(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 September 2020 (24.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/190138 A1

(51) International Patent Classification:
C12Q 1/6881 (2018.01) *C12Q 1/6886* (2018.01)

(21) International Application Number:
PCT/NL2020/050181

(22) International Filing Date:
18 March 2020 (18.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
19163837.8 19 March 2019 (19.03.2019) EP

(71) Applicant: **STICHTING EUROCLONALITY** [NL/NL]; Albinusdreef 2, 2333 ZA Leiden (NL).

(72) Inventors: **BRÜGGEMANN, Monika Ursula Helga**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **KOTROVÁ, Michaela**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **KNECHT, Henrik**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **DARZENTAS, Nikolaos**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **CAZZANIGA, Giovanni**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **DAVI, Frédéric Bernard Louis**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **VAN DONGEN, Jacobus Johannes Maria**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **GARCÍA-SANZ, Ramón**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **GONZALEZ DE CASTRO, David**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **GROENEN, Patricia Johanna Theodora Anneliese**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **HUMMEL, Michael Andreas**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **MACINTYRE-DAVI, Elizabeth Anne**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **STAMATOPOULOS, Konstantinos**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **POTT, Christiane**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **TRKA, Jan**; c/o Albinusdreef 2, 2333 ZA Leiden (NL). **LANGERAK, Anthonie Willem**; c/o Albinusdreef 2, 2333 ZA Leiden (NL).

(74) Agent: **WITMANS, H.A.**; V.O., P.O. Box 87930, 2508 DH Den Haag (NL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2020/190138 A1

(54) **Title:** MEANS AND METHODS FOR ACCURATELY ASSESSING CLONAL IMMUNOGLOBULIN (IG)/T CELL RECEPTOR (TR) GENE REARRANGEMENTS.

(57) **Abstract:** The invention relates to means and methods for assessing clonal immunoglobulin (IG)/T cell receptor (TR) gene rearrangements in a clinical, diagnostic and/or research setting. Provided is a quality control composition comprising a mixture of genomic DNA isolated from a set of nine cultured cell lines, said set comprising the B cell lines ALL/MIK (ALL), Raji (Burkitt lymphoma), REH (B cell precursor ALL), TMM (CML-BC / EBV+B-LCL), TOM-1 (B cell precursor ALL), WSU-NHL (B cell lymphoma) and the T cell lines JB6 (ALCL), Karpas299 (ALCL) and MOLT- 13 (ALL), or wherein one or more cell lines of said set is replaced with one or more other cell line(s) comprising the same IG/TR gene rearrangements. Also provided is a quality control composition consisting of essentially equimolar amounts of genomic DNA isolated from healthy human thymus, healthy human tonsil and healthy human peripheral blood mononuclear cells.

Title: Means and methods for accurately assessing clonal immunoglobulin (IG)/T cell receptor (TR) gene rearrangements.

The invention relates to the fields of immunology, immunogenetics and clinical diagnostics. In particular, it relates to means and methods for assessing clonal immunoglobulin (IG)/T cell receptor (TR) gene rearrangements in a clinical, diagnostic and/or research setting.

Specific antigen recognition by cells of the adaptive immune system (B cells, T cells) is mediated through receptors (immunoglobulin, IG, and T cell receptor, TR) that are uniquely formed during immune development in bone marrow and thymus, respectively. Through recombination of IG/TR loci a diverse (polyclonal) repertoire of unique IG/TR receptors is created. In certain autoimmune diseases this repertoire is skewed (oligoclonal), whereas in lymphoid malignancies receptors are largely identical (monoclonal)[1–7]. IG/TR rearrangements thus form unique genetic biomarkers (molecular signatures) for studying immune cells for clinical, diagnostic and research applications[8–11]. Classically, methods for immunogenetic analysis mostly concern fragment analysis and Sanger-based sequencing. The introduction of next-generation sequencing (NGS) makes deeper analysis of IG/TR rearrangements possible, with impact on the main immunogenetic applications: clonality assessment, minimal residual disease (MRD) detection, repertoire analysis[12–29].

Identification and assessment of clonal IG/TR gene rearrangements is a widely used tool for the diagnosis and follow-up of lymphoid malignancies[30–35]. NGS of IG/TR gene rearrangements is gaining popularity in clinical laboratories, as it avoids laborious design of patient-specific real-time quantitative (RQ)-PCR assays and provides the capability to sequence multiple rearrangements and rearrangement types within a single

sequencing run. Hence, several methods have already been described for high-throughput profiling of IG/TR rearrangements at diagnosis and follow-up in acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL) and other lymphoid malignancies.[16,17,23,24,36,37]

Potential applications for IG/TR NGS are identification of clonal IG/TR markers in diagnostic samples for subsequent analysis of minimal residual disease (MRD), but also the actual MRD analysis itself, in different lymphoid malignancies (mainly ALL, CLL, follicular lymphoma, mantle cell lymphoma, multiple myeloma). In addition, it can be applied for clonality diagnostics in the diagnostic process of lymphoproliferative disorders.

It has been found that NGS assays, especially those based on amplicons, pose major challenges. For example, multiple primers need to anneal under the same reaction conditions, while many technical variables may be introduced by sample preparation, library construction, sequencing and bioinformatics, potentially leading to inaccurate results[38]. Unfortunately, standardization and validation in a scientifically-controlled multicentre setting is still lacking. Particularly in a clinical context, strategies for standardisation of laboratory protocols and quality control (QC) of each component of an NGS assay are highly sought for.

Reference standards are essential for the evaluation of wet-lab and *in silico* NGS processes to ensure the analytical validity of test results prior to implementation of an NGS technology into clinical practice [29,39,40]. Reference DNA materials should be stable sources of rearrangements that can be sequenced and used for measuring qualitative and quantitative properties. However, the present inventors recognized that previously published standards have a limited scope and utility, since they (1) do not cover all relevant IG/TR loci, (2) do not report on the quality of the sequencing run or the performance of samples and primers, and/or (3) are synthetic constructs that may not reflect the complexity of native genomic DNA[23,41,42].

Therefore, they aimed at providing improved types of quality controls that can be readily integrated in existing systems for immunoprofiling IG/TR sequence data, in particular ARResT/Interrogate[43], an interactive web-based computational platform that can process and annotate large amounts of immunogenetic data, calculate several relevant statistics including on QC, and present results in the form of multiple interconnected visualizations. More specifically, they sought to provide a composition that can be directly added to a sample to undergo concurrent library preparation and sequencing, acting as in-tube qualitative and quantitative standard that is subjected to the same technical downstream variables as the accompanying samples. Furthermore, they aimed at providing a composition that allows to uniformly assess performance biases or unusual amplification shifts across all types of IG/TR gene rearrangements in a sequencing run by tracking primer usage and comparison with stored reference profiles.

To that end, the present inventors of the EuroClonality-NGS Working Group joined forces to develop, standardize and validate *in vitro* assays and bioinformatics for IG/TR NGS applications. This resulted in the provision of two novel types of quality controls; a central in-tube quality/ quantification control (cIT-QC) based on human B and T cell lines with well-defined IG/TR gene rearrangements, and a central polytarget quality control (cPT-QC) based on a standardised mixture of lymphoid specimens representing a full repertoire of IG/TR genes.

Accordingly, in a first aspect the invention provides a composition (herein also referred to as "central in-tube quality/quantification control" (cIT-QC)) comprising a mixture of genomic DNA isolated from a set of nine cultured cell lines, said set comprising the B cell lines ALL/MIK (B cell precursor ALL), Raji (Burkitt lymphoma), REH (B cell precursor ALL), TMM (CML-BC / EBV+B-LCL), TOM-1 (B cell precursor ALL), WSU-NHL (B cell lymphoma); and three T cell lines: JB6 (ALCL) , Karpas299 (ALCL) and

MOLT-13 (T-ALL), or wherein one or more cell lines of said set is replaced with one or more other cell line(s) comprising the same immunoglobulin (IG)/T cell receptor (TR) gene rearrangements, i.e. the same IG/TR rearrangement profile.

In a second aspect, the invention provides a composition (herein also referred to as "central polytarget quality control" (cPT-QC)) consisting of essentially equimolar amounts of genomic DNA isolated from healthy human thymus, healthy human tonsil and healthy human peripheral blood mononuclear cells.

Compositions according to the present invention are not known or suggested in the art. US2018/208984 A1 relates to a method for detecting IG/TR rearrangements using next-generation sequencing using a set of primers. A set of plasmids comprising known alleles, including TCR sequences of the T cell lines JB6, Karpas299 and MOLT-13, is used as a control. However, the control sample of US2018/208984 using cDNA prevents the inclusion of incomplete TR and IG rearrangements, because they are not transcribed into mRNA molecules. Such incomplete TR and IG rearrangements are explicit targets in the present invention, as they are complementary targets for clonality detection and MRD assessment. For example, unlike a control composition of the invention, US2018/208984 does not allow for the identification and quantification of the rearrangements TRB D-J, TRD V-D, TRD D-D, TRD D-J, IGK-Kde and IGH D-J.

Beccutti *et al.* (BMC Bioinformatics, Vol. 18, no. 1, 2017, pages 1-12) relates to a method for detecting IG/TR rearrangements using NGS. DNA isolated from buffy coat (comprising peripheral blood mononuclear cells) is used as a control. Beccutti *et al.* is silent about the use of DNA from additional tissue sources, let alone that it suggests to include tonsil and thymus genomic DNA in a polytarget quality control composition in order to include essential rearrangements that are not found in PBMCs.

A cIT-QC composition as provided herein has a number of unique and advantageous properties. First, with the selected set of only nine cell lines featuring a total of 46 rearrangements, it represents as few cell lines as possible, while covering each target by at least three different rearrangements, hence allowing for detecting ALL cells harbouring not only lineage-associated but also cross-lineage rearrangements. Second, the rearrangements are unambiguously detectable with Sanger sequencing and/or amplicon-based NGS. Third, the variable region of IGH gene rearrangements are unmutated and therewith avoid issues with primer annealing. Table 1 presents the full list of the 46 rearrangements.

With the use of genomic DNA, a composition of the invention explicitly avoids the usage of plasmids, which are known to pose a serious threat to contaminate PCR assays. Additionally, genomic DNA was chosen to optimally represent the patient samples for which the assay is intended for, and which also comprise genomic DNA.

Genomic DNA is readily isolated from the cell lines using established extraction protocols known in the art. In one embodiment, the DNA is obtained using a phenol-chloroform extraction protocol, followed by ethanol precipitation and elution in Tris ethylenediaminetetra-acetic acid (TE) buffer. The composition is suitably in a dry (e.g. lyophilized) form to be reconstituted with a liquid prior to use.

Table 1: Composition and IG/TR rearrangement profile of cell lines used in the cIT-QC composition.

Lineage	Entity	Cell line	IGH-DJ	IGH-VJ	IGK-VJ-Kde	intron-Kde	TRB-DJ	TRB-VJ	TRD	TRG
T	ALCL	JB6					D1 7/6/4 J2-2	V12-3=V12-4 6/14/4 J2-3		V10 7/12/12 J1=J2; V2 5/13/ J1=J2
T	ALCL	Karpas299					D1 /2/6 J1-6	V20-1 1/22/6 J2-7		V2 /13/4 JP2; V8 /2/5 J1=J2
T	T-ALL	MOLT-13					D1 //6 J1-5; D2 /4/3 J2-3	V10-1 6/18/1 J1-1	V1 1/9/ J1	V3 /8/9 J1=J2; V8 3//3 JP1
B	B cell precursor ALL	ALL/MIK		V3-72 16/24/ J4; V7-4-1 11/40/27 J4	V1D-39=V1-39 6/7/5 J3; V2D-24=V2-24 26/6/20 Kde	intron 4/2/ Kde; intron 4/6/1 Kde			V2 5/21/4 J29	V2 /5/8 JP1; V5 2/3/ JP1
B	B cell precursor ALL	REH		V3-15 1/21/5 J6	V2-29 5/4/ J4; V3D-20=V3-20 4/1/ Kde	intron 5// Kde		V20-1 1/2/26 J2-7	V2 3/22/5 J29; V2 7/3/ D3	V4 10/14/3 J1=J2; V9 1/2/3 J1=J2
B	Burkitt Lymphoma	Raji	D6-13 8/12/6 J1	V3-11=V3-21=V3-48 2/40/3 J4	V1-8 2/2/4 Kde					
B	CML-BC / EBV+B-LCL	TMM	D2-2 3/13/ J3	V1-24 /28/8 J5	V2D-30=V2-30 /7/3 Kde					
B	B cell precursor ALL	TOM-1		V4-55 1/17/10 J6					V2 3/3/2 D3; V2 8/4/ D3	V5 8//18 J1=J2
B	B cell lymphoma	WSU-NHL	D2-2 1/1/8 J4	V6-1 1/22/19 J6	V1D-17=V1-17 1//4 J4	intron 2//3 Kde				

In a preferred aspect, a composition comprises a mixture of about equal amounts of genomic DNA isolated from the selected set of cultured cell lines. For example, the cIT-QC composition is formulated to provide a test sample with the DNA of at least 40 cell copies, preferably at least 50 copies of each cell line. Whereas there is no maximum number of cell copies to be represented in the control sample, very high amounts of genomic DNA may consume to a considerable extent the sequencing power in the assay. In one embodiment, the cIT-QC composition contains about an equal number of cell line DNA copies of the selected set of cultured cell lines and is (formulated to be) reconstituted to a solution that contains the genomic DNA of 20-50 cell copies of each cell line per reaction.

The B and T cell lines for use in a cIT-QC composition provided herein can be obtained from any suitable (commercial) source. For example, the Raji cell line is DSMZ ACC 319, the REH cell line is DSMZ ACC 22, the TMM cell line is DSMZ ACC 95, the TOM-1 cell line is DSMZ ACC 578, the WSU-NHL cell line is DSMZ ACC 58, the Karpas299 cell line is DSMZ ACC 31 and/or the MOLT-13 cell line is DSMZ ACC 436. It is of course also possible to replace one or more of the cell lines shown in Table 1 with another cell line having good growth characteristics that contains (or is provided with) the same rearrangements. Hence, also encompassed is a composition comprising a mixture of genomic DNA isolated from a set of cultured cell lines which together cover the profile with 46 rearrangement types shown in Table 1. In particular, the invention provides a composition comprising a mixture of genomic DNA isolated from a set of nine cultured cell lines, said set comprising the B cell lines ALL/MIK (B cell precursor ALL), Raji (Burkitt lymphoma), REH (B cell precursor ALL), TMM (CML-BC / EBV+B-LCL), TOM-1 (B cell precursor ALL), WSU-NHL (B cell lymphoma) and the T cell lines JB6 (ALCL) , Karpas299 (ALCL) and MOLT-13 (T-ALL), or wherein one or more cell lines of said set is replaced with one or more other

(i.e. distinct from the nine cell lines recited), cell line(s) comprising the same immunoglobulin (IG)/T cell receptor (TR) gene rearrangements. Such an "equivalent" cell type comprises at least the same gene rearrangements depicted in Table 1, or IG/TR rearrangements of the same type, i.e. different CDR3. In a preferred aspect, the composition comprises genomic DNA isolated from the B cell lines ALL/MIK, REH and TOM-1, each comprising cross-lineage TR rearrangements.

In a preferred embodiment, the composition consists of a mixture of, preferably in about equal amounts, genomic DNA isolated from the B cell lines ALL/MIK, Raji, REH, TMM, TOM-1, WSU-NHL and the T cell lines JB6, Karpas299 and MOLT-13.

In a further aspect, the invention provides a cPT-QC composition consisting of essentially equimolar amounts of genomic DNA isolated from healthy human thymus, healthy human tonsil and healthy human peripheral blood mononuclear cells (PB-MNC). In other words, it consists of an equimolar mixture of 1/3 thymus, 1/3 tonsil, 1/3 PB-MNC DNA. As used herein, the term "healthy" refers to tissue obtained from a human subject that is known or presumed not to suffer from an underlying malignant immunological disease or disorder. In one aspect, thymus is obtained from young children through removal due to physical impossibility to reach the heart for surgery. It is preferred that, for each tissue, the genomic DNA is obtained from a number of different human individuals. For example, for each tissue the DNA of 3 to 10 human subjects is used. Since this composition represents a "standardised lymphoid specimen", it is suitably used as separate sample to be processed alongside test samples, it is preferably formulated to provide essentially the same amount of DNA as a regular sample that is tested. This typically ranges from 50 to 200 ng, preferably. In a specific aspect, the composition is dried e.g. lyophilized.

The expression "in about equal amounts" or "in essentially equal amounts" as used herein reflects the aim that each of the cell lines / lymphoid tissue samples is equally represented in the mixture of genomic DNA.

The invention also provides a diagnostic kit comprising a (first) container comprising a "central in-tube quality/quantification control" (cIT-QC) composition of the invention and / or a (second) container comprising a "central polytarget quality control" (cPT-QC) composition as herein disclosed. In one embodiment, the kit comprises at least a cIT-QC composition as herein disclosed. In another embodiment, the kit comprises at least a cPT-QC composition of the present invention. The cPT-QC composition may be packaged together with one or more further useful quality control composition(s). For example, the further control composition may comprise a mixture of genomic DNA isolated from a set of cultured cell lines which together cover the profile with 46 rearrangement types shown in Table 1, such that both quality controls can be used to monitor the assay performance when assessing clonal IG/TR gene rearrangements.

Preferably, the kit comprises both the cIT-QC and the cPT-QC compositions as herein disclosed. The kit may advantageously further comprise one or more reagents for detecting IG/TR gene rearrangements, such as a set of primers for amplicon-based NGS of IG/TR gene rearrangements. In a specific embodiment, the diagnostic kit comprises, in addition to one or both QC composition(s) provided herein, one or more primer sets for detecting one or more of the IG/TR gene rearrangements selected from the group consisting of IGH-VJ, IGH-DJ, IGK-VJ-Kde, TRB-VJ, TRB-DJ, TRD and TRG. Particularly preferred primers, e.g. for use in combination with the QC compositions, are those that have been optimized for NGS-based detection, such as the primers shown in Figure 5 or Table 3.

The invention also provides a set of primers for amplicon-based next-generation sequencing (NGS) of IG/TR gene rearrangements, comprising two or more of the primers selected from the primers shown in Figure 5. Preferably, the set comprises at least two primers for detecting one or more, preferably two or more, more preferably three or more, of the IG/TR gene rearrangements selected from the group consisting of IGH-VJ, IGH-DJ, IGK-VJ-Kde, TRB-VJ, TRB-DJ, TRD and TRG. In a specific aspect, the primer set comprises primers for detecting each of the IGH-VJ, IGH-DJ, IGK-VJ-Kde, TRB-VJ, TRB-DJ, TRD and TRG gene rearrangements. The primer sequences of Figure 5 may be provided with a universal primer sequences (such as M13 forward sequence M13 forward primer (-20): GTAAAACGACGGCCAGT; and/ or T7 universal primer: TAATACGACTCACTATAGGG;) or other universal primer sequences known in the art that do not hybridize to the target sequence, and/or other adaptor or bar code sequences.

In a specific aspect, a primer of the invention comprises a forward or reverse M13 sequence. In one embodiment, a primer sequence of Figure 5 is provided with a universal M13 tail at its 5′ end, preferably with the sequence GTAAAACGACGGCCAGT. In another embodiment, a primer sequence of Figure 5 is provided with a T7 promotor sequence. In a specific aspect, the invention provide a set of primers comprises two or more of the primers selected from the primers shown in Table 3.

The provision of the novel QC compositions of the invention has important implications for the quality control and quantitation strategies. As is demonstrated herein below, the cPT-QC composition is a valuable tool to monitor reproducibility of results and to identify primer perturbations and other deviations in the wet lab protocol, as they introduce detectable changes to the sequencing profile. The addition of the cPT-QC to each sequencing run allows to check the primer and assay performance after

sequencing. Accidental deviations in the concentrations of single primers within the multiplexed IG/TR primer sets can be detected, performance failures of single primers can be traced and consequences for the IG/TR analysis can be estimated by analysis of the cPT-QC data.

Additionally, the advantages and diverse utility of the cIT-QC are shown. In contrast to plasmids or synthetic reference templates, cIT-QC cell lines are particularly well suited to be used as control because they are sources of large quantities of genomic DNA and are commercially available. cIT-QC rearrangements represent 2/3 of the amplifiable rearrangement types over all eight primer sets, and thus offer an opportunity to highlight a number of issues, most obviously over-/under-amplification, but also bioinformatic misidentification. Additionally, cIT-QC rearrangements can replace buffy coat DNA for PCR stability without influencing the patient immune repertoire (since cIT-QC rearrangements are bioinformatically identified and by default excluded from the results).

The cIT-QC enables the conversion from reads to cells, which is of utmost importance for clinical use. Diagnostic material being analysed for MRD marker identification can show abundances of particular clonotypes that do not reflect the clonal composition of the sample. For example, if the diagnostic sample is highly infiltrated by a lymphoid malignancy that does not harbour a targetable rearrangement, the (few) residual lymphoid cells would generate the whole spectrum of detectable rearrangements; in such situations minor accompanying physiological B or T cell clones could be misassigned as clones with leukemic markers.

In addition to its use in marker identification, and as exemplarily shown for B and T cell depletion in aplastic follow-up samples, the cIT-QC is of utmost relevance for MRD quantification in samples on or after treatment, in particular if B or T cell directed therapy, which minimises the background of polyclonal gene rearrangements, was applied. If the relative

tumor burden is calculated by the ratio of leukemia-specific reads to all annotated reads without any normalisation, the quotient reflects the marker frequency only among cells carrying a particular type of rearrangement (e.g. IG rearrangements in the total pool of B cells present) and might thus heavily overestimate the actual tumor load[44].

Still further, the QC protocols can be readily embedded in ARResT/Interrogate, which informs users with reports and messages and allows them e.g. to include the QC-failed samples back into the analysis. The logic behind this is that the flag "fail" is an alarm that pre-defined QC criteria were not met, but it does not necessarily indicate that the data are fully corrupt. However, flagged data should always be used with caution, and dependent on the application or question.

The invention therefore also relates to the use of a composition according to the invention or a kit as described herein above in an assay for detecting IG/TR gene rearrangements. A person skilled in the art will recognize and appreciate the diverse range of applications. Only by way of example, the assay is a clinical diagnostic assay, preferably an assay for detecting clonality, identifying MRD markers and/or MRD monitoring and/or analyzing the (clonal) immune repertoire in a lymphoid malignancy.

A further embodiment relates to an *in vitro* method for detecting IG/TR gene rearrangements in at least one biological sample using NGS, comprising the conventional steps of sample preparation, PCR and/or library construction, sequencing and bioinformatics analysis, but characterized in that at least one of the QC compositions is used. For example, at least one biological sample is spiked with a cIT-QC composition, e.g. in an amount to provide the DNA of at least 40 cell copies of each cell line. Such use as in-tube qualitative and quantitative control enables the conversion from reads to cell correlates, which is of utmost importance for clinical use.

Alternatively or additionally, a cPT-QC composition is run as a separate sample in parallel to the at least one biological "test" sample(s), therewith serving as external control to check the primer and assay performance after sequencing.

Typically, the at least one biological sample is a clinically relevant sample. In one aspect, it is a sample for detection of clonality to support or exclude the diagnosis of malignant lymphoproliferation. In another aspect, it is a sample taken for MRD marker identification or for MRD monitoring analysis or for (clonal) immune repertoire analysis.

A method provided herein can be performed using standard means and protocols known in the art. In one embodiment, at least part of the method is performed using microfluidics technology. For example, the steps of sample preparation, PCR, library construction and/or sequencing is performed in a microfluidics device comprising one or more prestored reagents. Particularly preferred for use in a method of the invention is a centrifugal-microfluidic disk system (also known in the art as "centrifugal microfluidic biochip" or "centrifugal micro-fluidic biodisk") which is a type of lab-on-a-chip technology that can be used to integrate processes such as separating, mixing, reaction and detecting molecules of nano-size in a single piece of platform, including a compact disk or DVD. There are various typical units in a centrifugal microfluidic structure, including valves, volume metering, mixing and flow switching. These types of units can make up structures that can be used in a variety of ways. Before the molecules react with the reagents, they should be prepared for the reactions. The most typical is separation by centrifugal force. In the case of blood, for example, the sedimentation of blood cells from plasma can be achieved by rotating the biodisk for some time. After separation, all molecular diagnostic assays require a step of cell/viral lysis in order to release genomic and proteomic material for downstream processing. Typical lysis methods include chemical

and physical method. The chemical lysis method, which is the simplest way, uses chemical detergents or enzymes to break down membranes. The physical lysis can be achieved by using bead beating system on a disk. Lysis occurs due to collisions and shearing between the beads and the cells and through friction shearing along the lysis chamber walls.

In one aspect, the disk comprises pre-stored reagents for automated and integrated DNA extraction, PCR and/or library generation. See for example the review by Tang *et al*[45].

Exemplary disks for use in a method of the invention include those having one or more of the specific features as disclosed in patent application in the name of Hahn Schickard, such as WO2013/124258, WO2014/198703, WO2015/189280, WO2015/051950 and WO2017/191032.

In a method of the invention, the step of bioinformatic analysis advantageously comprises the use of a web-based, interactive application. For example, bioinformatic analysis comprises the use of a purpose-built bioinformatic application (such as ARResT/Interrogate, or equivalent) for the pre-processing of raw data, primer sequence analysis, immunogenetic annotation, post-processing of results, analysis and use of the cIT-QC (including for marker quantification), analysis and use of the cPT-QC (including for comparison to pre-analyzed stored reference datasets), reporting of / access to / visualization of results.

Herewith, the invention demonstrates the applicability of two reference / QC standards, which allow standardised analysis of IG/TR NGS data (e.g. using the NGS primer sets herein disclosed) with high reproducibility, accuracy and precision in marker identification. With ARResT/Interrogate, a complete *in silico* solution accompanying the *in vitro* assays is provided, which enables an analysis of IG/TR sequences including all quality criteria and quantification concepts necessary for valid marker identification in lymphoid malignancies.

LEGEND TO THE FIGURES

Figure 1. Study design: workflows of development and application for cIT-QC and cPT-QC, and schematic overview of test dataset based on a 96-well plate.

Figure 2. Schematic overview of the SOP for quality control and quantification in marker identification: library preparation, PCR & NGS, bioinformatics with ARResT/Interrogate.

Figure 3. Plots of relationships between cIT-QC and markers in the test dataset. **A.** Relationship between % abundances of reads for cIT-QC and markers (at the x-axis). For cIT-QC, the % denominator is reads with junction; for markers, the % denominator is what we term "usable" reads with junction, which excludes cIT-QC reads; this leads to sums of >100%. **B.** Abundance of markers before and after normalisation to percentage of cells. *normalisation may lead to values >100%.

Figure 4: Schematic overview of the workflow for multicentre validation of IG/TR NGS assays for MRD marker identification in ALL. The IG and TR gene rearrangements are amplified in a two-step approach using multiplex PCR assays. Each of the participating laboratories performed NGS-based IG/TR MRD marker identification in 10 patients with ALL. The central polytarget control (cPT-QC) composition of the invention was used to monitor primer performance, and central in-tube quality/quantification controls (cIT-QC) of the invention were spiked to each sample as library-specific quality control and calibrator. Pipetting was performed in a 96-well format. The data analysis was performed employing ARResT/Interrogate.

Figure 5: Schematic diagrams of rearrangements and primer sets, and histograms showing junctions nucleotide lengths for each investigated locus. 5A-1) Schematic diagrams of IGH-VJ and IGH-DJ rearrangements. The relative position of the VH family primers, DH family primers and

consensus JH primers is given according to their most 5´nucleotide upstream (-) or downstream (+) of the involved RSS.

5A-2) Histograms showing junction nucleotide lengths of complete IGH rearrangements (IGH-VJ tube) in a BCP-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the V-J genes combination.

5A-3) Histograms showing junction nucleotide lengths of incomplete IGH rearrangements (IGH-DJ tube) in a BCP-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the D-J genes combination.

5B-1) Schematic diagrams of IGK-VJ rearrangement and the two types of Kde rearrangements (V-Kde and intronRSS–Kde). The relative position of the VK, JK, Kde, and intronRSS (INTR) primers is given according to their most 5´nucleotide upstream (-) or downstream (+) of the involved RSS.

5B-2) Histograms showing junction nucleotide lengths of IGK-VJ and IGK-V-Kde rearrangements (IGK-VJ-Kde tube) in a B-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the V-J-Kde genes combination.

5B-3) Histograms showing junction nucleotide lengths of intron-Kde rearrangements (intron-Kde tube) in a BCP-ALL patient, cPT-QC, BC, thymus, and tonsil.

5C-1) Schematic diagrams of TRB-VJ rearrangement and DJ rearrangement. The relative position of the TRB V family primers, TRB D primers and the TRB J primers is given according to their most 5´nucleotide upstream (-) or downstream (+) of the involved RSS.

5C-2) Histograms showing junction nucleotide lengths of complete TRB rearrangements (TRB-VJ tube) in a T-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the V-J genes combination.

5C-3) Histograms showing junction nucleotide lengths of incomplete TRB rearrangements (TRB-DJ tube) in a T-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the D-J genes combination.

5D-1) Schematic diagrams of TRG V–J rearrangement and the relative

position of the TRG V and TRG J primers. The relative position of the TRG V primers and the TRG J primers is given according to their most 5´nucleotide upstream (-) or downstream (+) of the involved RSS.

5D-2) Histograms showing junction nucleotide lengths of TRG rearrangements (TRG tube) in a T-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the V-J genes combination.

5E-1) Schematic diagram of VD–JD, DD–JD, DD–DD, and VD–DD, VD-JA29 rearrangements, showing the positioning of VD, JD, DD, and JA29 primers, all combined in a single tube. The relative position of the Vd, Dd, and Jd primers is indicated according to their most 50 nucleotide upstream (-) or downstream (+) of the involved RSS.

5E-2) Histograms showing junction nucleotide lengths of TRD rearrangements (TRD tube) in a T-ALL patient, cPT-QC, BC, thymus, and tonsil. Bars are coloured according to the V-D-J genes combination.

Figure 6: Results of multicentre validation of assays for MRD marker identification in ALL. Left hand columns: Index sequences identified by Sanger sequencing. Right hand columns: Index sequences identified by NGS. Darkest colored sections of the columns reflect clonal sequences identified by both methods, lightest colored sections are sequences identified only by the respective method. Median colored sections are clonal sequences identified by both methods, but by NGS with an abundance of <5% after normalization.

EXPERIMENTAL SECTION

Example 1: Design and production of the central in-tube quality/quantification control (cIT-QC)

Sources and methods

In total, 59 human B (n=30) and T (n=29) lymphoid cell lines were obtained from the American Type Culture Collection (ATCC; www.lgcpromochem-atcc.com, Manassas, VA, USA) and the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ; www.dsmz.de, Braunschweig, Germany), or were derived from internal cell line banks. DNA from cultured cell lines was isolated using a phenol-chloroform extraction protocol, followed by ethanol precipitation and elution in Tris ethylenediaminetetra-acetic acid (TE) buffer. Alternatively, DNA was isolated with the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) according to manufacturer's protocol.

Identification of cell line-specific clonal IG/TR gene rearrangements

Each of the 59 cell lines was screened for clonal IG/TR gene rearrangements using the aforementioned EuroClonality-NGS assay with 100ng of DNA (quantified with Qubit 3.0, Thermo Fisher Scientific) from each cell line, without addition of buffy coat (BC). Paired-end sequencing (2x250bp) was performed on an Illumina MiSeq (Illumina, San Diego, CA, USA) with a final concentration of 7pM per library aiming for at least 2000 reads per sample. To avoid low-complexity library issues 10% PhiX control was added to each sequencing run.

Verification of cell line-specific clonal IG/TR gene rearrangements

Additional methods were used to verify the NGS-amplicon-identified cell line rearrangements:

1. A capture-based protocol, established within EuroClonality-NGS Working Group and covering the coding V, D and J genes of IG/TR loci[37]: in short, cell line DNA was fragmented and processed with the KAPA Hyperplus Kit with Library Amplification (Roche Sequencing Solutions, Pleasanton, CA, USA); hybridisation of libraries was performed with customised SeqCap EZ Choice Probes (Roche Sequencing Solutions, Pleasanton, CA, USA), developed based on Wren et al[37]; 2x150bp paired-end sequencing was performed on Illumina NextSeq.
2. Multiplex amplification and Sanger sequencing according to the BIOMED-2 protocol: PCR products were checked for fragment sizes and clonality in the QIAXCEL Advanced System[11,46]. Clonal PCR products were subjected to heteroduplex analysis and sequenced on either an ABI 3130 or ABI 3500 platform (Applied Biosystems, Foster City, CA, USA).

IG/TR rearrangement profiles of all cell lines, as obtained with the different methods, were compared.

Verification of cell line-specific gene rearrangements from human B and T cell lines via ddPCR

For cases with discrepant results between the three methods, IG/TR allele-specific PCR assays were designed for digital droplet PCR (ddPCR) (QX200TM Droplet DigitalTM PCR System, Bio-Rad) to verify the respective rearrangement. Absolute quantification of IG/TR gene rearrangements by ddPCR was performed using two different gDNA amounts (50ng, 100ng). Each experiment included a polyclonal buffy coat BC control and a no template control.

Allele-specific primers for clonal IG/TR rearrangements and probes for quantification were synthesized by Sigma Aldrich. All primers were cleaned by desalting, while hydrolysis probes containing a 5'-FAM/3'-TAMRA reporter dye were cleaned by HPLC. All oligonucleotides were resuspended

in TE buffer at a total strand concentration Ct = 100 µM and stored at -20°C before use.

ddPCR reactions were prepared in a volume of 20 µL using 10 µL by 2X ddPCR SuperMix (Bio-Rad Laboratories, Hercules, CA), testing two different amounts of cell line gDNA (50ng/500ng) quantified before with the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Waltham, MA), forward primer (FP) and reverse primer (RP), each at a final concentration 300 nmol/L, and FAM-labelled probes (100 nmol/L). Droplets were generated by the QX200 droplet generator (Bio-Rad) using 20 µL of the reaction mixture and 70 µL of the droplet generation oil for probes (Bio-Rad), located onto suitable holes in a DG8 cartridge (Bio-Rad). About 45 µL of the drop-oil mixture (12,000-20,000 drops) were transferred to a 96-well plate (Bio-Rad) and loaded on a DNA Engine Dyad Peltier Thermal Cycler with the following amplification protocol: 95°C for 10 min, followed by 40 cycles: denaturation at 94°C for 30 s; annealing at 60°C for 1 min; extension at 60°C for 1 min. PCR products were loaded into the QX200 droplet reader and analysed by QuantaSoft Version 1.2 (Bio-Rad Laboratories).

Cell line mixture preparation

Initially, quantification of DNA of selected B- and T-cell lines was done by Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Waltham, MA). Quantitative values were checked again by ddPCR-based quantification of the albumin housekeeping gene using 50-200 ng DNA/cell line in order to precisely determine the number of cells per µl of DNA. Primers and probe for albumin quantitation were synthesized by Sigma Aldrich. ddPCR was carried out according to the protocol described above, in duplicates for each cell line. After completion of the PCR, samples were analyzed in the Droplet Reader in terms of number of copies of cell lines per 20µl reaction volume. Based on the values from the ddPCR, the cell line DNA was diluted in TE buffer down to 400 copies/µl. Thereafter, another

ddPCR quantification was performed to check the dilution of each cell line DNA again. Two different volumes of the diluted cell line solution (0.5µl DNA [200 copies] and 2µl DNA [800 copies]) were used as input amount. With suitable quantitative values, cell line DNAs were further diluted and mixed with each other leading to 40 copies of each cell line being present in 2µl of the DNA mixture. This mixture was added to each sample as cIT-QC and subjected to simultaneous library preparation prior to sequencing.

Implementation of the cIT-QC

Bioinformatically, cIT-QC reads are identified using an immunogenetic annotation-based approach that is extremely fast while allowing for variations in sequence, avoiding compute-intensive and potentially inaccurate alignment-based approaches. In ARResT/Interrogate, the term 'spike-ins' is also used to refer to the cIT-QC.

Regarding QC, identification of at least one read per cIT-QC rearrangement and of at least as many total cIT-QC reads as total cells used is required, otherwise the sample is tagged as "QC-failed" (see below for how this is used in ARResT/Interrogate). The quantification factor - calculated by dividing total cIT-QC cells by total reads - is stored and applied in any case, thus still allowing the user to analyse the sample.

Quantification is based on applying the quantification factor to convert the read counts of a clonotype to cell counts, and then calculate the relative abundance against the total input cells.

Example 2: Design and production of the central polytarget quality control (cPT-QC)

Sources and methods

A cPT-QC composition was prepared that consists of genomic DNA isolated from healthy human thymus, healthy human tonsil and healthy human

peripheral blood mononuclear cells (DNA amounts mixed in a ratio 1:1:1). To that end, a (semi-)automated genomic DNA extraction was performed on cell suspensions obtained after dissecting and mincing tissues or Ficoll density blood separation.

The cPT-QC composition is suitable used to undergo NGS library preparation alongside the investigated samples. For the EuroClonality-NGS assay, this involves one cPT-QC sample per run, amplified in eight tubes.

Implementation of the cPT-QC

Primers are bioinformatically identified in the reads of each of the eight cPT-QC tubes of the run and their abundances compared to stored cPT-QC reference results using the test of proportions.

Stored reference results are the output of ARResT/Interrogate from the analysis of a cPT-QC sample. These results should be confirmed through replicate runs over time in each lab to accommodate for technical variability. The results and not the raw data are stored to ensure that the bioinformatic analysis is not compromised inadvertently by the user; this means that the results are updated with every major release of ARResT/Interrogate to ensure compatibility with new runs.

Issues with abundances of particular primers or a specific primer set are used to tag the corresponding cPT-QC samples plus all user samples of the same primer set as "QC-failed".

Replicate runs

As reproducibility is important for a QC of this type, replicate runs of the cPT-QC were performed. Relative abundances of 5' primers were compared employing the test of proportions.

Primer perturbation runs

To assess the usability of the cPT-QC to detect problems with primer performance, artificial perturbations of primer concentrations were created to simulate missing pipetting a primer or pipetting the wrong primer concentration.

First the 5´ primer usage was analysed in a cPT-QC sample and two primers of differing abundances were selected from each primer set, thereby skipping the intron-Kde primer set, which only has two primers; IGH-VJ-FR1-M-1, IGH-V-FR1-O-1; IGH-D-B-1, IGH-D-E-1; IGK-V-G-1, IGK-V-I-1; TRB-V-AD-1, TRB-V-G-1; TRB-D-A-1, TRB-D-B-1; TRG-V-F-1, TRG-V-E-1; TRD-D-A-1, TRD-V-B-1. Those primers were perturbed by fully excluding them from the primer pool and changing their concentration by reduction to 10% and increase to 200%. Relative abundances of 5' primers were compared between these perturbed sets and cPT-QC employing the test of proportions.

Creation of a test dataset

A dataset was created to evaluate and showcase the aforementioned concepts and functionalities, which consists of the following samples:

1. Four diagnostic bone marrow B-/T-ALL samples with high leukemic cell content (leukemic infiltration assessed by routine cytomorphology to be 60-80%).
2. Four samples of patients with B/T cell aplasia after B/T cell targeted treatment. The two samples with B cell aplasia were CLL samples after Rituximab (anti-CD20) treatment and the two samples with T cell aplasia were T-PLL (prolymphocytic leukemia) samples after Alemtuzumab (anti-CD52) treatment. In all these samples lineage-specific aplasia was confirmed by flow cytometry.

3. cPT-QC for all IG/TR primer sets, but with the TRB-VJ primer set results swapped with perturbed results from validation experiments as outlined above. To showcase the generic QC functionalities, <1000 random reads from one of the diagnostic samples were artificially chosen.

Methodology

The diagnostic samples and the cPT-QC were run with all primer sets, while the aplastic follow-up samples were only run with the corresponding primer sets, i.e. the IG sets for samples with B cell aplasia, and the TR sets for samples with T cell aplasia (as depicted in Figure 1; test dataset). Additionally, the follow-up samples were run without addition of buffy coat (BC) to test if the addition of the cIT-QC composition is sufficient to stabilise the samples for sequencing, without compromising their immunogenetic profile. To this end, the protocol visualised in Figure 2 was followed.

Primer performance assessment using the cPT-QC

The tests of proportions of 5' primer relative abundances, applied to the cPT-QC, BC, their replicates, and to the libraries with primer perturbations, showed that there is a clear difference in p-values between sets of un-perturbed and perturbed primers. In other words, the p-values of the differences in abundance of the perturbed primers are noticeably lower. Table 2 presents a simplified view of the results, focusing on the abundances of perturbed primers plus at least one other un-perturbed primer per primer set either to show their normal behavior or discuss their abnormal behavior. Percentage abundances of 5' primers across all primer sets. Top group of primers were perturbed; bottom group is a selection of primers that were left un-perturbed: one per primer set selected alphabetically, plus two examples where the primer behavior is of interest to the discussion (see text). Results are shown: in cPT-QC replicates (third column); against

samples where primers were excluded ("0%", fourth column), reduced to 10% (fifth column), increased to 200% (sixth column). Changes in percentages (indicated separately and as +/-) that led to the test of proportions.

At a p-value threshold of 1e-200, none of the primers are flagged in the cPT-QC, which highlights the reproducibility of the assay, while all the perturbed primers are flagged in the perturbed scenarios. In fact, the lowest p-value in the normal samples is 7.86e-142 for primer TRD-V-A-1 (Table 2), compared to multiple zero values in the perturbed comparisons (with a few exceptions, mainly for the 200% perturbation). Significant changes in abundance were also visible in other cells, with the most likely explanation that those primers were indirectly affected by perturbations of other primers. That is, a primer "taking over" when an initially abundant primer was excluded, such as IGH-V-FR1-D-1 when IGH-VJ-FR1-M-1 is perturbed either way especially since these primers amplify partially overlapping lists of genes.

Evaluation of QC aspects in ARResT/Interrogate

Information on the in silico quality control based on both the cPT-QC and cIT-QC is available in ARResT/Interrogate, with "QC-failed" samples excluded by default to warn and prevent the user from their unintended use. However, the user is notified and has the option to include them back in the analysis.

Generic quality control is also performed on samples, specifically to check for low number of raw reads and low percentage of reads with an identified junction. Such samples are also tagged as "QC-failed".

Table 2: cPT-QC: replicates and primer perturbations.

all numbers % abundances; rep:replicate; test of proportions vs cPT-QC rep1, dark grey:<1e-199, light grey:<1e-99

primers			cPT-QC			cPT-QC vs 0%		cPT-QC vs 10%		cPT-QC vs 200%	
primer set		primer name	rep1	<diff>	rep2	<diff	rep1	<diff	rep1	<diff	rep1
IGH-VJ-FR1	perturbed	IGH-V-FR1-M-1	27.44	-5.2	22.24	-26.71	0.7297	-25.42	2.017	+7.72	35.16
IGH-VJ-FR1	perturbed	IGH-V-FR1-O-1	1.184	-0.089	1.095	-1.121	0.06314	-1.116	0.06792	+1.681	2.865
IGH-DJ	perturbed	IGH-D-B-1:#1:14C	7.318	+0.12	7.438	-7.318	0	-7.271	0.0474	+7.152	14.47
IGH-DJ	perturbed	IGH-D-B-1:#2:14T	11.74	+0.48	12.22	-11.74	0.0006552	-11.65	0.06643	+11.48	23.22
IGH-DJ	perturbed	IGH-D-E-1:#4:14G22G	1.864	-0.099	1.765	-1.859	0.005364	-1.853	0.01096	-0.2	1.664
IGK-VJ-Kde	perturbed	IGK-V-G-1	6.08	+0.169	6.249	-5.954	0.1259	-5.911	0.169	+7.51	13.59
IGK-VJ-Kde	perturbed	IGK-V-I-1	8.848	+0.057	8.905	-8.799	0.0588	-8.496	0.3527	+11.86	20.71
TRB-VJ	perturbed	TRB-V-AD-1	31.76	+1.68	33.64	-31.41	0.3514	-26.86	4.905	+3.93	35.69
TRB-VJ	perturbed	TRB-V-G-1	10.09	-0.515	9.575	-10.06	0.02769	-9.869	0.2006	+1.76	11.85
TRB-DJ	perturbed	TRB-D-A-1	63.2	+0.85	64.15	-63.19	0.01025	-48.89	14.31	+6.53	69.73
TRB-DJ	perturbed	TRB-D-B-1	36.14	-1.36	34.78	-36.06	0.0784	-33.22	2.919	+12.71	48.85
TRD	perturbed	TRD-V-B-1	12.55	+2.33	14.88	-12.49	0.06142	-12.14	0.4108	+30.74	43.29
TRD	perturbed	TRD-D-A-1	64.6	+6.36	70.96	-64.51	0.0914	-62.44	2.162	-7.41	57.19
TRG	perturbed	TRG-V-E-1	3.515	-0.113	3.402	-3.512	0.003061	-3.455	0.05982	+5.547	9.062
TRG	perturbed	TRG-V-F-1	14.48	-0.08	14.4	-14.37	0.1087	-14.45	0.02861	+9.05	23.53
IGH-VJ-FR1		IGH-V-FR1-A-1	15.34	+1.7	17.04	-0.89	14.45	-3.65	11.69	+7.41	22.75
IGH-VJ-FR1		IGH-V-FR1-O-1	16.41	-1.62	14.79	+26.18	42.59	+22.64	39.05	-9.999	6.411
IGH-DJ		IGH-D-A-1:#1:6C	8.291	+1.512	9.803	+1.779	10.07	+1.258	9.549	-0.508	7.783
IGK-VJ-Kde		IGK-V-A-1	9.787	+0.08	9.867	+3.863	13.65	+3.403	13.19	+0.147	9.934
TRB-VJ		TRB-V-AB-1	1.423	+0.054	1.477	+1.48	2.903	+0.517	1.94	-0.069	1.354
TRD		TRD-V-A-1	14.37	-7.114	7.256	+9.44	23.81	+8.12	22.49	-4.508	9.862
TRG		TRG-V-A-1	18.71	+1.7	20.41	+3.06	21.77	+1.9	20.61	-2.63	16.08

Example 3: Marker identification & quantification

Abundances of lymphocyte subpopulations are frequently not available for samples of patients with lymphoid malignancies. Furthermore, as IG/TR NGS only reflects relative representation of the rearrangements, it was important to establish a calibrator, which would allow to normalise sequencing reads to input DNA cells. This is particularly important for tubes that exclusively cover rearrangements being present only in a minority of lymphoid cells (especially the TRD and intron-Kde tubes). TRD genes are not rearranged in normal B cells and are deleted in most TRαß cells. Therefore, oligoclonal TCRγδ T-cells might give rise to dominant clonotypes in the TRD NGS assay, in particular as the normal TCRγδ T cell repertoire is strikingly skewed during childhood. Here the cIT-QC-based abundance correction is of utmost importance to avoid miss-assignment of (minor) clonal TRD rearrangements from minor TCRγδ cell populations as leukemic rearrangements that would then serve as markers in further MRD analysis.

Analysis of the test dataset showed the utility of the cIT-QC in marker identification and quantification. Without the cIT-QC, both diagnostic and aplastic samples seem to be oligoclonal if simply based on the number of reads (Figure 3). However, the very high number of reads from only a very limited number of cIT-QC cells (120-440, dependent on number of cIT-QC rearrangements per primer set), in all aplastic and a few of the diagnostic samples, are an indirect, yet clear indication of the restricted numbers of patient-related input cells harbouring rearrangements of the particular IG/TR locus in those samples. From another perspective, the total read percentage of cIT-QC is much greater than those of rearrangements in these samples, suggesting that also the number of cIT-QC cells is greater than the number of patient-related input cells. Indeed, after quantification with the cIT-QC, cell abundances fall well below the thresholds implying clonality.

On the other hand, and as expected, in the diagnostic samples cIT-QC sequences constitute a minority. Hence, this implies that with the cIT-QC the abundance of a certain rearrangement can much more accurately be determined and recalculated to cell abundances.

Additionally, five experienced EuroMRD ALL reference laboratories performed IG/TR NGS in 50 diagnostic ALL samples, and compared results with those generated through routine IG/TR Sanger sequencing. A cPT-QC composition was used to monitor primer performance, and a cIT-QC composition was spiked into each sample as a library-specific quality control and calibrator. NGS identified 259 (average 5.2/sample, range 0-14) clonal sequences vs. Sanger-sequencing 248 (average 5.0/sample, range 0-14). The overall concordance between Sanger and NGS, including negative libraries, was 78%.

Example 4: Development and multicentre validation of IG/TR NGS assays for MRD marker identification in ALL

This example describes the development and design of an IG/TR assay, including bioinformatics, and its validation for MRD marker identification in acute lymphoblastic leukemia (ALL). Five EuroMRD ALL MRD reference laboratories performed IG/TR NGS in 50 diagnostic ALL samples, and compared results with those generated through routine IG/TR marker screening and Sanger sequencing. A cPT-QC composition was used to monitor primer performance, and a cIT-QC composition was spiked into each sample as a library-specific quality control and calibrator. The overall workflow of the validation study is shown in Figure 4.

MATERIALS AND METHODS

General concept of assay design

With the objective of developing a universal amplicon-based NGS approach for IG/TR sequence analysis at the DNA level, applicable in all lymphoid malignancies, assays for multiple IG/TR loci were designed for: IG heavy (IGH), IG kappa (IGK), TR beta (TRB), TR gamma (TRG), and TR delta (TRD), including complete and incomplete rearrangements whenever applicable. IG lambda (IGL) was excluded due to its limited complementarity to other IG loci and its reduced diversity. TR alpha (TRA) was excluded due to its high complexity, hampering a reasonable multiplex PCR approach at the DNA level.

The IGH locus is rearranged in two steps (Figure 5A). After initial coupling of a single IGH-D gene to an IGH-J gene, an IGH-V gene is joined to the incomplete IGH-DJ rearrangement, resulting in a complete IGH-VJ rearrangement. For amplification of complete IGH rearrangements, primers located in the FR1, FR2 and FR3 regions were designed, but here we only describe the FR1 assay for marker identification in ALL. IGH-DJ rearrangements were amplified in a separate multiplex PCR reaction. The IGK light chain locus is composed of functional IGKV and IGKJ genes, as well as the so-called kappa deleting element (Kde) that can rearrange to IGKV genes, or to a recombination signal sequence (RSS) in the IGKJ-IGKC intron, leading to functional inactivation of the IGK allele (Figure 5B). The IGKV forward primers were designed to be used in combination with IGKJ and Kde reverse primers in one multiplex reaction, whereas a second PCR was developed for the forward intron RSS and reverse Kde primers.

The TRB locus also features a two-step process with initial formation of incomplete TRB-DJ rearrangements followed by complete TRB-VJ rearrangements. Incomplete and complete TRB rearrangements were designed to be detected in two separate multiplex PCR reactions (Figure

5C). As TRG locus rearrangements are one-step VJ recombinations involving a limited number of TRGV and TRGJ genes, a single multiplex assay could be developed (Figure 5D). Finally, in the TRD locus, complete VJ rearrangements are preceded by DD, VD and DJ rearrangements. In addition, certain TRAV genes can rearrange to both TRDJ and TRAJ, whereas TRDV-TRAJ rearrangements, usually involving TRAJ29, can also occur. All of these rearrangements were designed to be amplified in one multiplex PCR assay (Figure 5E). The bioinformatic platform ARResT/Interrogate[43], already developed from the ground-up within the EuroClonality-NGS working group to assist with its multi-faceted activities, was further adapted for this study as described below.

Primer design and technical validation of primer performance

Primers were designed to be gene-specific, but in case of allelic variants, degenerate primers were designed to facilitate multiplexing. For the same reason, single mismatches in the middle or at the 5`-end of the primer were accepted. Table 3 shows the primer sequences comprising nucleotide sequences of Figure 5 and additional adapter sequences (forward or reverse). Those of skill in the art will realize that the Example is only illustrative and that many variations of the specific methods of the Example are possible. For example, there is no need to use the M13 sequences as part of the primers as used in the Example. This could be replaced by any other known sequence of DNA.

Table 3: Primer sequences for the 1st and 2nd step PCR in IG/TR NGS.

1st step PCR tube	Primer nomenclature	µM in PCR	Primer direction	Primer sequence with universal primer sequences attached (forward/reverse) 5' to 3'
TRB V-J	TRB-V-C-1	0.00625µM	5'	GTAAAACGACGGCCAGTTCGCTTCTCACCTGAATGCCC
	TRB-V-A-1	0.0125µM	5'	GTAAAACGACGGCCAGTCTCAGTTGAAAGGCCTGATGGA
	TRB-V-X-1	0.0125µM	5'	GTAAAACGACGGCCAGTGGAAGCATCCCTGATCGATTCT
	TRB-V-AA-1	0.0125µM	5'	GTAAAACGACGGCCAGTTCAGCTAAGTGCCTCCCAAATT
	TRB-V-B-1	0.025µM	5'	GTAAAACGACGGCCAGTAGTTCCAAATCGCTTCTCACCT
	TRB-V-F-1	0.025µM	5'	GTAAAACGACGGCCAGTTTCCCTAATCGATTCTCAGGGC
	TRB-V-J-1	0.025µM	5'	GTAAAACGACGGCCAGTTACAACTGCCAAAGGAGAGGTC
	TRB-V-L-1	0.025µM	5'	GTAAAACGACGGCCAGTTAAAGGAGAAGTCCCGAATGGC
	TRB-V-M-1	0.025µM	5'	GTAAAACGACGGCCAGTGGAGAAGTTCCCAATGGCTACA
	TRB-V-S-1	0.025µM	5'	GTAAAACGACGGCCAGTATAAAGGAGAAGTCCCCGATGG
	TRB-V-W-1	0.025µM	5'	GTAAAACGACGGCCAGTCTCTAGATGATTCGGGGATGCC
	TRB-V-Z-1	0.025µM	5'	GTAAAACGACGGCCAGTTGAAGCAGACACCCCTGATAAC
	TRB-V-AE-1	0.025µM	5'	GTAAAACGACGGCCAGTTGAGCGATTTTTAGCCCAATGC
	TRB-V-AG-1	0.025µM	5'	GTAAAACGACGGCCAGTACAAAGGAGAGATCTCTGATGGA
	TRB-J-A-1	0.025µM	3'	TAATACGACTCACTATAGGGCTACAACTGTGAGTCTGGTGCC
	TRB-J-B-1	0.025µM	3'	TAATACGACTCACTATAGGGCTACAACGGTTAACCTGGTCC
	TRB-J-C-1	0.025µM	3'	TAATACGACTCACTATAGGGTACAACAGTGAGCCAACTTCCC
	TRB-J-D-1	0.025µM	3'	TAATACGACTCACTATAGGGCAAGACAGAGAGCTGGGTTCC
	TRB-J-E-1	0.025µM	3'	TAATACGACTCACTATAGGGCTAGGATGGAGAGTCGAGTCCC
	TRB-J-F-1	0.025µM	3'	TAATACGACTCACTATAGGGCTGTCACAGTGAGCCTGGTC
	TRB-J-G-1	0.025µM	3'	TAATACGACTCACTATAGGGCCTTCTTACCTAGCACGGTGAG
	TRB-J-H-1	0.025µM	3'	TAATACGACTCACTATAGGGTTACCCAGTACGGTCAGCCTAG
	TRB-J-I-1	0.025µM	3'	TAATACGACTCACTATAGGGCTTACCGAGCACTGTCAGCC

TRB-J-J-1	0.025µM	3'	TAATACGACTCACTATAGGGCTTACCCAGCACTGAGAGCC
TRB-J-K-1	0.025µM	3'	TAATACGACTCACTATAGGGTCACCGAGCACCAGGAGCC
TRB-J-N-1	0.025µM	3'	TAATACGACTCACTATAGGGGAATCTCACCTGTGACCGTGAG
TRB-V-D-1	0.05µM	5'	GTAAAACGACGGCCAGTGGAAACTTCCCTGGTCGATTC
TRB-V-N-1	0.05µM	5'	GTAAAACGACGGCCAGTCAACGATCGGTTCTTTGCAGTC
TRB-V-O-1	0.05µM	5'	GTAAAACGACGGCCAGTTAAATCAGGGCTGCTCAGTGAT
TRB-V-P-1	0.05µM	5'	GTAAAACGACGGCCAGTCAGTGATCGGTTCTCTGCAGAG
TRB-V-R-1	0.05µM	5'	GTAAAACGACGGCCAGTCTTGAACGATTCTCCGCACAAC
TRB-V-V-1	0.05µM	5'	GTAAAACGACGGCCAGTCCGAGGATCGATTCTCAGCTAA
TRB-V-AB-1	0.05µM	5'	GTAAAACGACGGCCAGTGCCAAAGGAACGATTTTCTGCT
TRB-V-AI-1	0.05µM	5'	GTAAAACGACGGCCAGTAGGGAGATGTTCCTGAAGGGTA
TRB-V-AJ-1	0.05µM	5'	GTAAAACGACGGCCAGTCCTGAGGGGTACAGTGTCTCTA
TRB-V-AL-1	0.05µM	5'	GTAAAACGACGGCCAGTCAGAATCTCTCAGCCTCCAGAC
TRB-V-E-1	0.1µM	5'	GTAAAACGACGGCCAGTACTTCCCTGATCGATTCTCAGC
TRB-V-H-1	0.1µM	5'	GTAAAACGACGGCCAGTCTCAGGTCACCAGTTCCCTAAC
TRB-V-I-1	0.1µM	5'	GTAAAACGACGGCCAGTCCTAGATTTTCAGGTCGCCAGT
TRB-V-Q-1	0.1µM	5'	GTAAAACGACGGCCAGTCTCAACTAGACAAATCGGGGCT
TRB-V-U-1	0.1µM	5'	GTAAAACGACGGCCAGTATCGATTTTCTGCAGAGAGGCT
TRB-V-Y-1	0.1µM	5'	GTAAAACGACGGCCAGTCGGTATGCCCAACAATCGATTC
TRB-V-AC-1	0.1µM	5'	GTAAAACGACGGCCAGTCTGAAGGGTACAGCGTCTCTC
TRB-V-AH-1	0.1µM	5'	GTAAAACGACGGCCAGTTCCTCTGAGTCAACAGTCTCCA
TRB-V-AK-1	0.1µM	5'	GTAAAACGACGGCCAGTCTGAGGCCACATATGAGAGTGG
TRB-J-L-1	0.1µM	3'	TAATACGACTCACTATAGGGGAAAACTCACCCAGCACGGTC
TRB-J-M-1	0.1µM	3'	TAATACGACTCACTATAGGGTCACCCAGCACGGTCAGCC
TRB-V-G-1	0.15µM	5'	GTAAAACGACGGCCAGTGATTCTCAGGTCTCCAGTTCCC
TRB-V-K-1	0.15µM	5'	GTAAAACGACGGCCAGTTACCACTGGCAAAGGAGAAGTC
TRB-V-T-1	0.15µM	5'	GTAAAACGACGGCCAGTCAAAGGAGAAGTCTCAGATGGC

WO 2020/190138 PCT/NL2020/050181

	TRB-V-AD-1	0.15µM	5'	GTAAAACGACGGCCAGTTTTCTCATCAACCATGCAAGCC
	TRB-V-AF-1	0.15µM	5'	GTAAAACGACGGCCAGTGGAGATGCACAAGAAGCGATTC
TRB D-J	TRB-J-A-1	0.025µM	3'	TAATACGACTCACTATAGGCTACAACTGTGAGTCTGGTGCC
	TRB-J-B-1	0.025µM	3'	TAATACGACTCACTATAGGCTACAACGGTTAACCTGGTCC
	TRB-J-C-1	0.025µM	3'	TAATACGACTCACTATAGGGTACAACAGTGAGCCAACTTCCC
	TRB-J-D-1	0.025µM	3'	TAATACGACTCACTATAGGGCAAGACAGAGAGCTGGGTTCC
	TRB-J-E-1	0.025µM	3'	TAATACGACTCACTATAGGGCTAGGATGGAGAGTCGAGTCCC
	TRB-J-F-1	0.025µM	3'	TAATACGACTCACTATAGGGCTGTCACAGTGAGCCTGGTC
	TRB-J-G-1	0.025µM	3'	TAATACGACTCACTATAGGGCCTTCTTACCTAGCACGGTGAG
	TRB-J-H-1	0.025µM	3'	TAATACGACTCACTATAGGGTTACCCAGTACGGTCAGCCTAG
	TRB-J-I-1	0.025µM	3'	TAATACGACTCACTATAGGGCTTACCGAGCACTGTCAGCC
	TRB-J-J-1	0.025µM	3'	TAATACGACTCACTATAGGGCTTACCCAGCACTGAGAGCC
	TRB-J-K-1	0.025µM	3'	TAATACGACTCACTATAGGGTCACCGAGCACCAGGAGCC
	TRB-J-N-1	0.025µM	3'	TAATACGACTCACTATAGGGGAATCTCACCTGTGACCGTGAG
	TRB-D-A-1	0.1µM	5'	GTAAAACGACGGCCAGTCCTCCACTCCCCTCAAAGGA
	TRB-D-B-1	0.1µM	5'	GTAAAACGACGGCCAGTCAGACTAACCTCTGCCACCTG
	TRB-J-L-1	0.1µM	3'	TAATACGACTCACTATAGGGGAAAACTCACCCAGCACGGTC
	TRB-J-M-1	0.1µM	3'	TAATACGACTCACTATAGGGTCACCCAGCACGGTCAGCC
TRG	TRG-V-E-1	0.05µM	5'	GTAAAACGACGGCCAGTCAAGCATGAGGAGGAGCTGGAAATTG
	TRG-V-F-1	0.05µM	5'	GTAAAACGACGGCCAGTACGTCTACATCCACTCTCACC
	TRG-V-A-1	0.1µM	5'	GTAAAACGACGGCCAGTGCACAAGGAACAACTTGAGATTG
	TRG-V-B-1	0.1µM	5'	GTAAAACGACGGCCAGTTGGAAGCACAAGGAAGAACTTGAGAA
	TRG-V-C-1	0.1µM	5'	GTAAAACGACGGCCAGTGCACAGGGAAGAGCCTTAAATT
	TRG-V-D-1	0.1µM	5'	GTAAAACGACGGCCAGTCAGGAGGTGGAGCTGGATATT
	TRG-V-G-1	0.1µM	5'	GTAAAACGACGGCCAGTCTCTCACTTCAATCCTTACCATCAA
	TRG-V-H-1	0.2µM	5'	GTAAAACGACGGCCAGTGCTCACACTTCCACTTCCACTTTGAAAATAAAGT
	TRG-J-A-1	0.2µM	3'	TAATACGACTCACTATAGGAGTGTTGTTCCACTGCCAAAG

	TRG-J-B-1	0.2µM	3'	TAATACGACTCACTATAGGGGTTCCGGGACCAAATACCTTG
	TRG-J-C-1	0.2µM	3'	TAATACGACTCACTATAGGGGAGCTTAGTCCCTTCAGCAAATA
	TRG-J-D-1	0.2µM	3'	TAATACGACTCACTATAGGGCCTAGTCCCTTTTGCAAACG
TRD	TRD-V-A-1	0.2µM	5'	GTAAAACGACGGCCAGTGAATGCAAAAAGTGGTCGCTATTC
	TRD-V-B-1	0.2µM	5'	GTAAAACGACGGCCAGTTGCAAAGAACCTGGCTGTACT
	TRD-V-C-1	0.2µM	5'	GTAAAACGACGGCCAGTTGCAGATTTTACTCAAGGACGG
	TRD-V-D-1	0.2µM	5'	GTAAAACGACGGCCAGTGCAAAATGCAACAGAAGGTCG
	TRD-V-E-1	0.2µM	5'	GTAAAACGACGGCCAGTGATAAAAATGAAGATGGAAGATTCACTGT
	TRD-V-F-1	0.2µM	5'	GTAAAACGACGGCCAGTCTCCTTCAATAAAAGTGCCAAGC
	TRD-V-G-1	0.2µM	5'	GTAAAACGACGGCCAGTATTGAAAAGAAGTCAGGAAGACTAAGT
	TRD-V-H-1	0.2µM	5'	GTAAAACGACGGCCAGTTCCAGAAAGCAGCCAAATCC
	TRD-D-A-1	0.2µM	5'	GTAAAACGACGGCCAGTAGGGGTATTGTGGATGGCAG
	TRD-J-A-1	0.2µM	3'	TAATACGACTCACTATAGGGTTCCACAGTCACACGGGT
	TRD-J-B-1	0.2µM	3'	TAATACGACTCACTATAGGGGGTTCCACGATGAGTTGTGTT
	TRD-J-C-1	0.2µM	3'	TAATACGACTCACTATAGGGCACGAAGAGTTTGATGCCAGT
	TRD-J-D-1	0.2µM	3'	TAATACGACTCACTATAGGGGTTGTTGTACCTCCAGATAGGTT
	TRD-J-E-1	0.2µM	3'	TAATACGACTCACTATAGGGTGGCTAGAAACACTTACTTGCA
	TRD-D-B-1	0.2µM	3'	TAATACGACTCACTATAGGGCCCAGGGAAATGGCACTTTTG
IGH D-J	IGH-D-A-1	0.2µM	5'	GTAAAACGACGGCCAGTGATTCYGAACAGCCCCGAGTCA
	IGH-D-B-1	0.2µM	5'	GTAAAACGACGGCCAGTGATTTTGTGGGGGYTCGTGTC
	IGH-D-C-1	0.2µM	5'	GTAAAACGACGGCCAGTGTTTGRRGTGAGGTCTGTGTCA
	IGH-D-D-1	0.2µM	5'	GTAAAACGACGGCCAGTGTTTRGRRTGAGGTCTGTGTCACT
	IGH-D-E-1	0.2µM	5'	GTAAAACGACGGCCAGTCTTTTTGTGAAGGSCCCTCCTR
	IGH-D-F-1	0.2µM	5'	GTAAAACGACGGCCAGTGTTATTGTCAGGSGRTGTCAGAC
	IGH-D-G-1	0.2µM	5'	GTAAAACGACGGCCAGTGTTATTGTCAGGGGGTGYCAGRC
	IGH-D-H-1	0.2µM	5'	GTAAAACGACGGCCAGTGTTTCTGAAGSTGTCTGTRTCAC
	IGH-J-A-1	0.4µM	3'	TAATACGACTCACTATAGGGCTTACCTGAGGAGACGGTGACC

IGH V-J	IGH-V-FR1-B-1	0.1µM	5'	GTAAAACGACGGCCAGTGCAGTCTGGAGCAGAGGTGAAAA
	IGH-V-FR1-E-1	0.1µM	5'	GTAAAACGACGGCCAGTGAGGTGCAGCTGTTGGAGTC
	IGH-V-FR1-G-1	0.1µM	5'	GTAAAACGACGGCCAGTTCAGTGGGGCGCAGGACTGTT
	IGH-V-FR1-H-1	0.1µM	5'	GTAAAACGACGGCCAGTTCCAGGACTGGTGAAGCCTCC
	IGH-V-FR1-K-1	0.1µM	5'	GTAAAACGACGGCCAGTCCTCAGTGAAGGTTTCCTGCAAGG
	IGH-V-FR1-L-1	0.1µM	5'	GTAAAACGACGGCCAGTAAACCCACAGAGACCCTCACGCTGAC
	IGH-V-FR1-M-1	0.1µM	5'	GTAAAACGACGGCCAGTCTGGGGGGTCCCTGAGACTCTCCTG
	IGH-V-FR1-N-1	0.1µM	5'	GTAAAACGACGGCCAGTCTTCACAGACCCTGTCCCTCACCTG
	IGH-V-FR1-O-1	0.1µM	5'	GTAAAACGACGGCCAGTTCGCAGACCCTCTCACTCACCTGTG
	IGH-J-A-1	0.1µM	3'	TAATACGACTCACTATAGGGCTTACCTGAGGAGACGGTGACC
	IGH-J-B-1	0.1µM	3'	TAATACGACTCACTATAGGGCTCACCTGAGGAGACGGTGACC
	IGH-V-FR1-A-1	0.2µM	5'	GTAAAACGACGGCCAGTCTGGGGCTGAGGTGAAGAAG
	IGH-V-FR1-C-1	0.2µM	5'	GTAAAACGACGGCCAGTTCACCTTGAAGGAGTCTGGTCC
	IGH-V-FR1-D-1	0.2µM	5'	GTAAAACGACGGCCAGTAGGTGCAGCTGGTGGAGTC
	IGH-V-FR1-F-1	0.2µM	5'	GTAAAACGACGGCCAGTCCAGGACTGGTGAAGCCTTC
	IGH-V-FR1-I-1	0.2µM	5'	GTAAAACGACGGCCAGTGTACAGCTGCAGCAGTCAGG
	IGH-V-FR1-J-1	0.2µM	5'	GTAAAACGACGGCCAGTGCTGGTGCAATCTGGGTCTG
IGK-A	IGK-V-A-1	0.1µM	5'	GTAAAACGACGGCCAGTAAGTGGGGTCCCATCAAGGTTCAG
	IGK-V-B-1	0.1µM	5'	GTAAAACGACGGCCAGTAGTCCCATCTCGGTTCAGTGGCAG
	IGK-V-C-1	0.1µM	5'	GTAAAACGACGGCCAGTGAAACAGGGGTCCCATCAAGGTTC
	IGK-V-D-1	0.1µM	5'	GTAAAACGACGGCCAGTTCCCAGACAGATTCAGTGGCAGTG

	IGK-V-E-1	0.1µM	5'	GTAAAACGACGGCCAGTCTGGAGTGCCAGATAGGTTCAGTG
	IGK-V-F-1	0.1µM	5'	GTAAAACGACGGCCAGTCCCTGGAGTCCCAGACAGGTTCAG
	IGK-V-G-1	0.1µM	5'	GTAAAACGACGGCCAGTGCATCCCAGCCAGGTTCAGTG
	IGK-V-H-1	0.1µM	5'	GTAAAACGACGGCCAGTGTCCCTGACCGATTCAGTGGCA
	IGK-V-I-1	0.1µM	5'	GTAAAACGACGGCCAGTAATCCCACCTCGATTCAGTGGC
	IGK-V-J-1	0.1µM	5'	GTAAAACGACGGCCAGTCTCAGGGGTCCCCTCGAGGTT
	IGK-V-K-1	0.1µM	5'	GTAAAACGACGGCCAGTAGACACTGGGGTCCCAGCCA
	IGK-DE-A-1	0.1µM	3'	TAATACGACTCACTATAGGGGCAGCTGCAGACTCATGAGGAG
	IGk-J-A-1	0.1µM	3'	TAATACGACTCACTATAGGGACGTTTGATCTCCACCTTGGTCCC
	IGK-J-B-1	0.1µM	3'	TAATACGACTCACTATAGGGACGTTTGATATCCACTTTGGTCCC
	IGK-J-C-1	0.1µM	3'	TAATACGACTCACTATAGGGACGTTTAATCTCCAGTCGTGTCCC
IGK-B	IGK-INTR-A-1	0.1µM	5'	GTAAAACGACGGCCAGTGAGTGGCTTTGGTGGCCATGC
	IGK-DE-A-1	0.1µM	3'	TAATACGACTCACTATAGGGGCAGCTGCAGACTCATGAGGAG

2nd step PCR	Primer nomenclature	µM in PCR	Primer direction	Illumina primer sequence with M13 adapter (forward/reverse) 5' to 3'
forward	Ill-D501-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D502-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D503-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D504-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D505-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D506-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D507-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACTCAGAAGCGTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
	Ill-D508-F	0.2µM	5'	AATGATACGGCGACCACCGAGATCTACACTAGGCATGACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAAAACGACGGCCAGT
reverse	Ill-D701-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAGGG
	Ill-D702-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATTCTCCGCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAGGG
	Ill-D703-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTAATACGACTCACTATAGGG

I11-D704-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D705-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D706-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D707-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D708-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D709-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D710-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D711-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
I11-D712-R	0.2µM	3'	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Primer3[31], Primer Digital (PrimerDigital Ltd, Helsinki, Finland) MFEprimer-2.0[32] and Oligo (Molecular Biology Insights, Inc., Colorado, USA) were used for checking primer specificity and multiplexing. Primer design criteria were followed for all loci: primer melting temperature 57-63°C; comparable size of final amplicon; primer length 20-24nt; avoidance of primer dimers; minimal distance of 3´primer end to the junctional region of, preferably, >10-15bp to avoid false negativity for rearrangements with larger nucleotide deletions from the germline sequence; avoidance of regions with known single nucleotide polymorphisms to allow identical primer annealing for all alleles of the respective V, D or J genes; targeting of, preferably, all V, D and J genes known to be rearranged plus the intronRSS and Kde regions for IGK.

Following *in silico* design, primers were first tested in monoplex and multiplex reactions using primary patient samples or cell lines with defined rearrangements. In occasional cases where no such samples were available, healthy tonsil or mononuclear DNA samples were employed. Oligoclonal template pools were then created from mixtures of rearranged cell lines and diagnostic samples with defined rearrangements covering many different V, D and/or J genes. Alternatively, for some loci, plasmid pools were produced, covering as many different rearrangements as possible. These multi-target pools allowed fine-tuning of reaction conditions and/or primer concentrations to assess comparable amplification efficiencies. This iterative process of testing also led to a reduction of primers if these appeared redundant. Further multicentre testing was performed with a limited number of monoclonal and poly/oligoclonal samples and on different sequencing platforms, which allowed assessment of robustness of the primer mixes and protocols.

Since the assays were designed with the aim to be platform-independent, a two-step PCR was employed, enabling to switch the sequencing adaptors

and to reduce the total number of primers even if a large number of barcodes is necessary. Also, maximal amplicon lengths were defined with respect to the possible maximal sequencing read lengths of current sequencers. PCR conditions were optimized with the aim to find optimal conditions common for all reactions, thus allowing for parallel library preparation. Various numbers of PCR cycles in 1st and 2nd PCR, different polymerases and several library purification methods were tested and compared.

Although this study was exclusively performed on the Illumina MiSeq, the applicability of the same PCR panel on the IonTorrent instrument (ThermoFischer Scientific) was tested in a single-centre setting and a one-step Illumina MiSeq PCR approach was also tested in a single-center setting.

Multicentre validation of assays for MRD marker identification in ALL

Five experienced EuroClonality-NGS laboratories tested the robustness and applicability of the optimized assays for IG/TR marker identification in ALL in comparison to standard techniques. All laboratories (Bristol/London, Paris, Monza, Prague and Kiel) are members of the EuroMRD consortium and reference laboratories for ALL MRD analysis. Each of the participating laboratories performed NGS-based IG/TR MRD marker identification in 10 patients with B- or T-lineage ALL. A central standard operating procedure was strictly followed by all laboratories. The study was executed using the Illumina MiSeq (2×250bp v2 kit). NGS analyses were performed fully in parallel to conventional PCR plus Sanger sequencing of clonal products following standard guidelines[11]. For a part of the cases with unexplained discrepant results between the two methods, allele-specific PCR assays (either for digital droplet PCR or real-time quantitative PCR) were designed to clarify if the respective clonal rearrangement represented the leukemic

bulk. EuroMRD guidelines were used to design and interpret allele-specific PCR assays[33,34].

RESULTS

Primer design and technical validation of primer performance

Based on the results of the testing and validation phases, the final IG/TR primer mixes consist of eight tubes with 92 forward and 30 reverse primers, 15 of the latter being used in pairs of different tubes). Primer positions and sequences are presented in Figure 5 and Table 3.

Implementation of quality control compositions

Quality control of robust amplification, library preparation and sequencing are of utmost importance for these complex assays. Different primers need to work under the same reaction conditions, while additional variability can be introduced by sample characteristics and sequencing. Primer performance has to be monitored longitudinally, and for the exact estimation of clonal abundance it is important to correct for the number of sequencing reads per input molecule.

To address these issues, two types of quality control compositions were included: (i) the cIT-QC of Example 1 was spiked to each tube as library control and calibrator, and (ii) the cPT-QC of Example 2 was run in parallel to monitor general primer performance and sequencing.

Laboratory protocol

Primers were tailed with universal and T7-linker sequences, and divided over eight tubes (IGH-VJ, IGH-DJ, IGK-VJ-Kde, intron-Kde, TRB-VJ, TRB-DJ, TRG, TRD). The PCR protocol is summarized in Table 4. Sequencing libraries were prepared via a two-step PCR, each using a final reaction volume of 50µl with 100ng diagnostic DNA and 10ng of polyclonal DNA. For the cIT-QC, genomic DNA of 40 cell equivalents of each the 9 different cell lines were spiked into all samples. $MgCl_2$ was intended to be used at a final

concentration of 1.5 mM, but needed optimization for some tubes. Therefore, master-mixes for the 1st PCR were tube-specific, but the temperature profile was uniform for all tubes.

After the 1st round of PCR, gel electrophoresis was performed to check for the successful amplification of all targets. For TRB, gel extraction of the specific PCR products was performed prior to the 2nd PCR.
All first round PCR products, except for TRB, the PCR products were diluted 1:50 unless amplicons were very weak. The TRB PCR products and PCR products with weak amplicons were used undiluted. Master-mixes for the 2nd PCR and the temperature profiles were identical for all tubes (Table 4). Primers for the 2nd PCR contained sequencing adaptors and sequencing indexes (barcodes). Unique combination of forward and reverse indexes was used for each library. Three µl of undiluted TRB PCR products and 1µl of 1:50-diluted IGH, IGK, TRG, and TRD PCR products were amplified in the 2nd PCR.

Table 4: Standardized PCR protocol. (A) Reaction conditions of 1st and 2nd PCR. (B) PCR Cycling conditions.

A

1st PCR

	Stock concentration	IGH-VJ		IGH-DJ		IGK-VJ-Kde, intron-Kde		TRB-VJ, TRB-DJ		TRG		TRD	
		Final concentration	µl/library	Final concentration	µl/library	Final concentration	µl/library	Final concentration	µl/library	Final concentration	µl/library	Final concentration	µl/library
PCR Buffer II	10x	1x	5	1x	5	1x	5	1x	5	1x	5	1x	5
MgCl2	25 mM	2,5 mM	5	3 mM	6	1,5 mM	3	4 mM	8	4 mM	8	2 mM	4
dNTP-Mix	10 mM	0.2 mM	1	0.4 mM	2.0	0.2 mM	1	0.2 mM	1	0.2 mM	1	0,2mM	1
EagleTaq/AmpliTaq Gold	5 U/µl	1U/rxn	0.2	1.5U/rxn	0.3	1U/rxn	0.2	1U/rxn	0.2	1U/rxn	0.2	1 U/rxn	0.2

reaction volume: 50µl

2nd PCR

	Stock concentration	all tubes	
		Final concentration	µl/sample
PCR Buffer with	10x	1x	5
MgCl2	18 mM	1.8 mM	0
dNTP-Mix	10 mM	0.2 mM	1
Fast Start High Fidelity polymerase	5 U/µl	2.5U/rxn	0.5

reaction volume: 50µl

B

1st PCR

	initial denaturation	94°C	10 min
35 cycles	denaturation	94°C	1 min
	annealing	63°C	1 min
	extension	72°C	30 sec
	final extension	72°C	30 min
		12°C	∞

2nd PCR

	initial denaturation	95°C	2 min
20 cycles	denaturation	94°C	30 sec
	annealing	63°C	30 sec
	extension	72°C	30 sec
	final extension	72°C	5 min
		12°C	∞

Following 2nd PCR, products from all samples of a run were pooled in equimolar ratios into 8 tube-wise subpools and purified by gel-extraction (see Table 5 for the amplicon lengths). Finally, the subpools were pooled equimolarly into one final pool. Sequencing was performed on Illumina MiSeq sequencers, using 2×250bp v2 chemistry with a final concentration of 7 pM for the amplicon library and 10% PhiX control added to avoid low-complexity library issues.

Table 5: Mean size of PCR products after the 2nd PCR (containing the Illumina sequencing adaptors and barcodes).

Gene	Amplicon length (bp)
TRB-VJ	309-407
TRB-DJ	300-408
TRG	256-360
TRD	309-450
IGH-VJ	484-681
IGH-DJ	266-358
IGK-VJ-Kde	296-384
intron-Kde	309-382

Bioinformatic protocol

ARResT/Interrogate was the main bioinformatics platform used in this study, along with Vidjil[47] and IMGT[48] resources for specific aspects of this work. Demultiplexing was performed accepting no mismatches. Reads were annotated with EuroClonality-NGS primer sequences (to trim non-amplicon sequence, and for the cPT-QC-based quality control), paired-end joined, dereplicated, immunogenetically annotated[48], and classified into rearrangement types (complete and incomplete, and other special types like intron-Kde rearrangements), or "junction classes". Reads with no

rearrangement were excluded from the total read count used for relative abundances.

cIT-QC sequences described above were identified in the data through their immunogenetic annotation. Their counts served both as 'in-tube' control and for normalization per primer set: total cIT-QC cells are divided by cIT-QC total reads, the resulting factor used to convert rearrangement reads to cells, those cells divided by total input cells (15,000 in this example). Identified IG/TR sequences were defined as index sequences if abundance after cIT-QC normalisation exceeded 5%. ARResT/Interrogate can track the DNJ 3'stem of a junction, the sequence remaining stable during IGH or TRB clonal evolution in case of V-replacement or ongoing V to DJ rearrangements. The stem consists of the last ≤3nt of D (or of the NDN if no D is identifiable), any and all of N2 nucleotides, and the J nucleotides of the junction. This stem is available as a separate immunogenetic feature across all samples and thus able to link other features, e.g. clonotypes.

Multicentre validation of assays for MRD marker identification in ALL

Next, fifty ALL diagnostic samples (29 BCP-ALL and 21 T-ALL) were analysed for the multicentre validation study. Each of the five participating laboratories received preconfigured 96-well plates containing the different multiplexed NGS primer combinations per target (Figure 4).

In summary, 96 libraries were generated per lab (total of 480 libraries), and sequenced with a total output of 47M reads (≃ 9.2M/lab). Centralised analysis was performed with ARResT/Interrogate[43] using IMGT germline sequences[48] – further analyses and verifications were performed with Vidjil[47] and IMGT/V-QUEST[48].

Overall, 311 clonal IG/TR rearrangements (clonotypes) were identified, with a mean of 5.9 (0-14)/sample by NGS (a 5% threshold was applied for NGS after cIT-QC-based normalization) vs. 5.0 (0-14)/sample by Sanger, while

217 (45%) libraries demonstrated no clonotypes above threshold by either method. A total of 196/311 (63%) clonotypes were fully concordant between NGS and Sanger (Figure 6). NGS exclusively identified 63 index sequences, whereas 52 IG/TR Sanger sequences were not assigned as NGS index sequence by ARResT/Interrogate. 26 NGS pos/Sanger neg cases showed a clonal PCR product also in the respective low-throughput approach but subsequent Sanger sequencing failed due to polyclonal background, mixed sequences or weak PCR products. In an additional 6 Sanger neg/NGS pos cases, the respective primer was missing in the low-throughput approach. For the remaining 31 discrepancies no technical explanation for Sanger failure could be found (in 16/19 q/ddPCR evaluated cases the rearrangement was confirmed by ASO-PCR, in 3/16 on a subclonal level).

Conversely, 52 clonal IG/TR rearrangements were only detected by Sanger when the 5% NGS threshold was applied: for 5 sequences (1 TRG, 2 TRB-VJ, and 2 IGH-DJ) the relevant primer was not present in the NGS primer set, in 12 cases no explanation was found for the discrepancy. However, in the majority of discordant cases (35/52) the Sanger identified sequences (7 TRD, 8 TRB-VJ, 6 TRG, 4 TRB-DJ, 2 IGK-VJ-Kde, 5 IGH-VJ, 3 IGH-DJ) were also detectable by NGS, but with and abundance below 5%. In 36/39 q/ddPCR evaluated cases the rearrangement was confirmed by ASO-PCR, including all low NGS positive sequences, in 14/36 cases on a subclonal level. Overall concordance between Sanger and NGS, including negative libraries, was 78%.

In 12/29 B-lineage ALL samples the evolution of the dominant clonal IGH sequence was identified employing ARResT/Interrogate. The evolved clonotypes shared the DNJ stem with the dominant one, but the VND part of the rearrangement differed.

The assay performance was also analysed by standardized evaluation of QC samples (cIT-QC and cPT-QC). This showed a remarkably high intra- and

inter-lab consistency without statistically significant differences between the five labs.

Suitable modifications of the central SOP for MRD marker identification

During the process of multicentre validation, suitable modifications of the SOP were tested in particular laboratories as parallel actions.

One-step versus two-step PCR: The EuroClonality-NGS working group decided to use two-step PCR to enable switching of sequencing adaptors and to limit the total number of required primer batches even if a large number of barcodes is necessary. As first round PCR products are not barcoded, identification of contamination phenomena is hampered in this approach. Therefore, a one-step PCR was tested in a single center (Paris) as an alternative for laboratories that are able to maintain higher numbers of different primer batches. The one-step approach reduces the risk of contamination and thus favours use of NGS not only for marker identification, but also for MRD assessment. The standard operating procedures are shown in supplementary information.

Bead extraction: In our single target evaluation and validation phase, gel extraction of the specific TRB amplicons turned out to lead to more specific libraries compared to bead extraction. However, gel extraction is not used in all laboratories, therefore, in a later phase of the study bead purification of all libraries was also tested. Optimization of the purification processes led to comparable ratios of specific reads irrespective of the type of library purification.

Withdrawal of addition of polyclonal DNA to reaction mix: Polyclonal DNA was added to each reaction in order to prevent excessive primer dimer formation in samples lacking particular rearrangements. The addition of polyclonal DNA, however, alters the composition of polyclonal background of the samples and hampers the analysis of the immune repertoire. We

therefore performed testing on 4 samples with B and 4 samples with T cell aplasia and showed that addition of cIT-QC is sufficient to prevent the excessive formation of unspecific PCR products.

REFERENCES

1 Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983; **302**: 575–581.

2 Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature* 1988; **334**: 395–402.

3 Schlissel MS. Regulating antigen-receptor gene assembly. *Nat Rev Immunol* 2003; **3**: 890–899.

4 Lefranc M-P, Lefranc G. *The T cell receptor factsbook*. Academic Press, 2001.

5 Lefranc M-P, Lefranc G. *The immunoglobulin factsbook*. Academic Press, 2001.

6 Monroe JG, Dorshkind K. Fate Decisions Regulating Bone Marrow and Peripheral B Lymphocyte Development. In: *Advances in immunology*. 2007, pp 1–50.

7 von Boehmer H, Melchers F. Checkpoints in lymphocyte development and autoimmune disease. *Nat Immunol* 2010; **11**: 14–20.

8 Evans PAS, Pott C, Groenen PJTA, 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**: 207–214.

9 Brüggemann M, White H, Gaulard P, Garcia-Sanz R, Gameiro P, Oeschger S *et al.* Powerful strategy for polymerase chain reaction-based clonality assessment in T-cell malignancies Report of the BIOMED-2 Concerted Action BHM4 CT98-3936. *Leukemia* 2007; **21**: 215–221.

10 Langerak AW, Groenen PJTA, Brüggemann 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**: 2159–2171.

11 van Dongen JJM, LangerakAW, Brüggemann M, Evans PA, Hummel M, Lavender FL *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.

12 Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 2009; **1**: 12ra23.

13 DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* 2013; **31**: 166–169.

14 Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* 2009; **19**: 1817–1824.

15 Gawad C, Pepin F, Carlton VEH, Klinger M, Logan AC, David B *et al.* Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. 2012; **120**: 4407–4417.

16 Logan AC, Gao H, Wang C, Sahaf B, Jones CD, Marshall EL *et al.* High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. *Proc Natl Acad Sci U S A* 2011; **108**: 21194–21199.

17 Logan AC, Zhang B, Narasimhan B, Carlton V, Zheng J, Moorhead M *et al.* Minimal residual disease quantification using consensus primers and high-throughput IGH sequencing predicts post-transplant relapse in chronic lymphocytic leukemia. *Leukemia* 2013; **27**: 1659–1665.

18 Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR *et al.* Overlap and Effective Size of the Human CD8+ T Cell Receptor Repertoire. *Sci Transl Med* 2010; **2**: 47ra64-47ra64.

19 Wang C, Sanders CM, Yang Q, Schroeder HW, Wang E, Babrzadeh F *et al.* High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci* 2010; **107**: 1518–1523.

20 Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML *et al.* High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. *SciTranslMed* 2012; **4**: 134ra63.

21 Wu Y-C, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 2010; **116**: 1070–1078.

22 Bartram J, Goulden N, Wright G, Adams S, Brooks T, Edwards D *et al.* High throughput sequencing in acute lymphoblastic leukemia reveals clonal architecture of central nervous system and bone marrow compartments. *Haematologica* 2018; **103**: e110–e114.

23 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**: 5173–5180.

24 Ladetto M, Bruggemann M, Monitillo L, Ferrero S, Pepin F, Drandi D *et al.* Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. *Leukemia* 2014; **28**: 1299–1307.

25 Pulsipher MA, Carlson C, Langholz B, Wall DA, Schultz KR, Bunin N *et al.* IgH-V(D)J NGS-MRD measurement pre- and early post- allo-transplant defines very low and very high risk ALL patients. *Blood* 2015; **125**: 3501–3508.

26 Kotrova M, Muzikova K, Mejstrikova E, Novakova M, Bakardjieva-Mihaylova V, Fiser K *et al.* The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL. *Blood* 2015; **126**: 1045–7.

27 Langerak AW, Brüggemann M, Davi F, Darzentas N, Gonzalez D, Cazzaniga G *et al.* High throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. *J Immunol* 2017; **198**: 3765–3774.

28 Kotrova M, van der Velden VHJ, van Dongen JJM, Formankova R, Sedlacek P, Brüggemann M *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.

29 Kotrova M, Trka J, Kneba M, Brüggemann M. Is Next-Generation Sequencing the way to go for Residual Disease Monitoring in Acute Lymphoblastic Leukemia? *Mol Diagn Ther* 2017. doi:10.1007/s40291-017-0277-9.

30 Pott C. Minimal Residual Disease Detection in Mantle Cell Lymphoma: Technical Aspects and Clinical Relevance. *Semin Hematol* 2011; **48**: 172–184.

31 Ferrero S, Drandi D, Mantoan B, Ghione P, Omedè P, Ladetto M. Minimal residual

disease detection in lymphoma and multiple myeloma: Impact on therapeutic paradigms. Hematol. Oncol. 2011; **29**: 167–176.

32 Brüggemann M, Gökbuget N, Kneba M. Acute Lymphoblastic Leukemia: Monitoring Minimal Residual Disease as a Therapeutic Principle. *Semin Oncol* 2012; **39**: 47–57.

33 Brüggemann M, Raff T, Kneba M. Has MRD monitoring superseded other prognostic factors in adult ALL? *Blood* 2012; **120**: 4470–4481.

34 van Dongen JJM, Seriu T, Panzer-Grumayer ER, Biondi A, Pongers-Willemse MJ, Corral L *et al.* Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* 1998; **352**: 1731–1738.

35 Brüggemann M, Kotrova M. Minimal residual disease in adult ALL: technical aspects and implications for correct clinical interpretation. *Hematol Am Soc Hematol Educ Progr* 2017; : 13–21.

36 Logan AC, Vashi N, Faham M, Carlton V, Kong K, Buño I *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.

37 Wren D, Walker BA, Brüggemann M, Catherwood MA, Pott C, Stamatopoulos K *et al.* Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. Haematologica. 2017; **102**: e57–e60.

38 Hardwick SA, Deveson IW, Mercer TR. Reference standards for next-generation sequencing. Nat. Rev. Genet. 2017; **18**: 473–484.

39 Gargis AS, Kalman L, Lubin IM. Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. *J Clin Microbiol* 2016; **54**: 2857–2865.

40 Endrullat C, Glökler J, Franke P, Frohme M. Standardization and quality management in next-generation sequencing. Appl. Transl. Genomics. 2016; **10**: 2–9.

41 Kurtz DM, Green MR, Bratman S V., Scherer F, Liu CL, Kunder CA *et al.* Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood* 2015; **125**: 3679–3687.

42 Pulsipher MA, Carlson C, Langholz B, Wall DA, Schultz KR, Bunin N *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**: 3501–3509.

43 Bystry V, Reigl T, Krejci A, Demko M, Hanakova B, Grioni A *et al.* ARResT/Interrogate: an interactive immunoprofiler for IG/TR NGS data. *Bioinformatics* 2016; **33**: btw634.

44 Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR *et al.* Chimeric Antigen Receptor–Modified T Cells for Acute Lymphoid Leukemia. *N Engl J Med* 2013; **368**: 1509–1518.

45 Tang M, Wang G, Kong SK, Ho HP4. A Review of Biomedical Centrifugal Microfluidic Platforms. *Micromachines (Basel)* 2016; **7**: E26.

46 Langerak AW, Szczepański T, Van Der Burg M, Wolvers-Tettero ILM, Van Dongen JJM. Heteroduplex PCR analysis of rearranged T cell receptor genes for clonality assessment in suspect T cell proliferations. *Leukemia* 1997; **11**: 2192–2199.

47 Duez M, Giraud M, Herbert R, Rocher T, Salson M, Thonier F. Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing. *PLoS One* 2016; **11**: e0166126.

48 Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G, Kossida S. IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res* 2015; 43: D413-22.

Claims

1. A composition comprising a mixture of genomic DNA isolated from a set of nine cultured cell lines, said set comprising the B cell lines ALL/MIK (B cell precursor ALL), Raji (Burkitt lymphoma), REH (B cell precursor ALL), TMM (CML-BC / EBV+B-LCL), TOM-1 (B cell precursor ALL), WSU-NHL (B cell lymphoma) and the T cell lines JB6 (ALCL) , Karpas299 (ALCL) and MOLT-13 (T-ALL), or wherein one or more cell lines of said set is replaced with one or more other cell line(s) comprising the same immunoglobulin (IG)/T cell receptor (TR) gene rearrangements.

2. Composition according to claim 1, comprising a mixture of genomic DNA isolated from the B cell lines ALL/MIK, Raji, REH, TMM, TOM-1, WSU-NHL and the T cell lines JB6, Karpas299 and MOLT-13.

3. Composition according to claim 1 or 2, wherein said composition comprises essentially equal amounts of genomic DNA of each of said cell lines.

4. A composition consisting of essentially equimolar amounts of genomic DNA isolated from healthy human thymus, healthy human tonsil and healthy human peripheral blood mononuclear cells.

5. Composition according to claim 4, wherein, for each tissue, the genomic DNA is obtained from a number, preferably 3 to 10, different human individuals.

6. Diagnostic kit comprising a container comprising a composition according to any one of claims 1 to 3, and / or a container comprising a composition according to claim 4 or 5.

7. Kit according to claim 6, comprising a first container comprising a composition according to any one of claims 1 to 3, and a second container comprising a composition according to claim 4 or 5.

8. Kit according to claim 6 or 7, further comprising one or more reagents for detecting immunoglobulin (IG)/T cell receptor (TR) gene rearrangements.

9. Kit according to claim 8, comprising a set of primers for amplicon-based next-generation sequencing (NGS) of IG/TR gene rearrangements.

10. Kit according to claim 8 or 9, comprising primer sets for detecting one or more of the IG/TR gene rearrangements selected from the group consisting of IGH-VJ, IGH-DJ, IGK-VJ-Kde, TRB-VJ, TRB-DJ, TRG and TRD.

11. Kit according to claim 9 or 10, comprising one or more of the primers selected from the primers shown in Figure 5, preferably one or more of the primers selected from the primers shown in Table 3.

12. A set of primers for amplicon-based next-generation sequencing (NGS) of IG/TR gene rearrangements, comprising two or more of the primers selected from the primers shown in Figure 5.

13. Set of primers according to claim 12, comprising two or more of the primers selected from the primers shown in Table 3.

14. The use of a composition according to any one of claims 1-5, a kit according to any one of claims 6-11, and/or a primer set according to claim 12 or 13 in an assay for detecting IG/TR gene rearrangements.

15. The use according to claim 14, wherein said assay is a clinical diagnostic assay, preferably an assay for detecting clonality, identifying

minimal residual disease (MRD) markers and/or MRD monitoring and/or analyzing the (clonal) immune repertoire in a lymphoid malignancy.

16. An *in vitro* method for detecting IG/TR gene rearrangements in at least one biological sample using NGS, comprising the steps of sample preparation, PCR and/or library construction, sequencing and bioinformatics analysis, wherein the at least one biological sample is spiked with a composition according to any one of claims 1 to 3, and/or wherein a composition according to claim 4 or 5 is run as a sample parallel to the at least one biological sample(s).

17. Method according to claim 16, wherein the at least one biological sample is a clinically relevant sample, preferably a sample for detection of clonality to support or exclude the diagnosis of malignant lymphoproliferation, or a sample taken for MRD marker identification or for MRD monitoring analysis or for (clonal) immune repertoire analysis.

18. Method according to claim 17 or 18, wherein at least part of the method is performed using a microfluidics device.

19. Method according to claim 18, wherein said microfluidics device comprises a centrifugal-microfluidic disk system, preferably wherein the disk comprises pre-stored reagents for automated and integrated DNA extraction, PCR and/or library generation.

20. Method according to any one of claims 17-19, wherein the step of bioinformatic analysis comprises the use of a web-based, interactive application for pre-processing of raw data, primer sequence analysis, immunogenetic annotation, post-processing of results, analysis and use of the cIT-QC (including for marker quantification), analysis and use of the cPT-QC (including for comparison to pre-analyzed stored reference datasets), reporting of / access to / visualization of results.

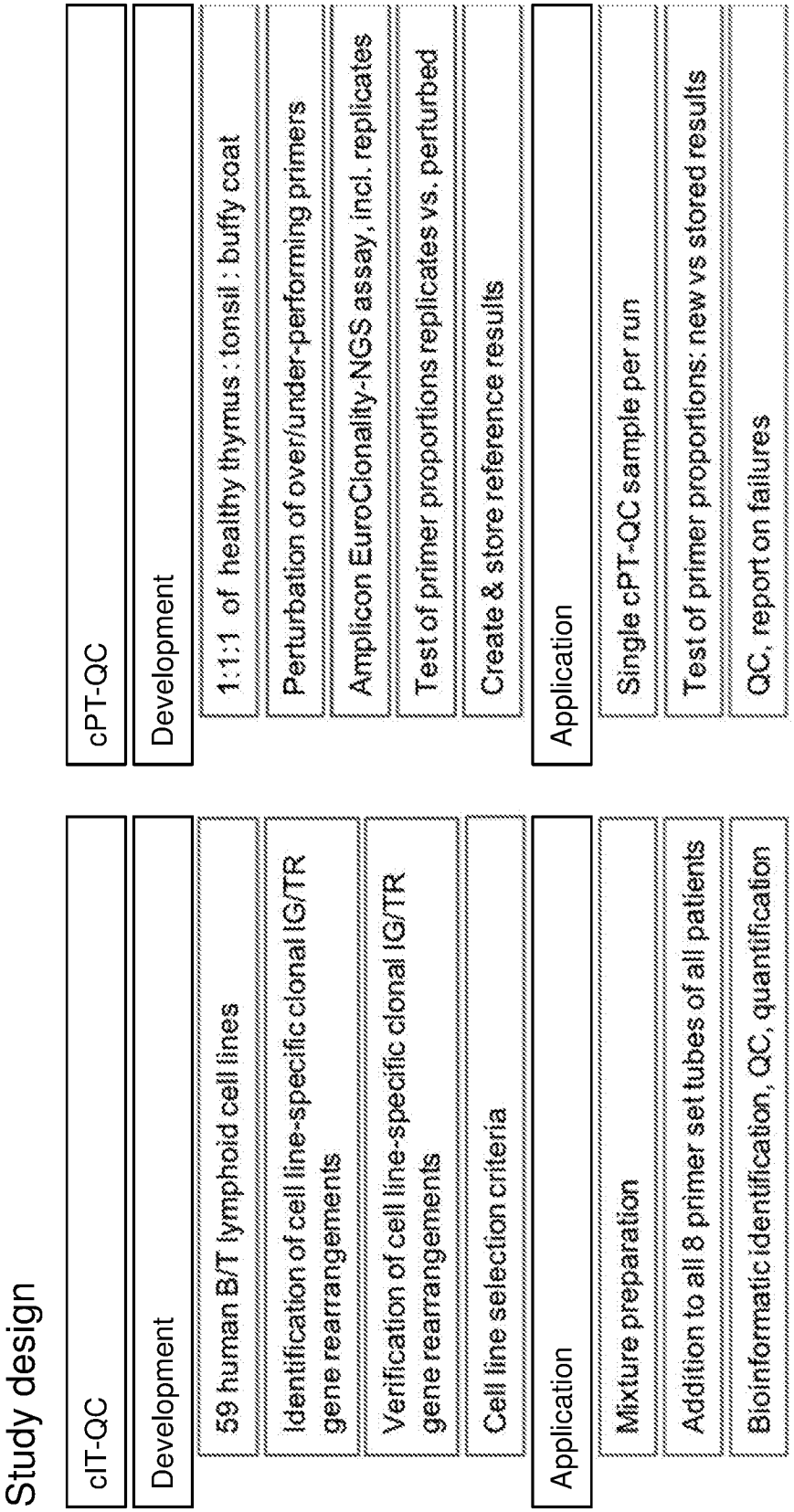

Study design
cIT-QC
Development
59 human B/T lymphoid cell lines
Identification of cell line-specific clonal IG/TR gene rearrangements
Verification of cell line-specific clonal IG/TR gene rearrangements
Cell line selection criteria
Application
Mixture preparation
Addition to all 8 primer set tubes of all patients
Bioinformatic identification, QC, quantification
cPT-QC
Development
1:1:1 of healthy thymus : tonsil : buffy coat
Perturbation of over/under-performing primers
Amplicon EuroClonality-NGS assay, incl. replicates
Test of primer proportions replicates vs. perturbed
Create & store reference results
Application
Single cPT-QC sample per run
Test of primer proportions: new vs stored results
QC, report on failures

Fig. 1 (1-2)

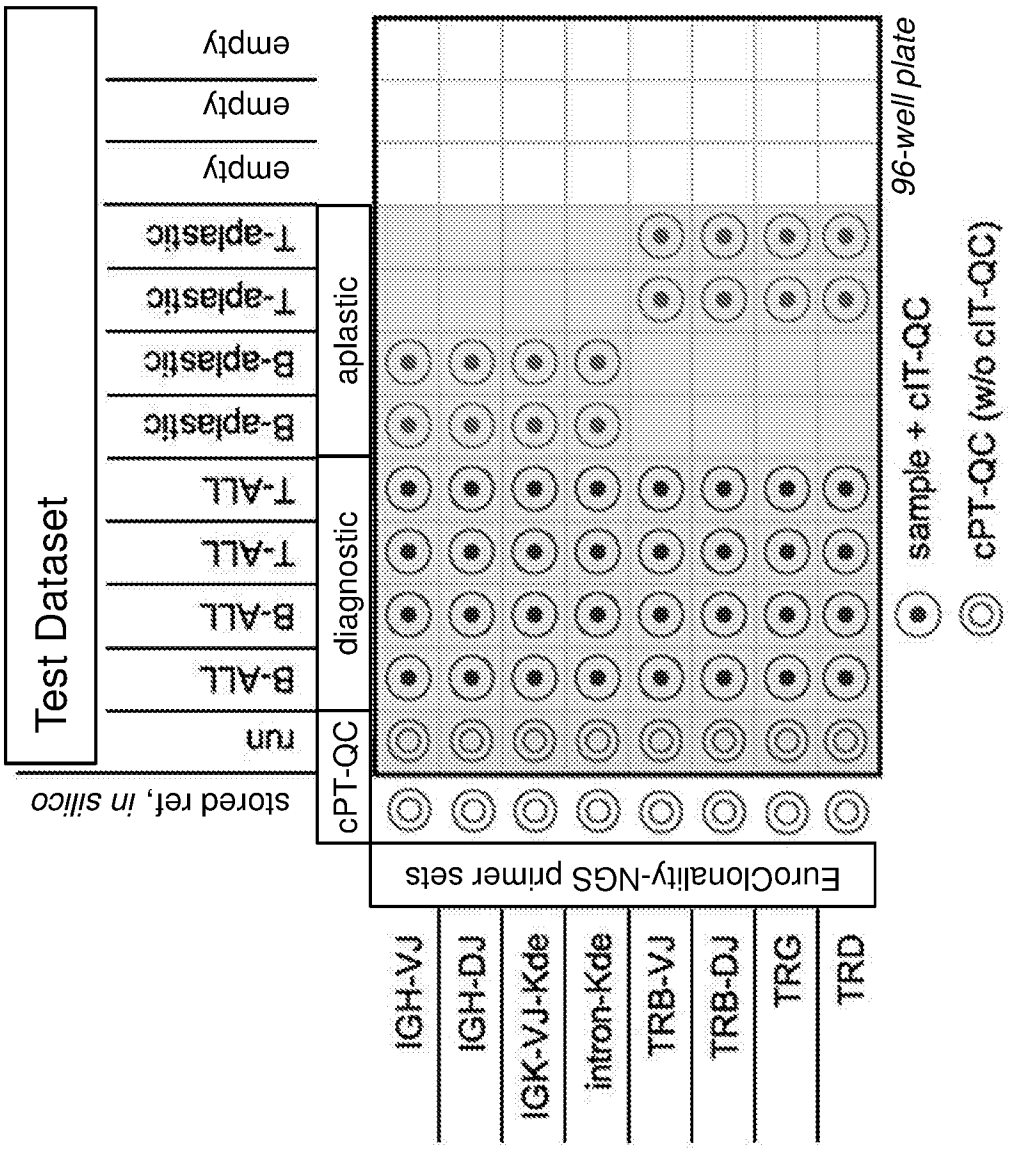
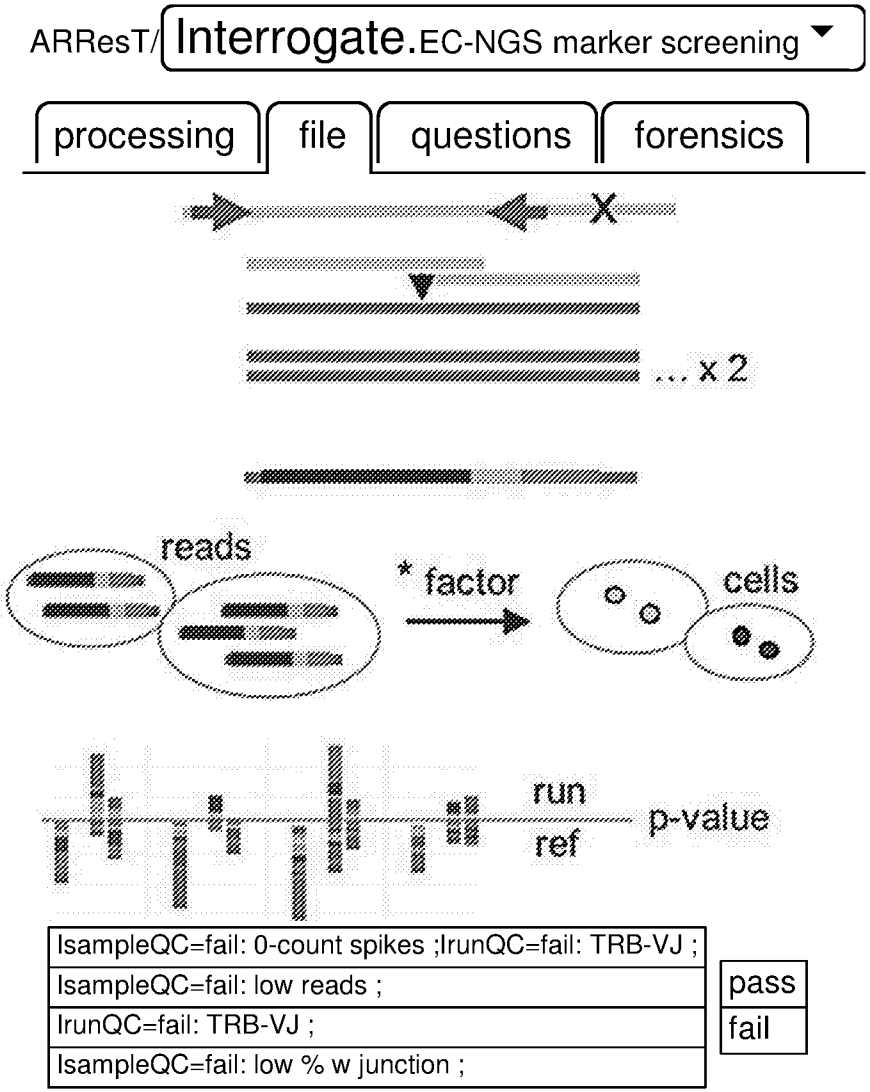

Test Dataset
stored ref, in silico
cPT-QC
run
B-ALL
B-ALL
T-ALL
T-ALL
B-ALL
B-aplastic
B-aplastic
T-aplastic
T-aplastic
empty
empty
empty
diagnostic
aplastic
EuroClonality-NGS primer sets
IGH-VJ
IGH-DJ
IGK-VJ-Kde
intron-Kde
TRB-VJ
TRB-DJ
TRG
TRD
96-well plate
sample + cIT-QC
cPT-QC (w/o cIT-QC)

Fig. 1 (2-2)

[see accompanying manuscript, Brüggemann et al, 2018]

Fig. 2 (1-2)



bioinformatics | ARResT/Interrogate

arrest.tools/interrogate contact@arrest.tools

process
- primer analysis (tagging, trimming)
- paired-end joining, saving of unjoined
- dereplication

compile junctions
- alignment vs germlines, junction analysis

equivalent analysis of cIT-QC sequences

profile
- clustering and clonotype identification
- cIT-QC analysis, quantification factors
- \ converting reads to cells
- loading of stored reference cPT-QC data
- \ run QC with stored vs new cPT-QC primer %s

reporting
- full run and sample reports and logs
- tagging tubes as 'QC-failed'

Fig. 2 (2-2)

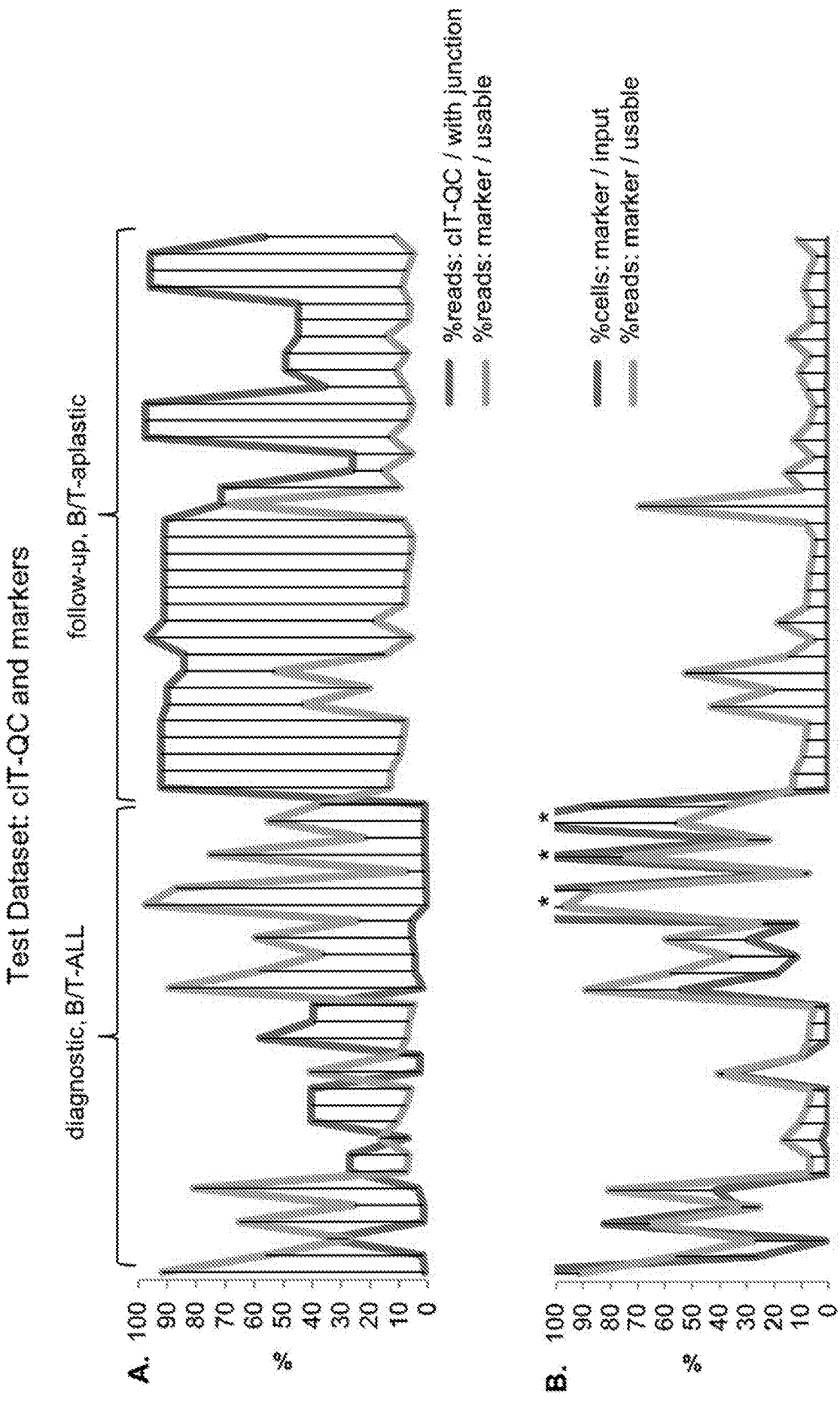

Test Dataset: clT-QC and markers
diagnostic, B/T-ALL
follow-up, B/T-aplastic
%reads: clT-QC / with junction
%reads: marker / usable
%cells: marker / input
%reads: marker / usable
A.
B.
%
%
100
90
80
70
60
50
40
30
20
10
0
*
*
*

Fig. 3

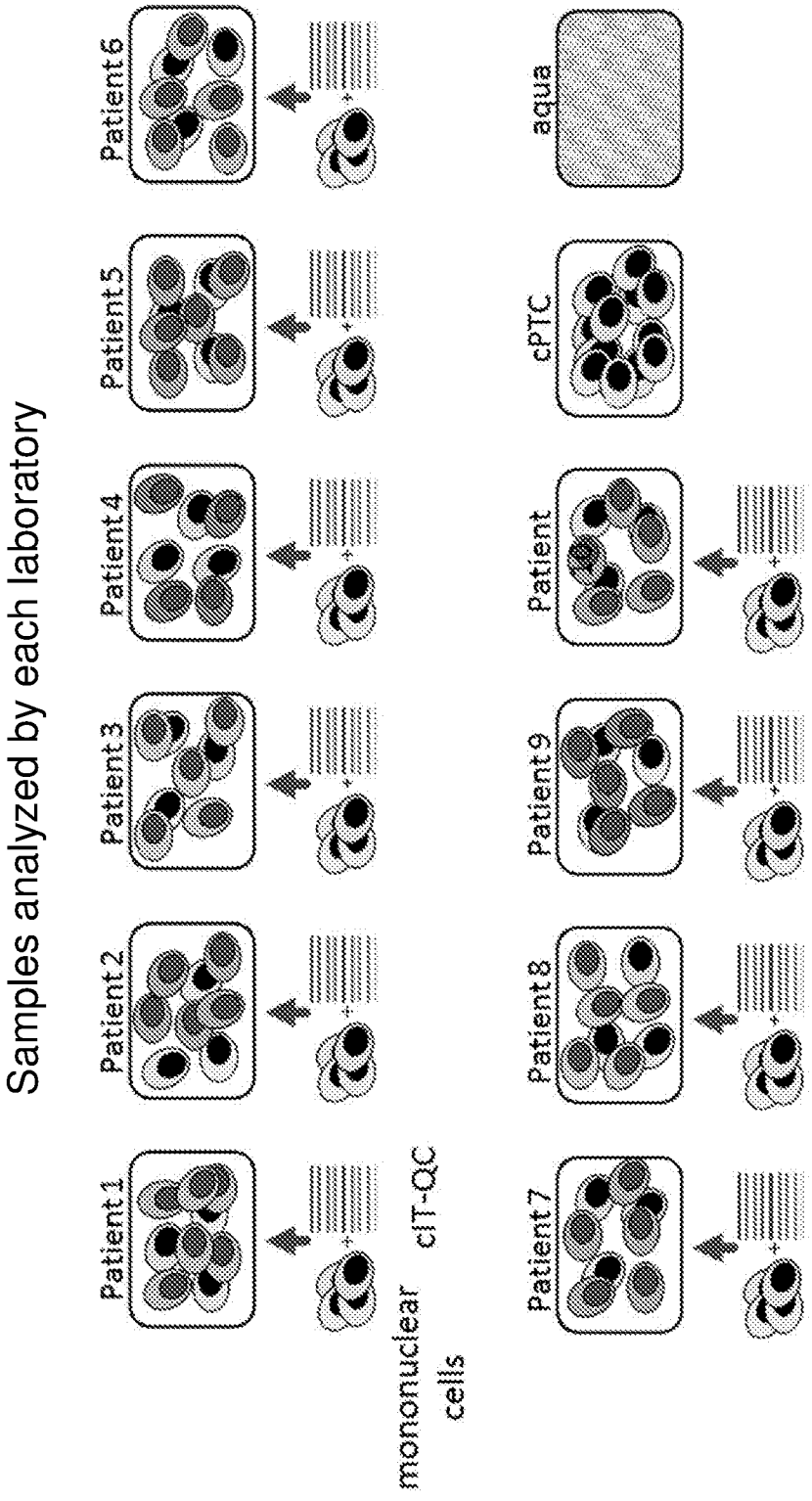

Samples analyzed by each laboratory
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient
cPTC
aqua
mononuclear cells
cIT-QC

Fig. 4 (1-4)

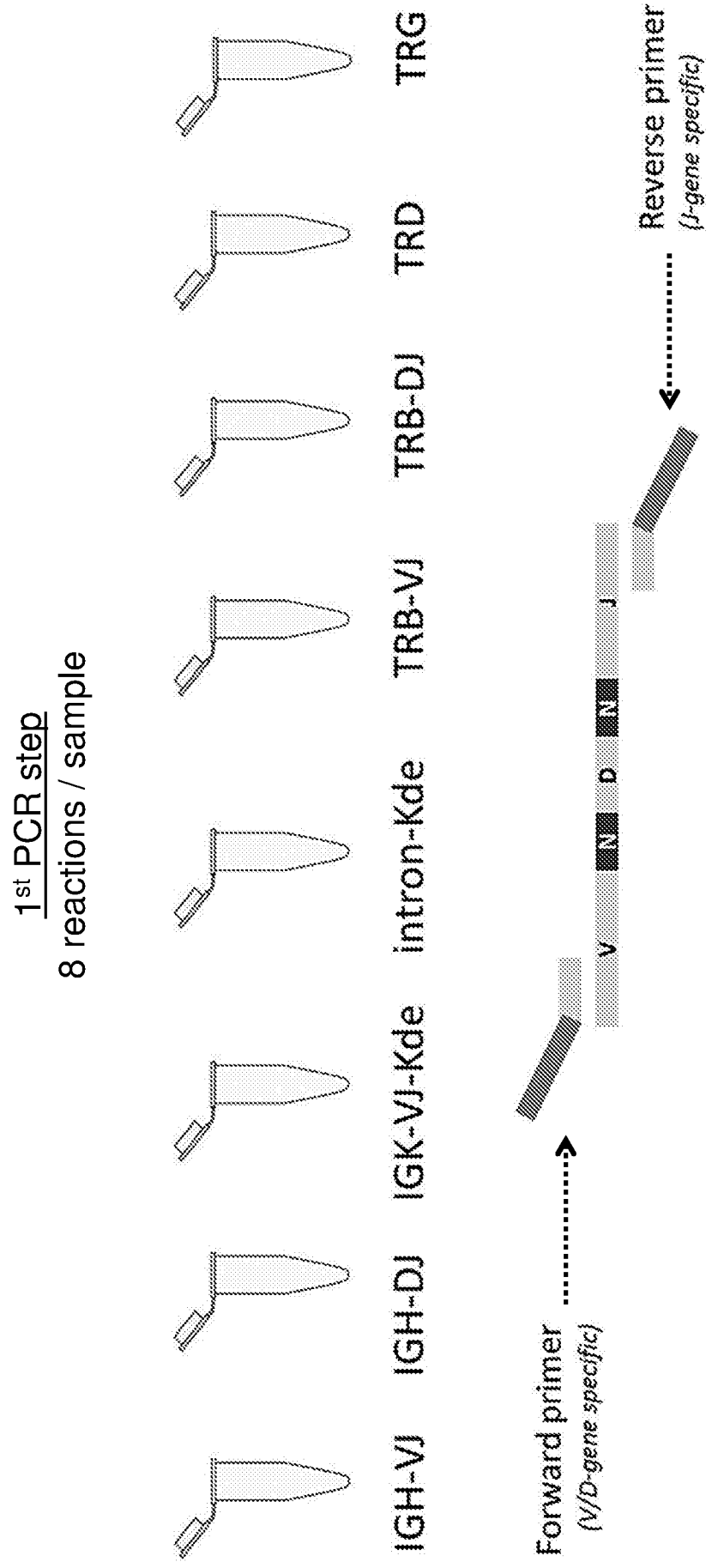

1st PCR step
8 reactions / sample
IGH-VJ
IGH-DJ
IGK-VJ-Kde
intron-Kde
TRB-VJ
TRB-DJ
TRD
TRG
Forward primer
(V/D-gene specific)
Reverse primer
(J-gene specific)
V
N
D
N
J

Fig. 4 (2-4)

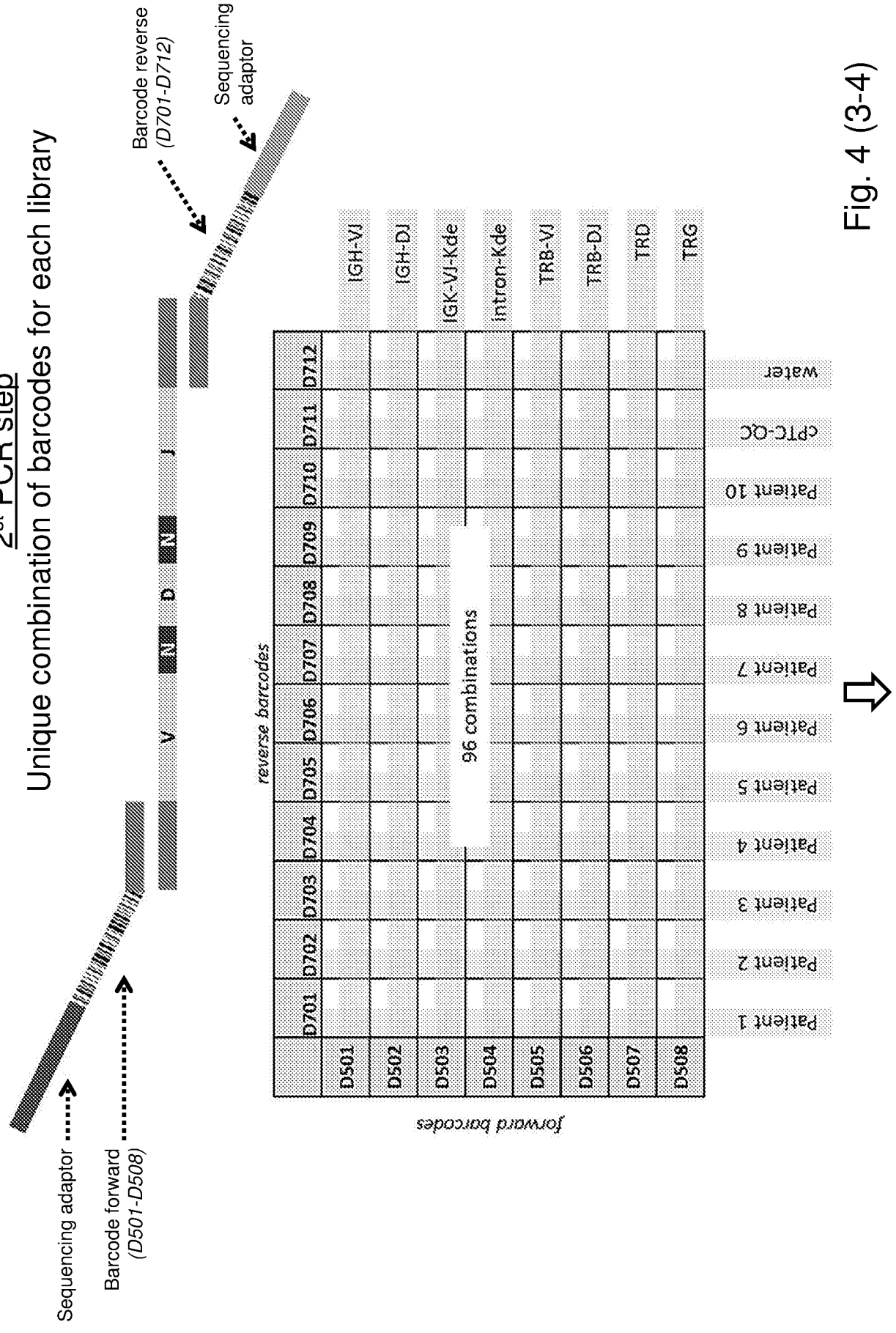
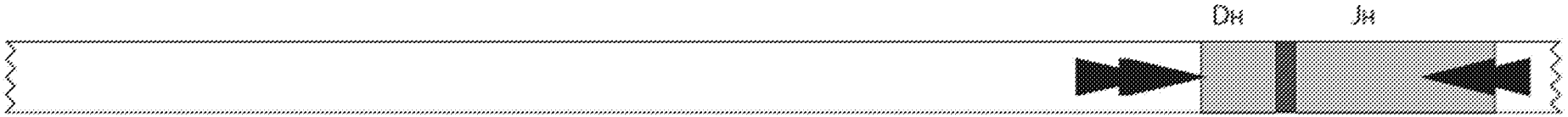

2st PCR step
Unique combination of barcodes for each library
Barcode reverse (D701-D712)
Sequencing adaptor
Sequencing adaptor
Barcode forward (D501-D508)
V
N
D
N
J
reverse barcodes
D701
D702
D703
D704
D705
D706
D707
D708
D709
D710
D711
D712
D501
D502
D503
D504
D505
D506
D507
D508
96 combinations
forward barcodes
IGH-VJ
IGH-DJ
IGK-VJ-Kde
intron-Kde
TRB-VJ
TRB-DJ
TRD
TRG
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient 10
cPTC-QC
water

Fig. 4 (3-4)

Final library
V
N
D
N
J
Bioinformatics
process
primer analysis (tagging, trimming)
paired-end joining, saving of unjoined
dereplication
... x 2
compile junctions
alignment vs germlines, junction analysis
profile
clustering and clonotype identification
clT-QC analysis, quantification factors
\ converting reads to cells
reads
* factor
cells

Fig. 4 (4-4)

IGH-V-FR1-A-1	(IGHV1-2*01)	(-277)	CTGGGGCTGAGGTGAAGAAG	CTTACCTGAGGAGACGGTGACC	(+68)	(IGHJ6*01) IGH-J-A-1
IGH-V-FR1-B-1	(IGHV5-51*01)	(-282)	GCAGTCTGGAGCAGAGGTGAAAA	CTCACCTGAGGAGACAGTGACC	(+58)	(IGHJ2*01) IGH-J-B-1
IGH-V-FR1-C-1	(IGHV2-5*01)	(-297)	TCACCTTGAAGGAGTCTGGTCC			
IGH-V-FR1-D-1	(IGHV3-7*01)	(-295)	AGGTGCAGCTGGTGGAGTC			
IGH-V-FR1-E-1	(IGHV3-23*01)	(-296)	GAGGTGCAGCTGTTGGAGTC			
IGH-V-FR1-F-1	(IGHV4-4*01)	(-272)	CCAGGACTGGTGAAGCCTTC			
IGH-V-FR1-G-1	(IGHV4-34*01)	(-278)	CAGTGGGGCGCAGGACTGTT			
IGH-V-FR1-H-1	(IGHV4-4*01)	(-272)	CCAGGACTGGTGAAGCCTCC			
IGH-V-FR1-I-1	(IGHV6-1*01)	(-302)	GTACAGCTGCAGCAGTCAGG			
IGH-V-FR1-J-1	(IGHV7-4-1*01)	(-286)	GCTGGTGCAATCTGGGTCTG			
IGH-V-FR1-K-1	(IGHV1-45*01)	(-250)	CCTCAGTGAAGGTTTCCTGCAAGG			
IGH-V-FR1-L-1	(IGHV2-26*01)	(-265)	AAACCCACAGAGACCCTCACGCTGAC			
IGH-V-FR1-M-1	(IGHV3-16*01)	(-256)	CTGGGGGGTCCCTGAGACTCTCCTG			
IGH-V-FR1-N-1	(IGHV4-31*01)	(-259)	CTTCACAGACCCTGTCCCTCACCTG			
IGH-V-FR1-O-1	(IGHV6-1*01)	(-263)	TCGCAGACCCTCTCACTCACCTGTG			



IGH-D-A-1	(IGHD1-1*01)	(-44)	GATTCYGAACAGCCCCGAGTCA	CTTACCTGAGGAGACGGTGACC	(+68)	(IGHJ6*01) IGH-J-A-1
IGH-D-B-1	(IGHD2-2*01)	(-58)	GATTTTGTGGGGGYTCGTGTC			
IGH-D-C-1	(IGHD3-3*01)	(-58)	GTTTGRRGTGAGGTCTGTGTCA			
IGH-D-D-1	(IGHD3-3*01)	(-58)	GTTTRGRRTGAGGTCTGTGTCACT			
IGH-D-E-1	(IGHD4-4*01)	(-43)	CTTTTTGTGAAGGSCCCTCCTR			
IGH-D-F-1	(IGHD5-5*01)	(-47)	GTTATTGTCAGGSGRTGTCAGAC			
IGH-D-G-1	(IGHD5-5*01)	(-47)	GTTATTGTCAGGGGGTGYCAGRC			
IGH-D-H-1	(IGHD6-6*01)	(-45)	GTTTCTGAAGSTGTCTGTRTCAC			

Fig. 5A-1

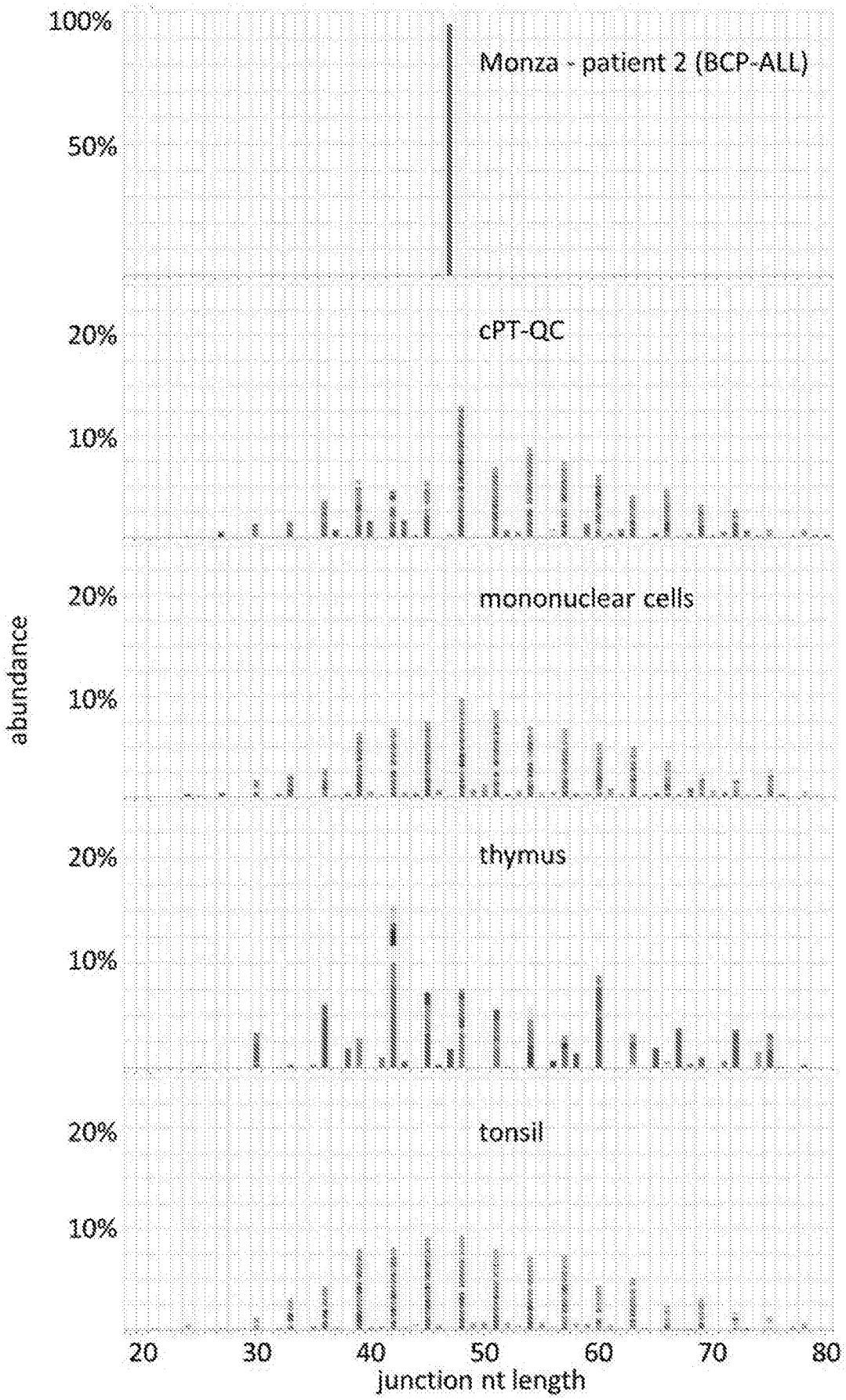

Monza - patient 2 (BCP-ALL)
cPT-QC
mononuclear cells
thymus
tonsil
100%
50%
20%
10%
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5A-2

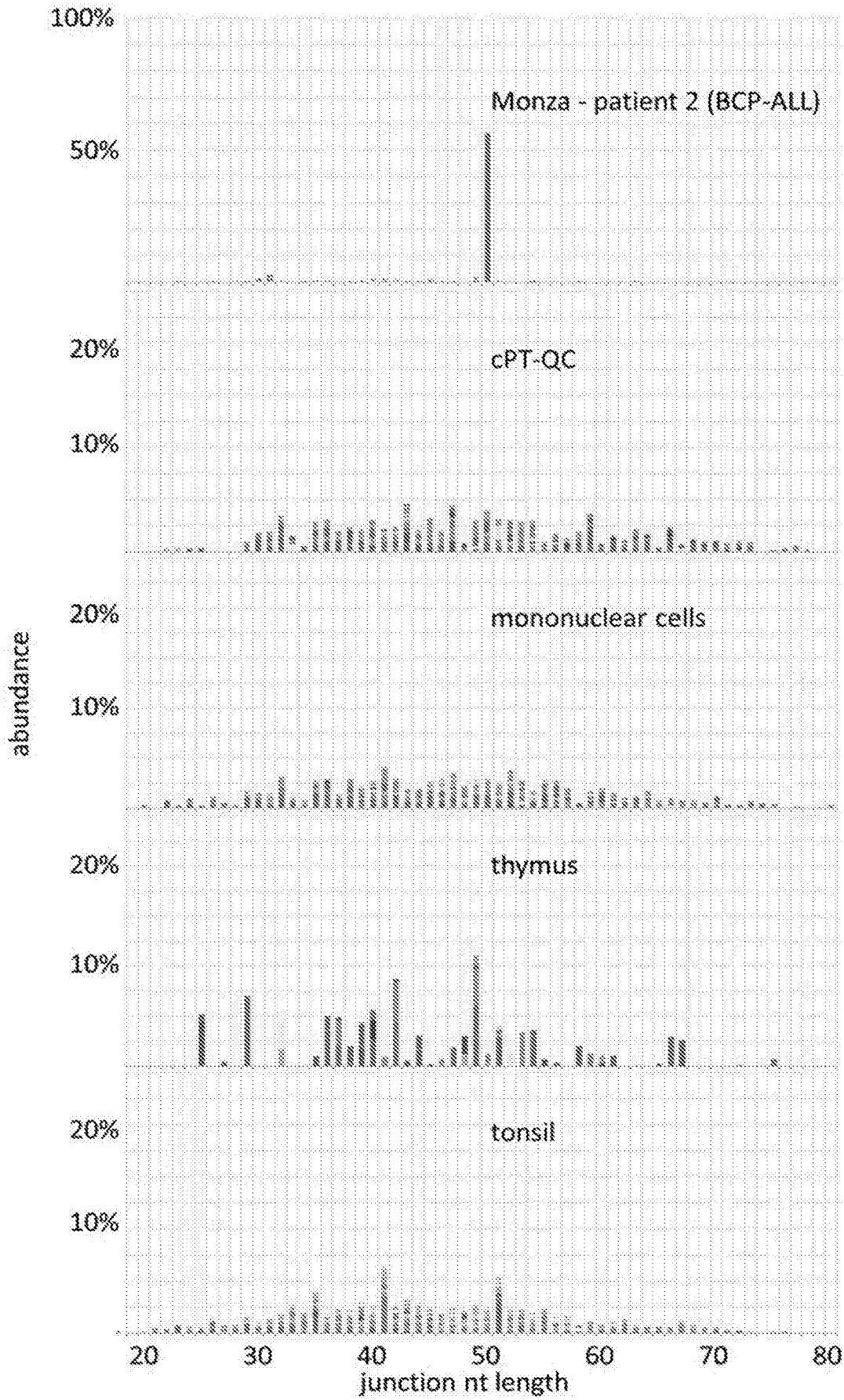

100%
50%
Monza - patient 2 (BCP-ALL)
20%
10%
cPT-QC
20%
10%
mononuclear cells
20%
10%
thymus
20%
10%
tonsil
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5A-3

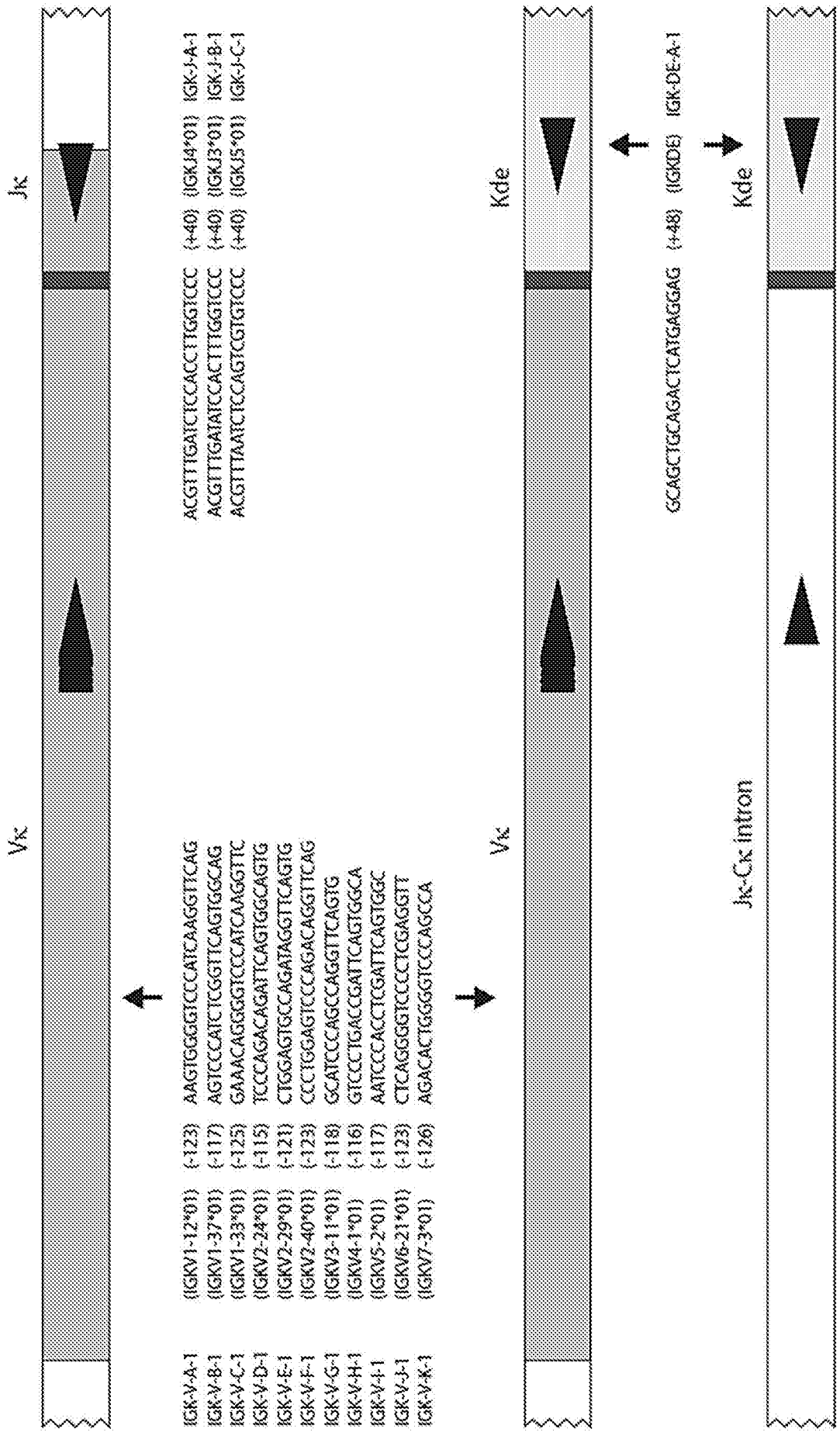

Jκ
Vκ
IGK-V-A-1 (IGKV1-12*01) (-123) AAGTGGGGGTCCCATCAAGGTTCAG
IGK-V-B-1 (IGKV1-37*01) (-117) AGTCCCATCTCGGTTCAGTGGCAG
IGK-V-C-1 (IGKV1-33*01) (-125) GAAACAGGGGGTCCCATCAAGGTTC
IGK-V-D-1 (IGKV2-24*01) (-115) TCCCCAGACAGATTCAGTGGCAGTG
IGK-V-E-1 (IGKV2-29*01) (-121) CTGGAGTGCCAGATAGGTTCAGTG
IGK-V-F-1 (IGKV2-40*01) (-123) CCCTGGAGTCCCAGACAGGTTCAG
IGK-V-G-1 (IGKV3-11*01) (-118) GCATCCCAGCCAGGTTCAGTG
IGK-V-H-1 (IGKV4-1*01) (-116) GTCCCTGACCGAATTCAGTGGCA
IGK-V-I-1 (IGKV5-2*01) (-117) AATCCCCACCTCGATTCAGTGGC
IGK-V-J-1 (IGKV6-21*01) (-123) CTCAGGGGTCCCCTCGAGGTT
IGK-V-K-1 (IGKV7-3*01) (-126) AGACACTGAGGGTCCCAGCCA
ACGTTTGATCTCCACCTTGGTCCC (+40) (IGKJ4*01) IGK-J-A-1
ACGTTTGATATCCACTTTGGTCCC (+40) (IGKJ3*01) IGK-J-B-1
ACGTTTAATCTCCAGTCGTGTCCC (+40) (IGKJ5*01) IGK-J-C-1
Kde
Vκ
GCAGCTGCAGACTCATGAGGAG (+48) (IGKDE) IGK-DE-A-1
Kde
Jκ-Cκ intron
IGK-INTR-A-1 (Intron) (-111) GAGTGGCTTTGGTGGCCATGC

Fig. 5B-1

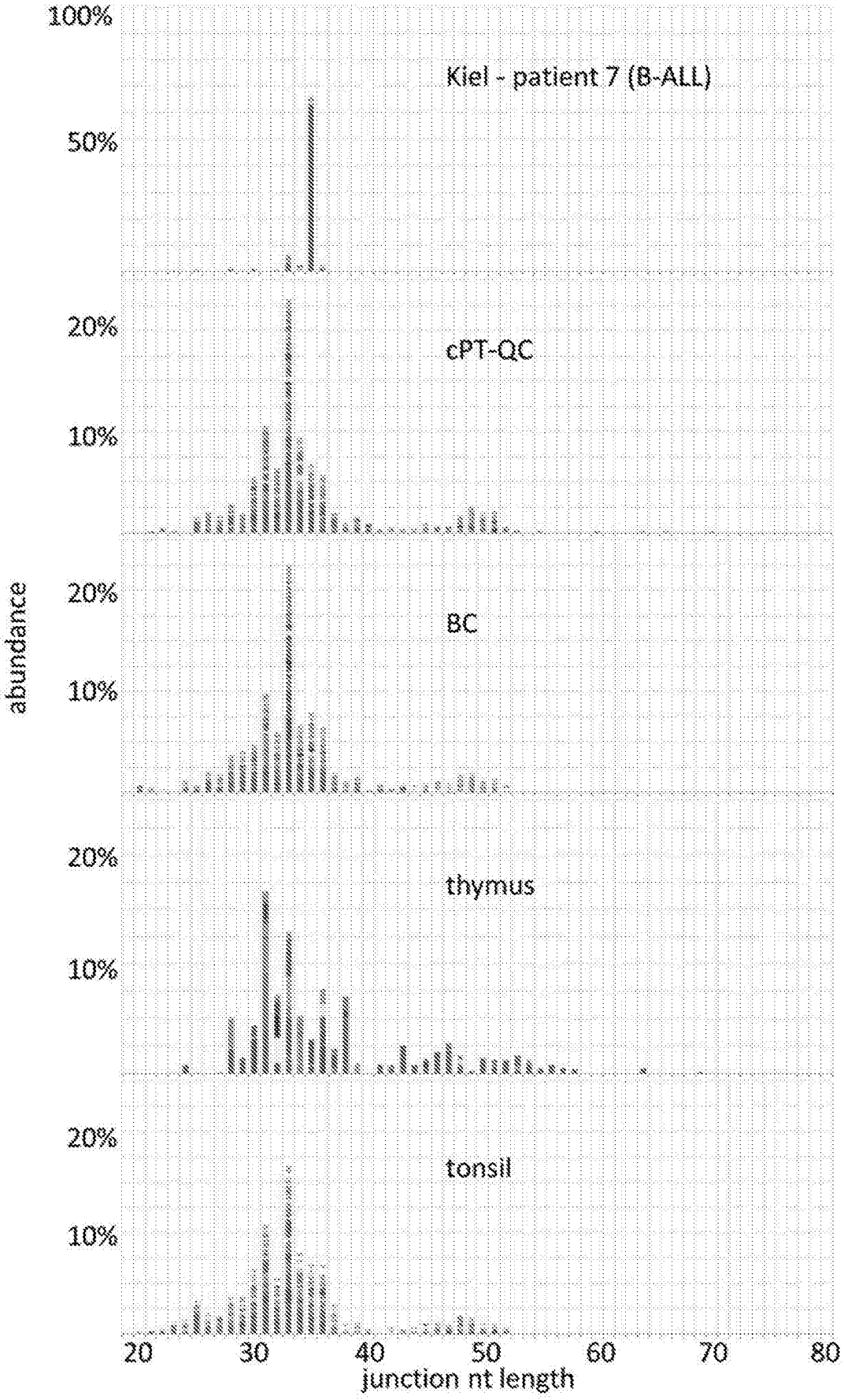

Kiel - patient 7 (B-ALL)
cPT-QC
BC
thymus
tonsil
100%
50%
20%
10%
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5B-2

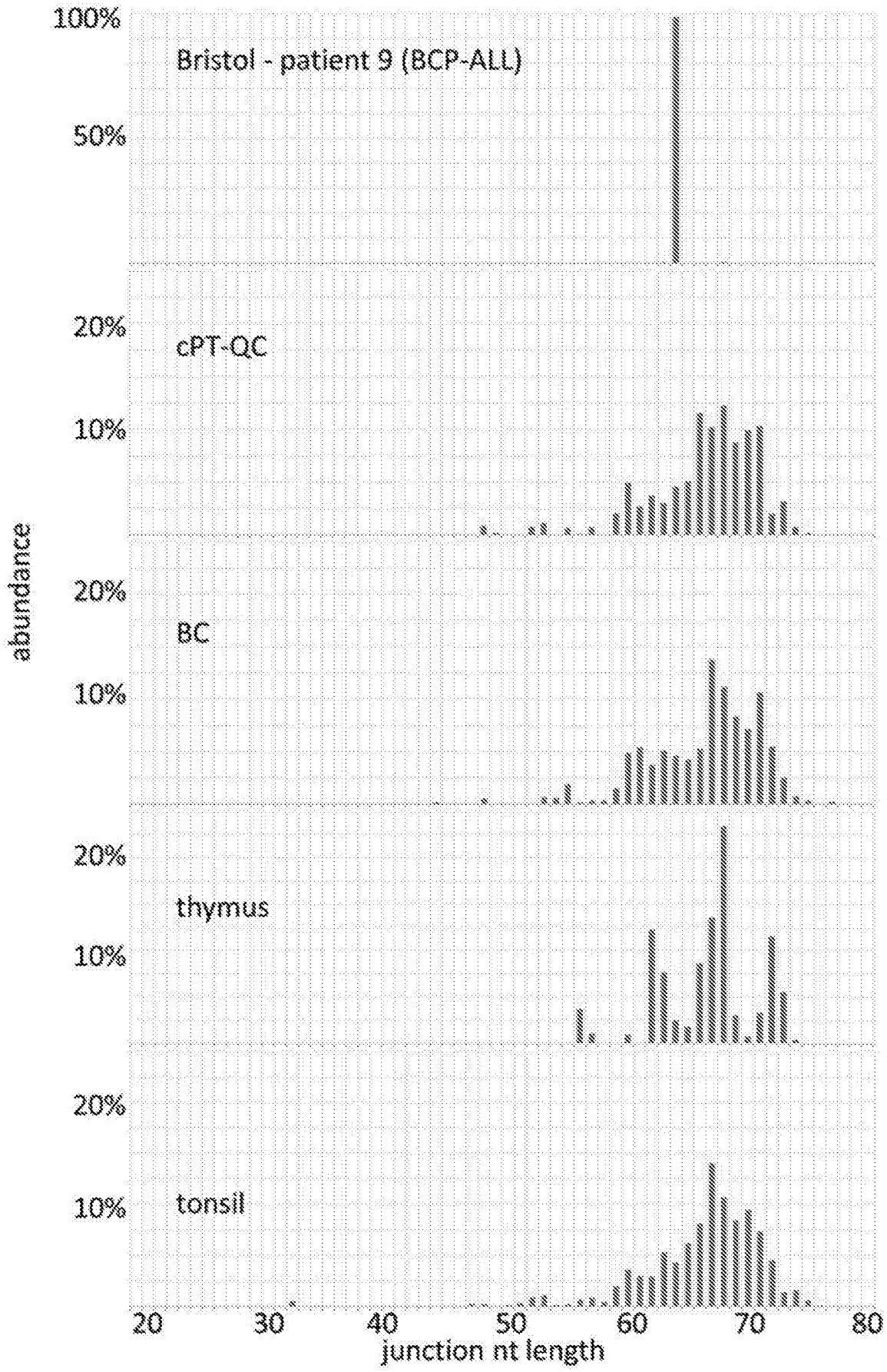

Bristol - patient 9 (BCP-ALL)
cPT-QC
BC
thymus
tonsil
100%
50%
20%
10%
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5B-3

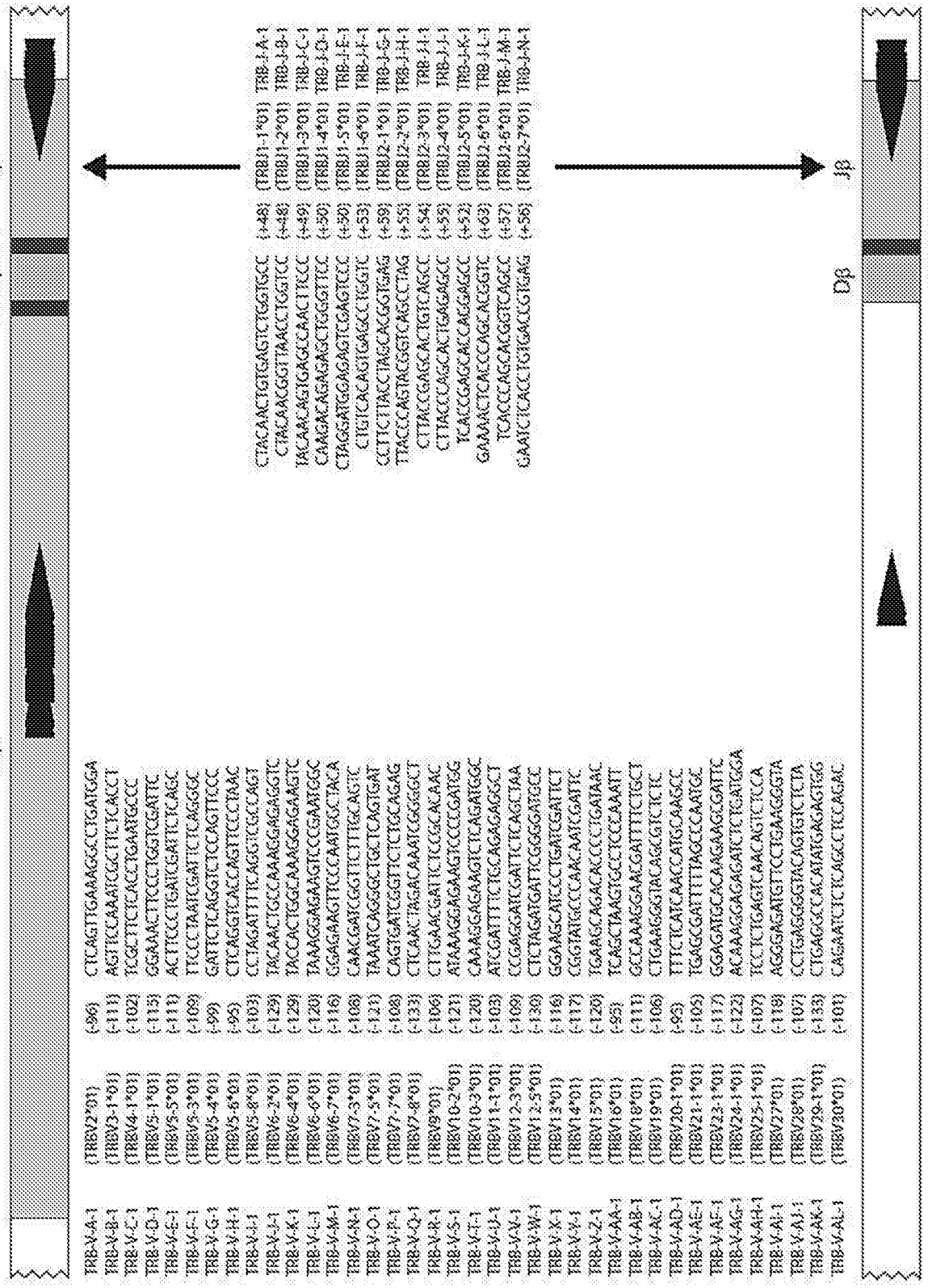

Vβ
Dβ
Jβ
TRBV2*01 (-96) CTCAGTTGAAAGGCCTGATGGA
TRBV3-1*01 (-111) AGTTCCAAATCGCTTCTCACCT
TRBV4-1*01 (-102) TCGCTTCTCACCTGAATGCCC
TRBV5-1*01 (-115) GGAAACTTCCCTGGTCGATTC
TRBV5-5*01 (-111) ACTTCCCTGATCGATTCTCAGC
TRBV5-3*01 (-109) TTCCCTAATCGATTCTCAGGGC
TRBV5-4*01 (-99) GATTCTCAGGTCTCCAGTTCCC
TRBV5-6*01 (-95) CTCAGGTCACCAGTTCCCTAAC
TRBV5-8*01 (-103) CCTAGATTTTCAGGTTCGCCAGT
TRBV6-2*01 (-129) TACAACTCGCCAAAGGAGAGGTC
TRBV6-4*01 (-129) TACCACTGGCAAAGGAGAAGTC
TRBV6-6*01 (-120) TAAAGGAGAAGTCCCGAATGGC
TRBV6-7*01 (-116) GGAGAAGTTCCCAATTCGCTACA
TRBV7-3*01 (-106) CAACGATCGGTTCTTTGCAGTC
TRBV7-5*01 (-121) TAAATCAGGGCTGCTCAGTGAT
TRBV7-7*01 (-108) CAGTGATCGGTTCTCTGCAGAG
TRBV7-8*01 (-133) CTCAACTAGACAAATCGGGGCT
TRBV9*01 (-106) CTTGAACGATTCTCCGCACAAC
TRBV10-2*01 (-121) ATAAAGGAGAAGTCCCCGATGG
TRBV10-3*01 (-120) CAAAGGAGAAGTCTCAGATGGC
TRBV11-1*01 (-103) ATCGATTTCTCGCAGAGAGGCT
TRBV11-3*01 (-109) CCGAAGGATCGATTCTCAGCTAA
TRBV12-3*01 (-130) CTCTAGATGATTCGGGGATGCC
TRBV12-5*01 (-116) GGAAGCATCCCTGATCGATTCT
TRBV13*01 (-117) CGGTATGCCCAACAATCGATTC
TRBV14*01 (-120) TGAAGCAGACACCCCTGATAAAC
TRBV15*01 (-95) TCAGCTAAGTGCCTCCCAAATT
TRBV16*01 (-111) GCCAAAGGAACGATTTTCTGCT
TRBV18*01 (-106) CTGAAGGGTACAGCGTCTCTC
TRBV19*01 (-95) TTTCTCATCAACCATGCAAGCC
TRBV20-1*01 (-105) TGAGCGATTTTTAGCCCAATGC
TRBV21-1*01 (-117) GGAGATGCACAAGAAGCGATTC
TRBV23-1*01 (-122) ACAAAGGAGAGATCTCTGATGGA
TRBV24-1*01 (-107) TCCTCTGAGTCAACAGTCTCCA
TRBV25-1*01 (-119) AGGGAGATGTTCCTGAAGGGTA
TRBV27*01 (-107) CCTGAGGGGTACAGTGTCTCTA
TRBV28*01 (-133) CTGAGGCCACATATGAGAGTGG
TRBV30*01 (-101) CAGAATCTCTCAGCCTCCAGAC
TRBJ1-1*01 (+48) CTACAACTGTGAGTCTGGTGCC
TRBJ1-2*01 (+48) CTACAACGGTTAACCTGGTCC
TRBJ1-3*01 (+49) TACAACAGTGAGCCAACTCCC
TRBJ1-4*01 (+50) CAAGACAGAGAGCTGGGTTCC
TRBJ1-5*01 (+50) CTAGGATGGAGAGTCGAGTCCC
TRBJ1-6*01 (+53) CTGTTCACAGTGAGCCTGGGTC
TRBJ2-1*01 (+59) CCTTCTTACCTAGCACGGTGAG
TRBJ2-2*01 (+55) TTACCCAGTACGGGTCAGCCTAG
TRBJ2-3*01 (+54) CTTACCGAGCACTGTCAGCC
TRBJ2-4*01 (+55) CTTACCCAGCACTGAGAGCC
TRBJ2-5*01 (+53) TCACCGAGCACCAGGAGCC
TRBJ2-6*01 (+63) GAAAACTCACCCAGCACGGTC
TRBJ2-7*01 (+57) TCACCCAGCACGGTCAGCC
TRBJ2-7*01 (+56) GAATCTCACCTGTGACCGTGAG
TRBJ1-1 TRB-J-A-1
TRBJ1-2 TRB-J-B-1
TRBJ1-3 TRB-J-C-1
TRBJ1-4 TRB-J-D-1
TRBJ1-5 TRB-J-E-1
TRBJ1-6 TRB-J-F-1
TRBJ2-1 TRB-J-G-1
TRBJ2-2 TRB-J-H-1
TRBJ2-3 TRB-J-I-1
TRBJ2-4 TRB-J-J-1
TRBJ2-5 TRB-J-K-1
TRBJ2-6 TRB-J-L-1
TRBJ2-6 TRB-J-M-1
TRBJ2-7 TRB-J-N-1
TRD-D-A-1 (TRBD1*01) (-114) CCTCCACTCCCCTCAAAGGA
TRD-D-B-1 (TRBD2*01) (-117) CAGACTAACCTCTGCCACCTG

Fig. 5C-1

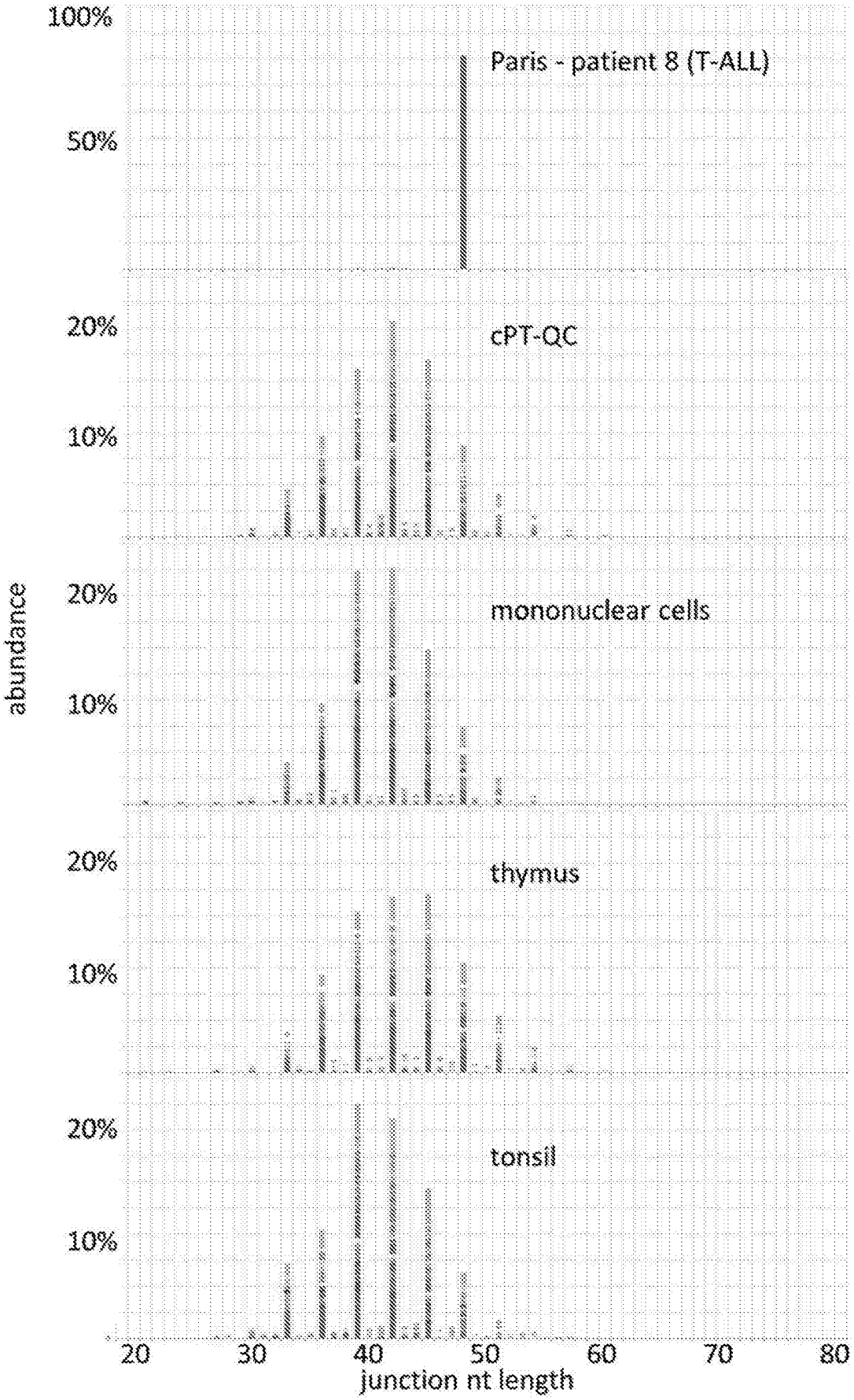

100%
50%
Paris - patient 8 (T-ALL)
20%
10%
cPT-QC
20%
10%
mononuclear cells
20%
10%
thymus
20%
10%
tonsil
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5C-2

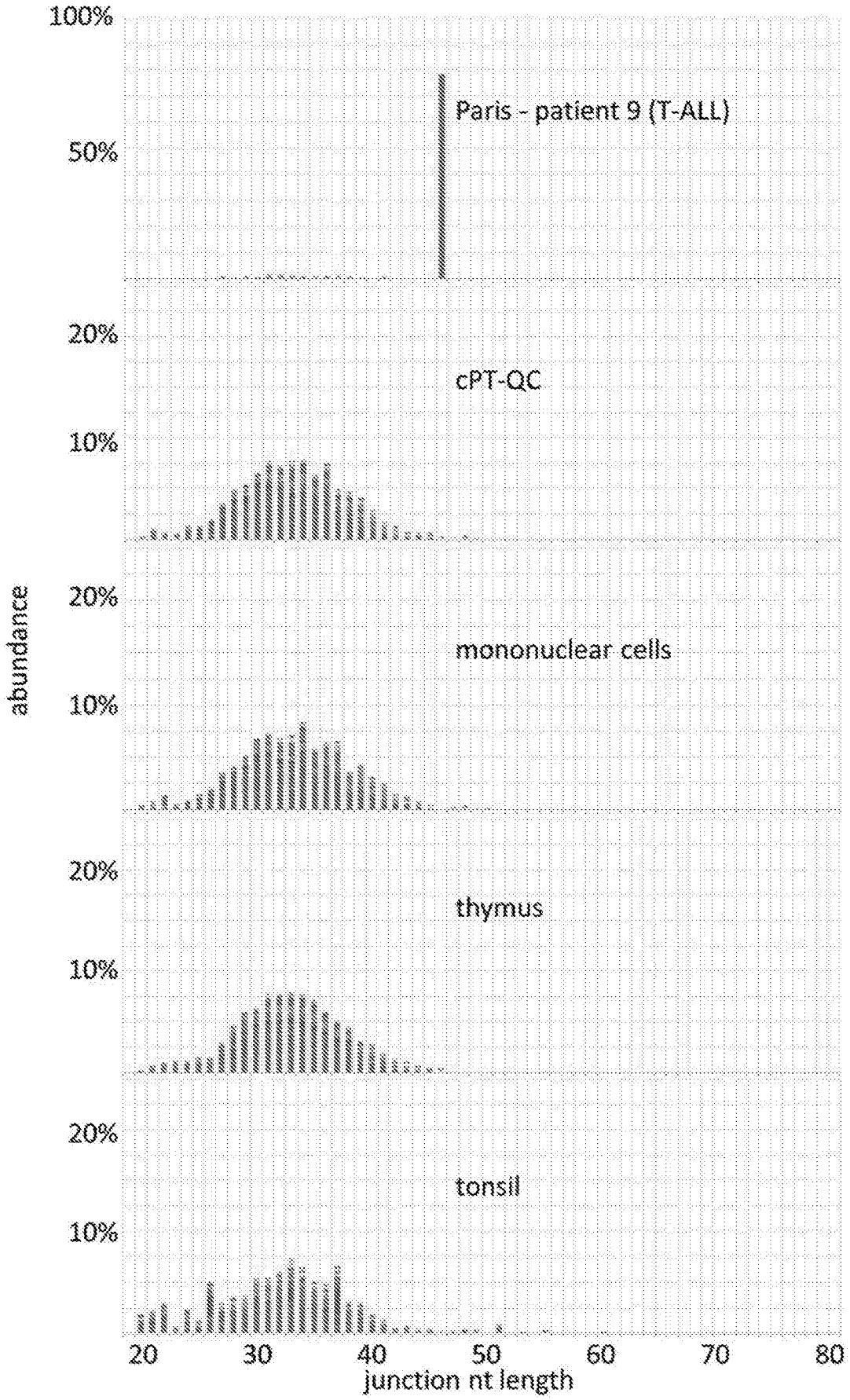

100%
50%
Paris - patient 9 (T-ALL)
20%
cPT-QC
10%
20%
mononuclear cells
10%
20%
thymus
10%
20%
tonsil
10%
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5C-3

V_γ
J_γ
TRG-V-A-1 (TRGV2*01) (-83) GCACAAGGAACAACTTGAGATTG
TRG-V-B-1 (TRGV4*01) (-88) TGGAAGCACAAGGAAGAACTTGAGAA
TRG-V-C-1 (TRGV8*01) (-83) GCACAGGGAAGAGCCTTAAATT
TRG-V-D-1 (TRGV3*01) (-82) CAGGAGGTGGAGCTGGATATT
TRG-V-E-1 (TRGV7) (-85) CAAGCATGAGGAGGAGCTGGAAATTG
TRG-V-F-1 (TRGV9*01) (-78) ACGTCTACATCCACTCTCACC
TRG-V-G-1 (TRGV10*01) (-79) CTCTCACTTCAATCCTTACCATCAA
TRG-V-H-1 (TRGV11*01) (-80) CTTCCACTTCCACTTTGAAAATAAAGT
AGTGTTGTTCCACTGCCAAAG (+39) (TRGJ1*01) TRG-J-A-1
GTTCCGGGACCAAATACCTTG (+45) (TRGJP*01) TRG-J-B-1
GAGCTTAGTCCCTTCAGCAAATA (+44) (TRGJP1*01) TRG-J-C-1
CCTAGTCCCTTTTGCAAACG (+41) (TRGJP2*01) TRG-J-D-1

Fig. 5D-1

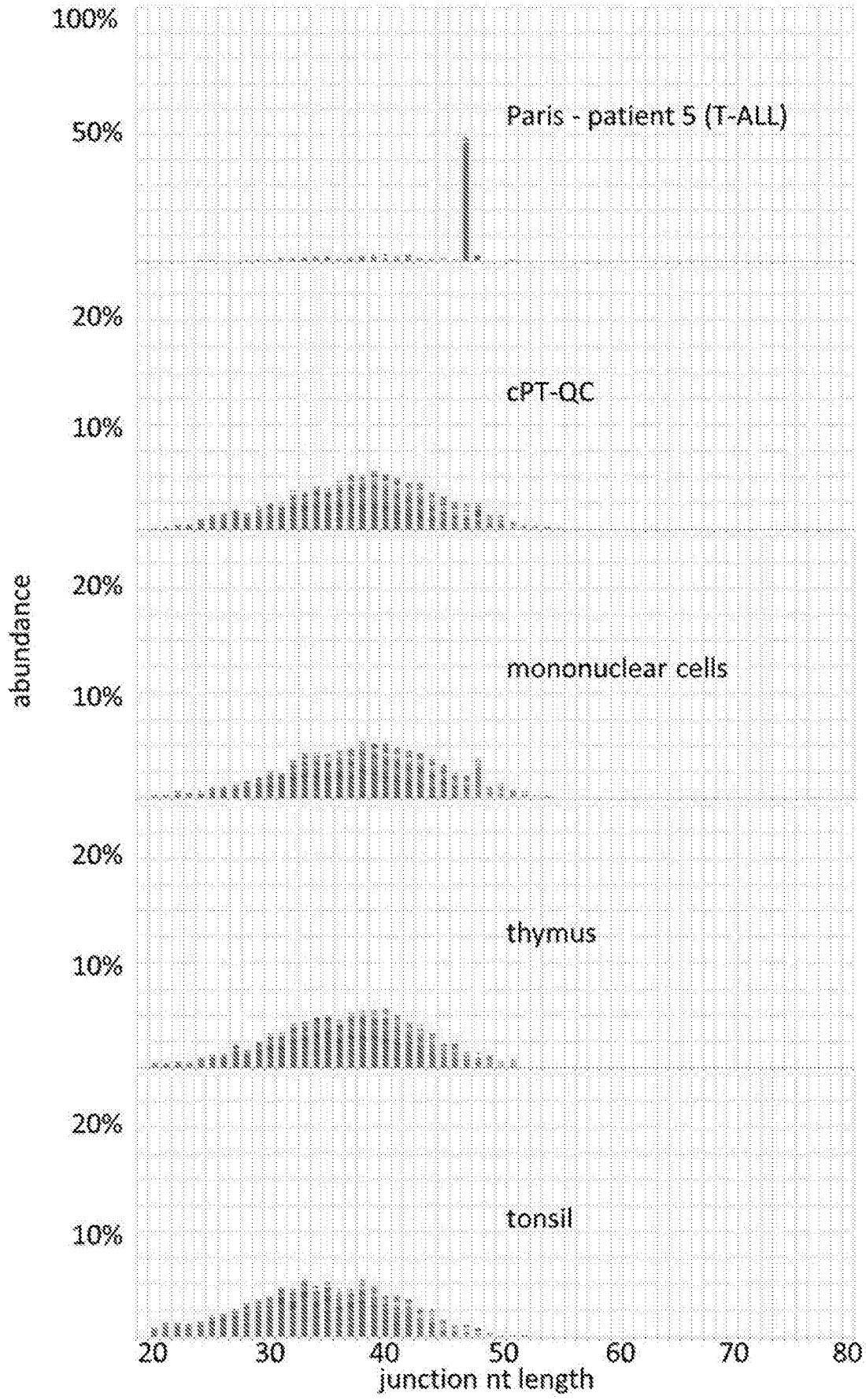

100%
50%
Paris - patient 5 (T-ALL)
20%
10%
cPT-QC
20%
10%
mononuclear cells
20%
10%
thymus
20%
10%
tonsil
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5D-2

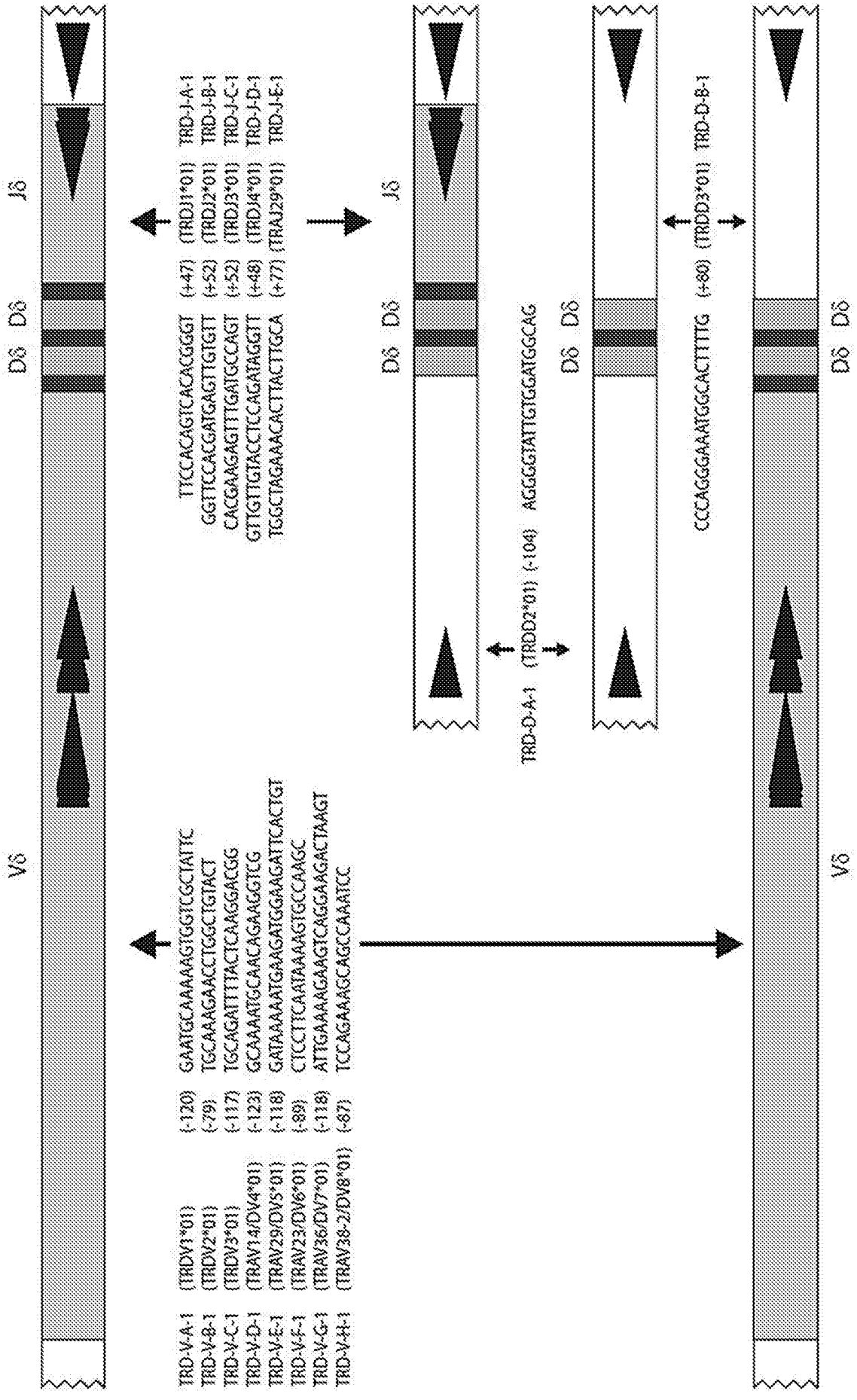

Vδ
Dδ Dδ
Jδ
TRD-V-A-1 (TRDV1*01) (-120) GAATGCAAAAAGTGGTCGCTATTC
TRD-V-B-1 (TRDV2*01) (-79) TGCAAAGAACCTGGCTGTACT
TRD-V-C-1 (TRDV3*01) (-117) TGCAGATTTTACTCAAGGACGG
TRD-V-D-1 (TRAV14/DV4*01) (-123) GCAAAATGCAACAGAAGGTCG
TRD-V-E-1 (TRAV29/DV5*01) (-118) GATAAAAATGAAGATGGAAGATTCACTGT
TRD-V-F-1 (TRAV23/DV6*01) (-89) CTCCTTCAATAAAAGTGCCAAGC
TRD-V-G-1 (TRAV36/DV7*01) (-118) ATTGAAAAGAAGTCAGGAAGACTAAGT
TRD-V-H-1 (TRAV38-2/DV8*01) (-87) TCCAGAAAGCAGCCAAATCC
TTCCACAGTCACACGGGT (+47) (TRDJ1*01) TRD-J-A-1
GGTTCCACGATGAGTTGTGTT (+52) (TRDJ2*01) TRD-J-B-1
CACGAAGAGTTTGATGCCAGT (+52) (TRDJ3*01) TRD-J-C-1
GTTGTTGTACCTCCAGATAGGTT (+48) (TRDJ4*01) TRD-J-D-1
TGGCTAGAAACACTTACTTGCA (+77) (TRAJ29*01) TRD-J-E-1
TRD-D-A-1 (TRDD2*01) (-104) AGGGGTATTGTGGATGGCAG
CCCAGGGAAATGGCACTTTTG (+80) (TRDD3*01) TRD-D-B-1

Fig. 5E-1

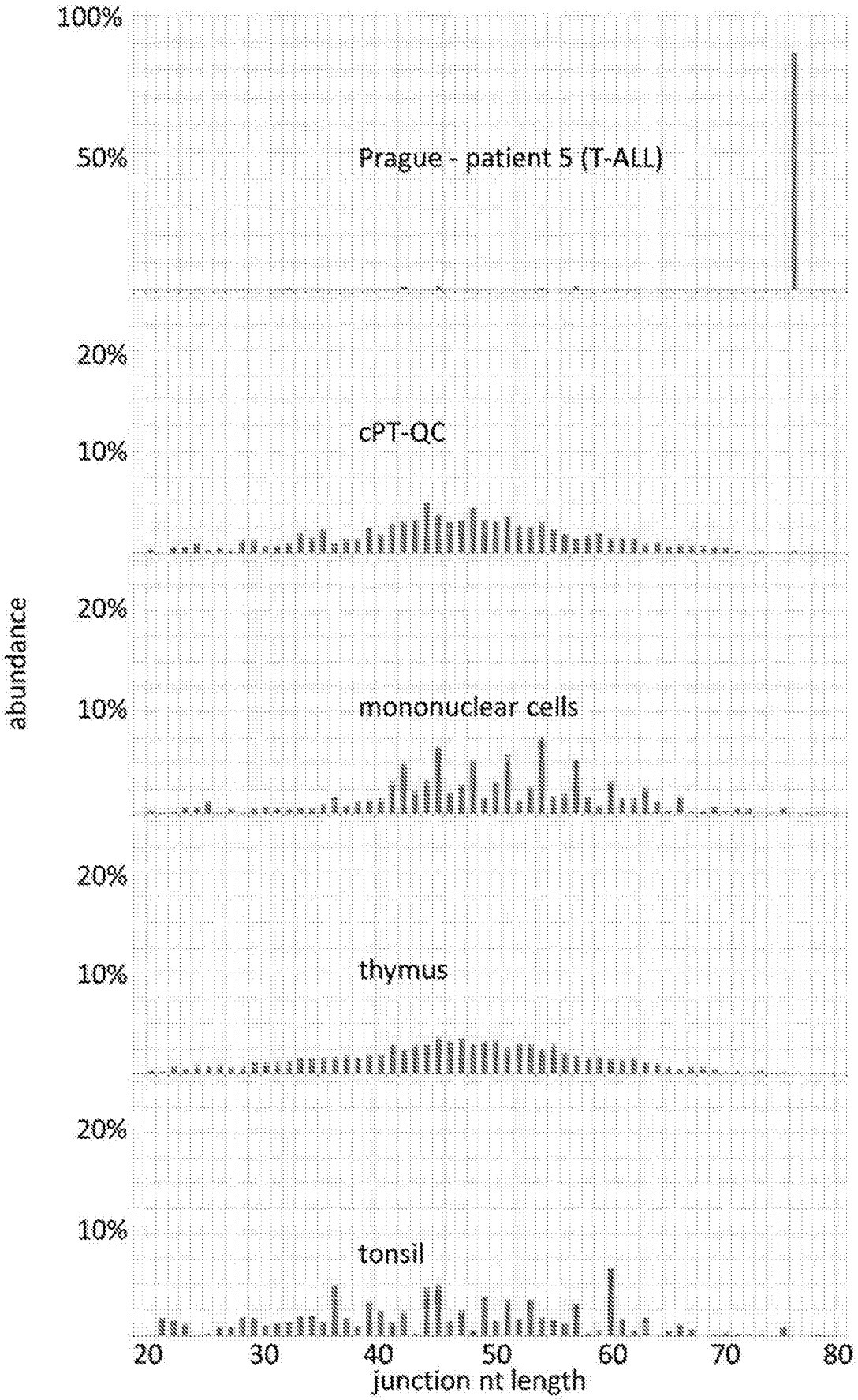

100%
50%
Prague - patient 5 (T-ALL)
20%
10%
cPT-QC
20%
10%
mononuclear cells
20%
10%
thymus
20%
10%
tonsil
abundance
20
30
40
50
60
70
80
junction nt length

Fig. 5E-2

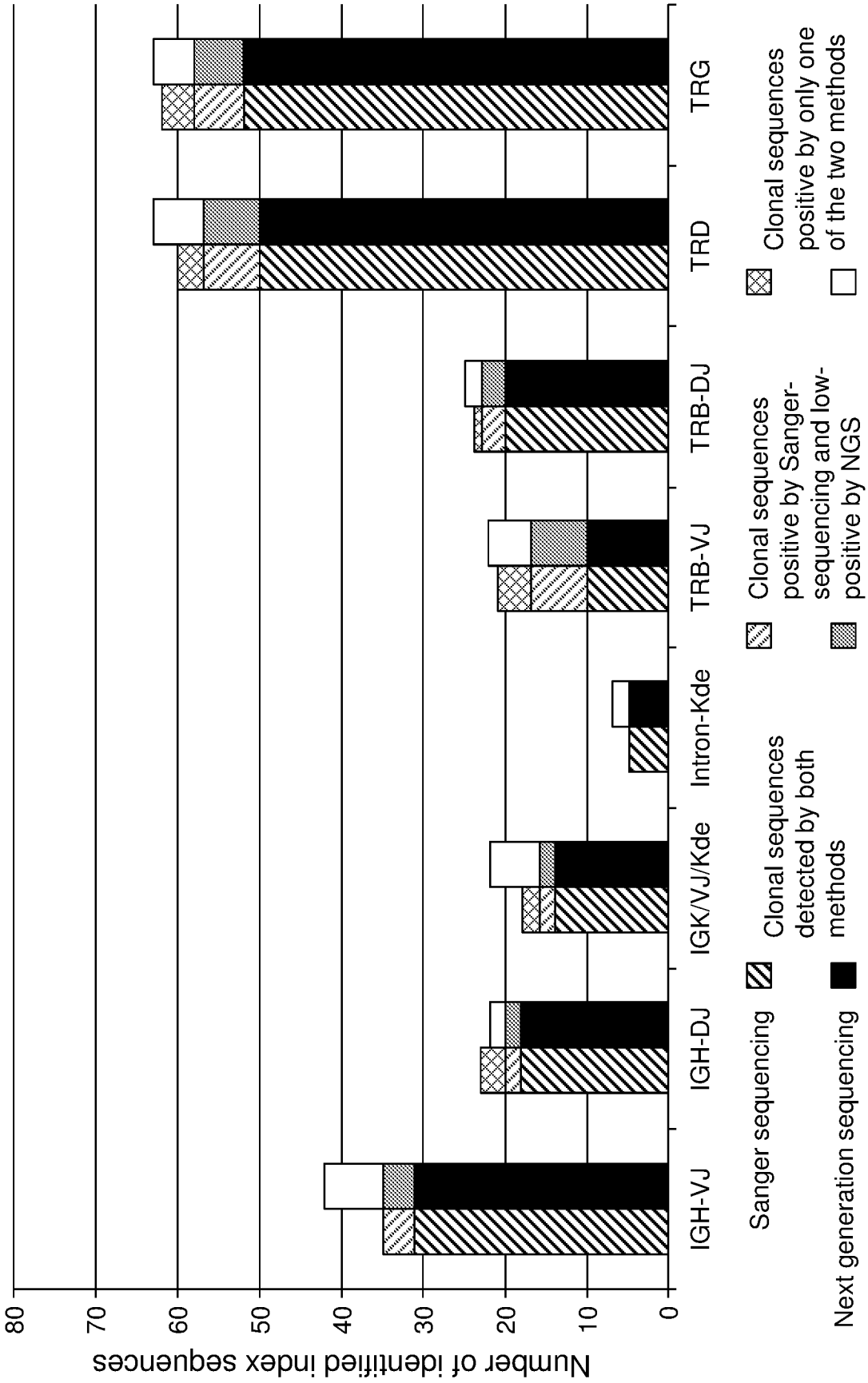

Number of identified index sequences
0
10
20
30
40
50
60
70
80
IGH-VJ
IGH-DJ
IGK/VJ/Kde
Intron-Kde
TRB-VJ
TRB-DJ
TRD
TRG
Sanger sequencing
Next generation sequencing
Clonal sequences detected by both methods
Clonal sequences positive by Sanger-sequencing and low-positive by NGS
Clonal sequences positive by only one of the two methods

Fig. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2020/050181

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6881 C12Q1/6886
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2018/208984 A1 (LOONEY TIMOTHY [US] ET AL) 26 July 2018 (2018-07-26) para. 23-25, 40-47, 339, 341, 379 -----	1-11, 14-20
A	MARCO BECCUTI ET AL: "HashClone: a new tool to quantify the minimal residual disease in B-cell lymphoma from deep sequencing data", BMC BIOINFORMATICS, BIOMED CENTRAL LTD, LONDON, UK, vol. 18, no. 1, 23 November 2017 (2017-11-23), pages 1-12, XP021250948, DOI: 10.1186/S12859-017-1923-2 abstract; p. 3, left-hand col., 5th para.; p. 6, left-hand col., 3rd para. ----- -/--	1-11, 14-20

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 June 2020	03/08/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ripaud, Leslie

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2020/050181

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Michael A Pulsipher ET AL: "IgH-V(D)J NGS-MRD measurement pre-and early post-allotransplant defines very low-and very high-risk ALL patients", , 28 May 2015 (2015-05-28), XP055614052, DOI: 10.1182/blood-2014-12- Retrieved from the Internet: URL:http://www.bloodjournal.org/content/bloodjournal/125/22/3501.full.pdf [retrieved on 2019-08-20] p. 3502, right-hand col. -----	1-11, 14-20
A	Malek Faham ET AL: "Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia", Blood, 1 January 2012 (2012-01-01), pages 5173-5180, XP055247552, DOI: 10.1182/blood-2012-07-444042 Retrieved from the Internet: URL:http://www.bloodjournal.org/content/bloodjournal/120/26/5173.full.pdf p. 5174 -----	1-11, 14-20
A	ASH A ALIZADEH ET AL: "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling", NATURE, MACMILLAN JOURNALS LTD., ETC.¦, vol. 403, no. 6769, 3 February 2000 (2000-02-03), pages 503-511, XP008127002, ISSN: 0028-0836, DOI: 10.1038/35000501 the whole document -----	1-11, 14-20
X,P	HENRIK KNECHT ET AL: "Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS", LEUKEMIA, 21 June 2019 (2019-06-21), XP055613554, London ISSN: 0887-6924, DOI: 10.1038/s41375-019-0499-4 the whole document -----	1-11, 14-20

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NL2020/050181

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-8, 16-20(completely); 9-11, 14, 15(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-8, 16-20(completely); 9-11, 14, 15(partially)

   concerns compositions comprising mixture of genomic DNA from different cells and their use in a method for detecting immunoglobulin (IG) or T-cell receptor (TR) gene rearrangements.

   ---

2. claims: 12, 13(completely); 9-11, 14, 15(partially)

   concerns a set of primers suitable for amplicon-based next-generation sequencing of IG/TR gene rearrangements and its use in a method for detecting said rearrangements.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/NL2020/050181

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018208984 A1	26-07-2018	CN 110249060 A	17-09-2019
		EP 3571320 A2	27-11-2019
		US 2018208984 A1	26-07-2018
		WO 2018136562 A2	26-07-2018