where n is the number of independent replicates. Thus, even relatively high error rates can be reduced effectively, such as widely used triplicates in qPCR.

Type of material used in MRD measurements

Another pertinent question relates to the specific type of material on which to base MRD measurements: are amplified immunoglobulin sequences based on bulk or mononuclear cells (MNCs) from bone marrow cells comparable to those from purified B lymphocytes? Currently, no studies consistently indicate that isolation of B cells using a cell sorter, or a magnetic separation system, is required to reach a high level of sensitivity or that it outperforms analyses on MNCs, and most current studies using sequencing of clonal gene rearrangements as an MRD tool do not rely on cell sorting. Also relevant is whether to base MRD measurements on peripheral blood, BM, circulating cell-free tumor DNA (ctDNA), and so forth. Selecting appropriately is emphasized by reports of differential levels of detection, such as BM versus peripheral blood in CLL. Many lymphoma subtypes are aleukemic and often have no bone marrow involvement. In this setting, ctDNA MRD detection is eagerly warranted, and presumable sensitivity issues become even more challenging. Recently, detection of circulating DNA from multiple myeloma cells has gained interest because of the ease of sampling compared with bone marrow biopsy. Although the detection has been backed by several reports [48–52], it has also been concluded that there is generally no correlation between circulating tumor DNA and BM MRD levels [53]. In contrast, a direct correlation between clonal concentration in BM and peripheral MNCs, evaluated by sequencing, has also been reported [54].

From a clinical perspective, other problems may arise. A patient brought into deep remission potentially experiences severe therapeutic lymphocyte depletion. In such instances, the sorting of B cells may only provide a small amount of DNA. Whereas bulk or MNC DNA from BM, peripheral blood, or other sources may provide enough material to reach the desired sensitivity, it is methodologically different from sorted cells, and the results cannot be directly compared. Although cell sorting may potentially increase the sensitivity, it may lead to a loss of material and present additional sample requirements, such as in cryopreservation. Thus, choosing a strategy or assay suited for a specific purpose is imperative. To provide an example, one assay can be based on calculating the fraction of clonal sequences relative to total nucleated cells, as mentioned previously; another may instead define MRD as clonal cells per million or rely on a spike-in control equivalent to a fixed number of cells. Unarguably, the different approaches will also perform differently under MRD circumstances in which the patient may be depleted of B cells.

Another ongoing discussion relates to the suitability of formalin-fixed paraffin-embedded (FFPE) tissue specimens. Even though thorough studies have found that FFPE of shorter amplicon length can be used diagnostically to establish an MRD marker [42,55], it cannot currently be recommended for the application of sensitive clinical MRD measurement because of the requirements regarding concentration, DNA integrity, and number of usable reads. As determined by Arcila et al. [39], FFPE samples were generally of lower coverage and had a higher failure rate. Finally, as an alternative to DNA, RNA is another choice of nucleic acid, which is not explored in depth here. One potential advantage of RNA is that the clonal sequence may be highly expressed, as explored recently in a cohort of 258 diagnostic childhood B-cell ALL cases [16]. Another study found that although the leukemic clone could be identified in both DNA and RNA, the transcriptional clonal frequencies were generally lower [56]. Caution is warranted because not all rearrangements are transcribed, as reported by several studies [57–59]. As RNA sequencing has often been performed using shorter reads in its more general application, the robustness becomes apparent from the results presented by Blachly et al. [60], who reported striking concordance between Sanger sequencing and NGS using 91 and 50 nucleotide paired-end reads. In that study, no targeted enrichment was performed and only a minor fraction of the reads mapped to the IGH region of the genome.

Standardization of NGS MRD

It has become clear that standardization is important to the future clinical applicability of MRD detection employing NGS. Such work is currently being performed by the EuroClonality–NGS consortium, which, in the recently published standardizations of IGH and TCR NGS, has presented protocols and guidelines for MRD marker identification and, not directly, the detection of MRD [61]. However, it seems that the pieces are currently being put together at a rapid pace; Kotrova et al. [62] presented a protocol on behalf of the EuroClonality–NGS Working Group, which may help to increase assay reproducibility and leverage the NGS MRD field. The consortium also provides software for the analysis and interpretation of large amounts of immunogenetic data [63]. Although many researchers may implement commercial assays, such as LymphoTrack and clonoSEQ, there is a need for standardized general guidelines for the implementation of NGS MRD targeting clonal gene rearrangements.

Currently, this still leaves several issues regarding "how low you can go," by practical means, largely unsettled. Standardizing all aspects of such assays, including guidelines for reaching a stable sensitivity, is a huge undertaking still raising many questions, both technically and biologically, and will expectedly continue for years.

Conclusions

Much work has been focused on designing and standardizing primers, but the guidelines involving type and sufficient amount of input material, adequate number of on-target sequencing reads, sequence similarity threshold for determining clonotypes, cross-contamination, and other factors have not been addressed to the same extent. A critical assessment of the practically feasible sensitivity is often lacking or is approached using serial dilutions to determine a high sensitivity. Although agreeing that NGS provides high sensitivity and specificity, we find that there is a need for a more critical and standardized evaluation of MRD detection. We suggest that guidelines for MRD assessment using NGS must stringently include the depth of coverage. The total number of sequencing reads attained is not a reliable measure as reads may be unspecific targets because of the competitive binding of primers in DNA samples with a small amount of lymphocyte DNA.

Such guidelines, and reports of the results, must at least include the amount of DNA used in the assay, but also considered in subsequent steps such as PCR and library preparation; the actual number of reads for each sequenced sample and the effective depth of coverage for clonal reads; transparent calculations of the achieved MRD sensitivity; and the use of mismatch thresholds or strategy when evaluating sequence deviations from the diagnostic clone. Other points raised pertain to the most optimal biological sample type for a given disorder, potential enrichment strategies such as MNC, and the possibility that such measures may potentially be detrimental for the assay, such as retrieving a small number of cells by cell sorting. Most often, the cellular composition changes during treatment.

Table 2. Selected issues and concerns when detecting measurable disease by targeting immunoglobulins

Subject	Consideration
Type of material used	Is the biological sample suitable in terms of the type of leukemia or lymphoma? Is the quality sufficient for the intended sensitivity? For example, is it derived from mononuclear cells, sorted cells, FFPE, etc.?
Biological differentiation	If the investigated malignancy is very immature, there is a risk that the rearrangement has not taken place.
Using RNA instead of DNA	Potentially nontranscribed rearrangements, decreased stability, and a higher degree of mismatch. However, the number of transcripts may also be substantial.
Amount of material	Does the number of cells and DNA support the desired sensitivity level?
Implementing cell sorting	Does sorting provide improved resolution, lead to loss of material or more complex workflow, or change the interpretation of sensitivity level?
Effects of subsampling	Subsampling cannot be avoided because even drawing a sample or sequencing can be regarded as downsampling or Poisson sampling. The probability of drawing the MRD clone may be described by the Poisson distribution.
Implementation of assay replicates	Replicates may lower the effective assay failure rate and may influence the sensitivity, but may also increase the running costs.
Cross-contamination	Is there a risk of cross-contamination? Is this risk higher in research projects, where longitudinal samples may be included in the same PCR and sequencing chip?
Sensitivity level	What sensitivity level is needed for the assay? Does the amount of DNA and sequencing reads reflect this?
PCR efficiencies	Is the PCR amplification efficiency of the clonotype comparable to that of the background?
Sequencing platform	Which sequencing platform should be chosen? What potential difference in error rate is expected?
Read depth and quality	Does the number of reads comply with the intended sensitivity level? How many reads are unmapped? Is the error rate constant between 5′ and 3′, and does it affect the assay sensitivity?
Specificity	Do the sample quality, nucleotide error rate, algorithm mismatch threshold, amplicon, sequencing read length, etc., support the intended specificity?
Clonotype mismatch	What number of nucleotide mismatches is acceptable and what algorithm should be chosen to define a clonotype?
MRD calculations	How is MRD defined and is it transparent? Does it depend on the total number of cells or a spike-in control? Does it implement a probability distribution, potential skew in PCR efficiencies, etc.?
Background level	At what read threshold does an identified rearrangement define a clonotype in comparison to the background level?
Testing sensitivity by serial dilutions	Does the serial dilution assay adequately describe the actual sensitivity of the assay?

Also, in research settings, when several longitudinally acquired samples may be analyzed in the same batch, the risk of cross-contamination or sample carryover must be evaluated.

Because the scope of this perspective is limited to raise awareness of some of the pitfalls when performing detection of residual malignant lymphocytes with rearranged immunoglobulins by NGS, it is not possible to provide an exhaustive review of the topics. In summary, we have touched on some of the considerations provided in Table 2. Importantly, what the investigator must realize is that from the point when a sample is drawn from the patient, the sensitivity of the method is affected in all subsequent laboratory steps toward the final processed sequencing output. In an oversimplified example for the sake of clarity, one may visualize the extraction of approximately 6.5 μg of DNA, that is, 1 million cells. The MRD level is 10^{-6} so if the probability of drawing a sample of one or more malignant cells is theoretically 63%, then this may be reduced further to 40% when sequencing to a depth of coverage of 1 million (see Figure 2B). Sampling 3 million cells and obtaining three times the depth theoretically averages this figure to 90%; however, many other factors, such as quality, source of material, and PCR kinetics, influence the assay. Van Dongen et al. [2] and other researchers mentioned here have already touched on the issue of lack of standardization, but discussion of the topic is still required several years after the introduction of the NGS/HTS MRD assays. We advocate that transparent information, such as depth of coverage and the MRD calculation method, is mandatory when reporting the sensitivity of the assay.

References

1. Monter A, Nomdedeu JF. ClonoSEQ assay for the detection of lymphoid malignancies. Expert Rev Mol Diagn. 2019;19:571–578.
2. van Dongen JJ, van der Velden VH, Bruggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. Blood. 2015;125:3996–4009.
3. Sanchez R, Ayala R, Martinez-Lopez J. Minimal residual disease monitoring with next-generation sequencing methodologies in hematological malignancies. Int J Mol Sci. 2019;20:2832.
4. Kotrova M, Trka J, Kneba M, Bruggemann M. Is next-generation sequencing the way to go for residual disease monitoring in acute lymphoblastic leukemia? Mol Diagn Ther. 2017;21:481–492.
5. Langerak AW, Bruggemann M, Davi F, et al. High-throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. J Immunol. 2017;198:3765–3774.
6. Kruse A, Abdel-Azim N, Kim HN, et al. Minimal residual disease detection in acute lymphoblastic leukemia. Int J Mol Sci. 2020;21:1054.
7. Pulsipher MA, Carlson C, Langholz B, et al. IgH-V(D)J NGS-MRD measurement pre- and early post-allotransplant defines very low- and very high-risk ALL patients. Blood. 2015;125:3501–3508.
8. Wood B, Wu D, Crossley B, et al. Measurable residual disease detection by high-throughput sequencing improves risk stratification for pediatric B-ALL. Blood. 2018;131:1350–1359.
9. Thompson PA, Srivastava J, Peterson C, et al. Minimal residual disease undetectable by next-generation sequencing predicts improved outcome in CLL after chemoimmunotherapy. Blood. 2019;134:1951–1959.
10. Bader P, Kreyenberg H, Henze GH, et al. Prognostic value of minimal residual disease quantification before allogeneic stem-cell transplantation in relapsed childhood acute lymphoblastic leukemia: the ALL-REZ BFM Study Group. J Clin Oncol. 2009;27:377–384.
11. Leisch M, Jansko B, Zaborsky N, Greil R, Pleyer L. Next generation sequencing in AML—on the way to becoming a new standard for treatment initiation and/or modulation? Cancers (Basel). 2019;11:252.
12. van Dongen JJ, Langerak AW, Bruggemann M, 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;17:2257–2317.
13. Beishuizen A, Verhoeven MA, van Wering ER, Hahlen K, Hooijkaas H, van Dongen JJ. Analysis of Ig and T-cell receptor genes in 40 childhood acute lymphoblastic leukemias at diagnosis and subsequent relapse: implications for the detection of minimal residual disease by polymerase chain reaction analysis. Blood. 1994;83:2238–2247.
14. Meleshko AN, Belevtsev MV, Savitskaja TV, Potapnev MP. The incidence of T-cell receptor gene rearrangements in childhood B-lineage acute lymphoblastic leukemia is related to immunophenotype and fusion oncogene expression. Leuk Res. 2006;30:795–800.
15. Nosaka T, Kita K, Miwa H, et al. Cross-lineage gene rearrangements in human leukemic B-precursor cells occur frequently with V-DJ rearrangements of IgH genes. Blood. 1989;74:361–368.
16. Li Z, Jiang N, Lim EH, et al. Identifying IGH disease clones for MRD monitoring in childhood B-cell acute lymphoblastic leukemia using RNA-Seq. Leukemia. 2020;34:2418–2429.
17. Cavagna R, Guinea Montalvo ML, Tosi M, 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 (Basel). 2020;12:1505.
18. Gawad C, Pepin F, Carlton VE, et al. Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. Blood. 2012;120:4407–4417.
19. Gonzalez D, Gonzalez M, Alonso ME, et al. Incomplete DJH rearrangements as a novel tumor target for minimal residual disease quantitation in multiple myeloma using real-time PCR. Leukemia. 2003;17:1051–1057.
20. Yao Q, Bai Y, Orfao A, Chim CS. Standardized minimal residual disease detection by next-generation sequencing in multiple myeloma. Front Oncol. 2019;9:449.
21. Ladetto M, Bruggemann M, Monitillo L, et al. Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. Leukemia. 2014;28:1299–1307.
22. Faham M, Zheng J, Moorhead M, et al. Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. Blood. 2012;120:5173–5180.
23. van Dongen JJ, Seriu T, Panzer-Grumayer ER, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. Lancet. 1998;352:1731–1738.
24. Thol F, Kolking B, Damm F, et al. Next-generation sequencing for minimal residual disease monitoring in acute myeloid

leukemia patients with FLT3-ITD or NPM1 mutations. Genes Chromosomes Cancer. 2012;51:689–695.
25. Logan AC, Vashi N, Faham M, et al. Immunoglobulin and T cell receptor gene high-throughput sequencing quantifies minimal residual disease in acute lymphoblastic leukemia and predicts post-transplantation relapse and survival. Biol Blood Marrow Transplant. 2014;20:1307–1313.
26. Martinez-Lopez J, Lahuerta JJ, Pepin F, et al. Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. Blood. 2014;123:3073–3079.
27. Borowitz MJ, Wood BL, Devidas M, et al. Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232. Blood. 2015;126:964–971.
28. Attal M, Lauwers-Cances V, Hulin C, et al. Lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma. N Engl J Med. 2017;376:1311–1320.
29. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia. 2017;31:2094–2103.
30. Kotrova M, van der Velden VHJ, van Dongen JJM, et al. Next-generation sequencing indicates false-positive MRD results and better predicts prognosis after SCT in patients with childhood ALL. Bone Marrow Transplant. 2017;52:962–968.
31. Takamatsu H, Takezako N, Zheng J, et al. Prognostic value of sequencing-based minimal residual disease detection in patients with multiple myeloma who underwent autologous stem-cell transplantation. Ann Oncol. 2017;28:2503–2510.
32. Cheng S, Inghirami G, Cheng S, Tam W. Simple deep sequencing-based post-remission MRD surveillance predicts clinical relapse in B-ALL. J Hematol Oncol. 2018;11:105.
33. Levis MJ, Perl AE, Altman JK, et al. A next-generation sequencing-based assay for minimal residual disease assessment in AML patients with FLT3-ITD mutations. Blood Adv. 2018;2:825–831.
34. Perrot A, Lauwers-Cances V, Corre J, et al. Minimal residual disease negativity using deep sequencing is a major prognostic factor in multiple myeloma. Blood. 2018;132:2456–2464.
35. Drandi D, Alcantara M, Benmaad I, et al. Droplet Digital PCR Quantification of mantle cell lymphoma follow-up samples from four prospective trials of the European MCL Network. Hemasphere. 2020;4:e347.
36. Levine RL, Valk PJM. Next-generation sequencing in the diagnosis and minimal residual disease assessment of acute myeloid leukemia. Haematologica. 2019;104:868–871.
37. Schranz K, Hubmann M, Harin E, et al. Clonal heterogeneity of FLT3-ITD detected by high-throughput amplicon sequencing correlates with adverse prognosis in acute myeloid leukemia. Oncotarget. 2018;9:30128–30145.
38. Rawstron AC, Fazi C, Agathangelidis A, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. Leukemia. 2016;30:929–936.
39. Arcila ME, Yu W, Syed M, 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;21:330–342.
40. Bartram J, Mountjoy E, Brooks T, et al. Accurate sample assignment in a multiplexed, ultrasensitive, high-throughput sequencing assay for minimal residual disease. J Mol Diagn. 2016;18:494–506.
41. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011;27:2601–2602.
42. Scheijen B, Meijers RWJ, Rijntjes J, et al. Next-generation sequencing of immunoglobulin gene rearrangements for clonality assessment: a technical feasibility study by EuroClonality-NGS. Leukemia. 2019;33:2227–2240.
43. Takamatsu H. Clinical value of measurable residual disease testing for multiple myeloma and implementation in Japan. Int J Hematol. 2020;111:519–529.
44. Ayala R, Onecha E. Next generation sequencing as the new gold standard for minimal residual disease detection in B-ALL. J Lab Precision Med. 2018;3. 97–97.
45. Rustad EH, Boyle EM. Monitoring minimal residual disease in the bone marrow using next generation sequencing. Best Pract Res Clin Haematol. 2020;33:101149.
46. Ziller MJ, Hansen KD, Meissner A, Aryee MJ. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods. 2015;12:230–232.
47. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. Nat Rev Genet. 2014;15:56–62.
48. Kis O, Kaedbey R, Chow S, et al. Circulating tumour DNA sequence analysis as an alternative to multiple myeloma bone marrow aspirates. Nat Commun. 2017;8:15086.
49. Mishima Y, Paiva B, Shi J, et al. The mutational landscape of circulating tumor cells in multiple myeloma. Cell Rep. 2017;19:218–224.
50. Manzoni M, Pompa A, Fabris S, et al. Limits and applications of genomic analysis of circulating tumor DNA as a liquid biopsy in asymptomatic forms of multiple myeloma. Hemasphere. 2020;4:e402.
51. Gerber B, Manzoni M, Spina V, et al. Circulating tumor DNA as a liquid biopsy in plasma cell dyscrasias. Haematologica. 2018;103:e245–e248.
52. Rustad EH, Coward E, Skytoen ER, et al. Monitoring multiple myeloma by quantification of recurrent mutations in serum. Haematologica. 2017;102:1266–1272.
53. Mazzotti C, Buisson L, Maheo S, et al. Myeloma MRD by deep sequencing from circulating tumor DNA does not correlate with results obtained in the bone marrow. Blood Adv. 2018;2:2811–2813.
54. Vij R, Mazumder A, Klinger M, et al. Deep sequencing reveals myeloma cells in peripheral blood in majority of multiple myeloma patients. Clin Lymphoma Myeloma Leuk. 2014;14. 131–139 e131.
55. Langerak AW, Groenen PJ, Bruggemann M, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. Leukemia. 2012;26:2159–2171.
56. Wu J, Jia S, Wang C, et al. Minimal residual disease detection and evolved IGH clones analysis in acute B lymphoblastic leukemia using IGH deep sequencing. Front Immunol. 2016;7:403.
57. Tan KT, Ding LW, Sun QY, et al. Profiling the B/T cell receptor repertoire of lymphocyte derived cell lines. BMC Cancer. 2018;18:940.
58. Plevova K, Francova HS, Burckova K, et al. Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. Haematologica. 2014;99:329–338.
59. Lombardo KA, Coffey DG, Morales AJ, et al. High-throughput sequencing reveals novel features of immunoglobulin gene rearrangements in Burkitt lymphoma. Blood Adv. 2017;1:1261–1262.

60. Blachly JS, Ruppert AS, Zhao W, et al. Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 2015;112:4322–4327.
61. Knecht H, Reigl T, Kotrova M, et al. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. Leukemia. 2019;33:2254–2265.
62. Kotrova M, Darzentas N, Pott C, Bruggemann M, EuroClonality NGSWG. Next-generation sequencing technology to identify minimal residual disease in lymphoid malignancies. Methods Mol Biol. 2021;2185:95–111.
63. Bystry V, Reigl T, Krejci A, et al. ARResT/Interrogate: an interactive immunoprofiler for IG/TR NGS data. Bioinformatics. 2017;33:435–437.