SUPPLEMENTARY APPENDIX

Antigen receptor sequencing of paired bone marrow samples shows homogeneous distribution of acute lymphoblastic leukemia subclones

Prisca M.J. Theunissen,[1] David van Zessen,[1,2] Andrew P. Stubbs,[2] Malek Faham,[3] Christian M. Zwaan,[4] Jacques J.M. van Dongen[1] and Vincent H.J. Van Der Velden[1]

[1]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, the Netherlands; [2]Department of Bioinformatics, Erasmus MC, University Medical Center Rotterdam, the Netherlands; [3]Adaptive Biotechnologies Corp., South San Francisco, CA, USA and [4]Department of Pediatric Oncology, Sophia Children's Hospital/Erasmus MC, University Medical Center Rotterdam, the Netherlands

 doi:10.3324/haematol. 2017.171454

Received: April 21, 2017.

Accepted: August 24, 2017.

Pre-published: August 31, 2017.

Correspondence: v.h.j.vandervelden@erasmusmc.nl

SUPPLEMENT

Antigen receptor sequencing of paired bone marrow samples shows homogeneous distribution of acute lymphoblastic leukemia subclones

Prisca M.J. Theunissen[1], David van Zessen[1, 2], Andrew P. Stubbs[2], Malek Faham[3], Michel Zwaan[4], Jacques J.M. van Dongen[1,$] and Vincent H.J. Van Der Velden[1]

[1]Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [2]Department of Bioinformatics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [3]Adaptive Biotechnologies Corp, South San Francisco, CA; [4]Department of Pediatric Oncology, Sophia Children's Hospital/Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands.

[$]Present address: department of Immunohematology & Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands

Correspondence: Vincent H.J. Van Der Velden, Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands; Tel: +31 107044253; E-mail: v.h.j.vandervelden@erasmusmc.nl

SUPPLEMENTARY METHODS

Estimation of degree of blood dilution for the bone marrow samples

As shown in Supplementary Table 1, the level of the leukemia in the blood is typically much lower than that in the bone marrow. Therefore generally the majority of the sequences detected in the bone marrow is likely to originate from the bone marrow. To estimate the degree of peripheral blood contamination for the bone marrow samples, we analyzed the percentage of T-cells in the bone marrow and peripheral blood samples using flow cytometry. The median %T cells in BM was 3% (range: 0,1%-8%), whereas the median %T cells in PB was 19% (range: 13-68%). The median ratio between the %T cells in BM and the %T cells in PB was 0.07 (range: 0.007-0.44), suggesting that blood dilution generally was less than 10%.

Calculation of absolute read count and frequency

To determine the absolute read count of a sequence, i.e. the read count of a sequence corrected for PCR amplification, a known number of reference sequences (spikes) was added to each PCR reaction (i.e. per locus). These spike, three plasmids with known sequences, were added to the original genomic DNA (not the library). This resulted in a factor of difference between the number of reference sequences that was added and the read count after sequencing. Per gene rearrangement type, this factor was used to correct the read count of each patient sequence, resulting in an absolute read count. [1] Each absolute read count therefore is assumed to be derived from one cell. To determine the frequency of a sequence, the individual read count of that sequence was divided by the total read count for that gene rearrangement type in a sample. However, to calculate the frequency of a sequence from the IGK locus or TRD locus, the individual read count of the sequence was divided by the sum of the Vκ-Jκ, Vκ-KDE and INTRSS-KDE total read counts and the sum of the Vδ-Dδ and Dδ-Dδ total read counts, respectively.

Frequencies of IG/TR gene rearrangements in NGS versus SB or PCR

PCR and SB determine the frequency of leukemic rearrangements within all IG and TR genes of the tested MNC's, including germline (non-rearranged) IG and TR genes. In contrast, NGS determines the frequency of a leukemic rearrangement within the rearranged genes of the same locus, without measurement of the germline genes. The number of non-leukemic MNC's is often higher in PB than in BM (in other words the tumor load in BM is generally higher than in PB). Consequently, the frequency of a leukemic rearrangement as determined by PCR or SB is often lower in PB than in BM, whereas using NGS, the frequency is generally similar between PB and BM.

PRISCA tool in Galaxy

As input file, a list of all unique sequences followed by the associated sample ID, patient ID, gene rearrangement type, frequency and absolute read count was used. During a step-wise bioinformatics process, sequences that were identical between the paired/triplicate samples were linked. Importantly,

sequences were defined as identical if, after trimming for primer and adapter sequences, all nucleotides were similar and thus showing 100% homology. Only a difference in length of the sequence was accepted (e.g. a sequence which aligns 100% to another sequence, but is two nucleotides shorter, is considered identical). The graphical output from this tool displays the frequency of each sequence and the presence in only one or more samples. PRISCA was developed in R (version 3.1.2) and on Galaxy (version v15.10). The PRISCA tool is freely available for use at https://bioinf-galaxian.erasmusmc.nl/galaxy without the need to login and it can be installed on an existing Galaxy server through the Galaxy toolshed [2].

NGS clonal variant comparison pipeline and data analysis

Background read count thresholds were determined based on the maximum absolute read count of normal lymphocyte-derived rearrangements in regenerating BM and normal BM samples. Since read counts of normal lymphocyte-derived rearrangements are expected to be lower in diagnosis BM than in regenerating BM or normal BM (due to overgrow of the leukemic population), these thresholds can be reliably applied to BCP-ALL BM samples at diagnosis. In order to use the thresholds in the analysis of BCP-ALL BM samples at diagnosis, the thresholds were expressed as a frequency of the total reads of the involved IG/TR gene rearrangement.

Analysis criteria

The IG/TR gene rearrangements at diagnosis were analyzed according to the following criteria: First, rearrangements were not further analyzed in case of a total read count of less than 1,000 (since no IG or TR gene rearrangements with a frequency above the background-threshold were present in these cases). Second, rearrangements were classified as ‘paired’ when present in both corresponding (BM-BM or BM-PB) samples. Third, a rearrangement was regarded as leukemia-derived, if the frequency of the rearrangement or the frequency of at least one of the paired rearrangements was higher than the threshold. Fourth, a leukemia-derived rearrangement was considered to be derived from an index clone if the frequency was >5% and from a subclone if the frequency was <5% [1]. Fifth, a patient was considered oligoclonal if: 1. More than two leukemic rearrangements of the same rearrangement type were found, or 2. Two leukemic rearrangements of the same rearrangement type were found and their frequency differed with more than a factor 5 (to take variation due to PCR amplification into account).

SUPPLEMENTAL RESULTS

We assessed the frequency of leukemic sequences detected in the bone marrow samples of the leukemia patients (and having a normalized read count above the defined thresholds) in each of the normal/regenerating control samples as a way to estimate the specificity of those sequences for leukemia. In total, 12 leukemic rearrangements were observed in one or more control samples, with a median absolute read count of 3 (range: 0-17). Data are summarized in **Supplementary Table 2**.

SUPPLEMENTARY TABLE

Supplementary Table 1. Patient and sample characteristics

	ID number	Gender	Age at diagnosis	Immunophenotypic classification	Cytogenetic aberrancy	Tumor load BM[a]	Tumor load PB[a]	WBC (10^9/L)
Left BM – Right BM	15760	m	12	common-ALL	iAMP	99%	80%	18.6
	15773	f	9	common-ALL	ETV6-RUNX1	91%	50%	7.2
	15803	m	1	pre-B-ALL	TCF3-PBX1	85%	64%	6.8
	16079	m	8	common-ALL	ETV6-RUNX1 hyperdiploid	77%	2%	4.7
	16261	f	3	common-ALL	ETV6-RUNX1	96%	2%	2.9
	16278	f	3	common-ALL	-	84%	35%	6.3
	16300	f	4	common-ALL	ETV6-RUNX1	93%	73%	49.9
BM - PB	10147	f	18	common-ALL	-	65%	41%	NA
	14008	f	7	common-ALL	hyperdiploid	97%	24%	4.0
	14803	f	4	pre-B-ALL	hyperdiploid	56%	40%	4.2
	15507	f	0	pro-B-ALL	MLL-ENL	100%	99%	0.8
	9781	f	5	common-ALL	hyperdiploid	87%	5%	NA

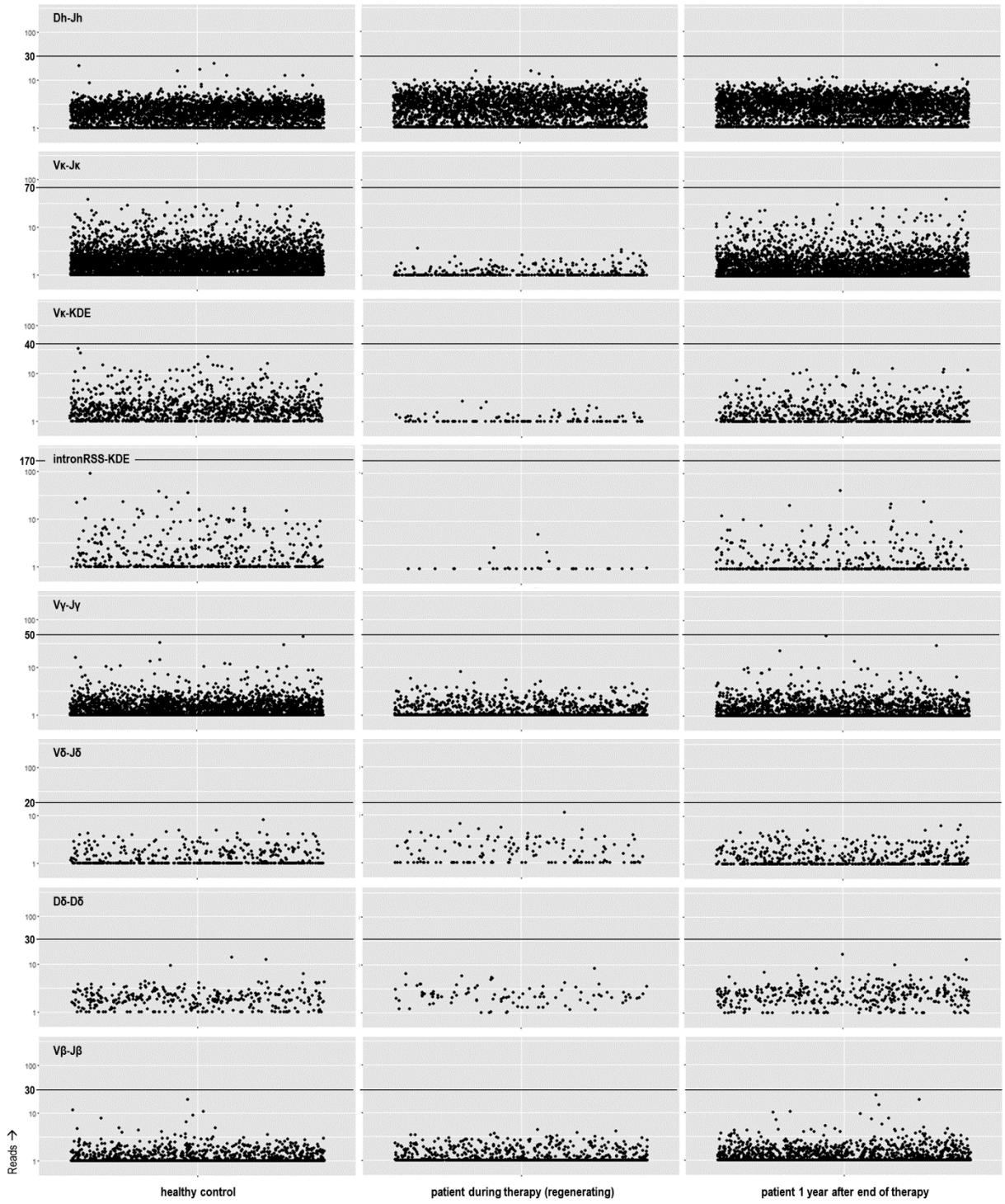
[a] Tumor load was determined by flowcytometric immunophenotyping. NA=not available.

Supplementary Table 2. Frequencies of leukemic sequences in normal/ regenerating bone marrow

Patient	V gene	J gene	BM left		BM right		read counts observed in control samples								
			reads	frequency	reads	frequen	ctrl 1	ctrl 2	ctrl 3	ctrl 4	ctrl 5	ctrl 6	ctrl 7.1	ctrl 7.2	ctrl 7.3
15803	TRBV20-1	TRBJ1-2	96,0	4,57	41,4	3,18	5,1	8,7	2,9	4,3	4,1	4,9	6,0	6,6	8,1
16079	TRBV27	TRBJ2-4	147,7	2,02	115,2	1,31	4,2	4,5	3,3	3,0	2,9	3,9	2,1	6,8	7,0
15803	TRDV2	TRDD3	2895,6	37,81	3461,5	37,59		3,3	16,7				6,0		
16079	TRGV9	TRGJ1	321,5	1,06	209,4	0,59			0,8				1,1	1,2	
16079	TRGV8	TRGJ1	68,7	0,23	54,2	0,15		1,3		0,9					
15773	IGHV2-70	IGHJ5	563,7	0,78	182,7	0,21				2,5					
15773	TRGV3	TRGJ2	18171,2	46,26	17570,0	47,09		0,6							
16079	TRBV12-3	TRBJ2-7	4773,2	65,18	5480,0	62,34					0,2				
16079	TRBV27	TRBJ2-3	2083,3	28,44	2622,5	29,85					0,1				
16079	TRGV9	TRGJ1	50,3	0,17	139,2	0,39							0,7		
16079	TRGV9	TRGJ2	112,6	0,37	168,1	0,47							0,5		
16261	IGKV2-30	KDEA1	44,0	0,22	349,4	1,26			0,7						

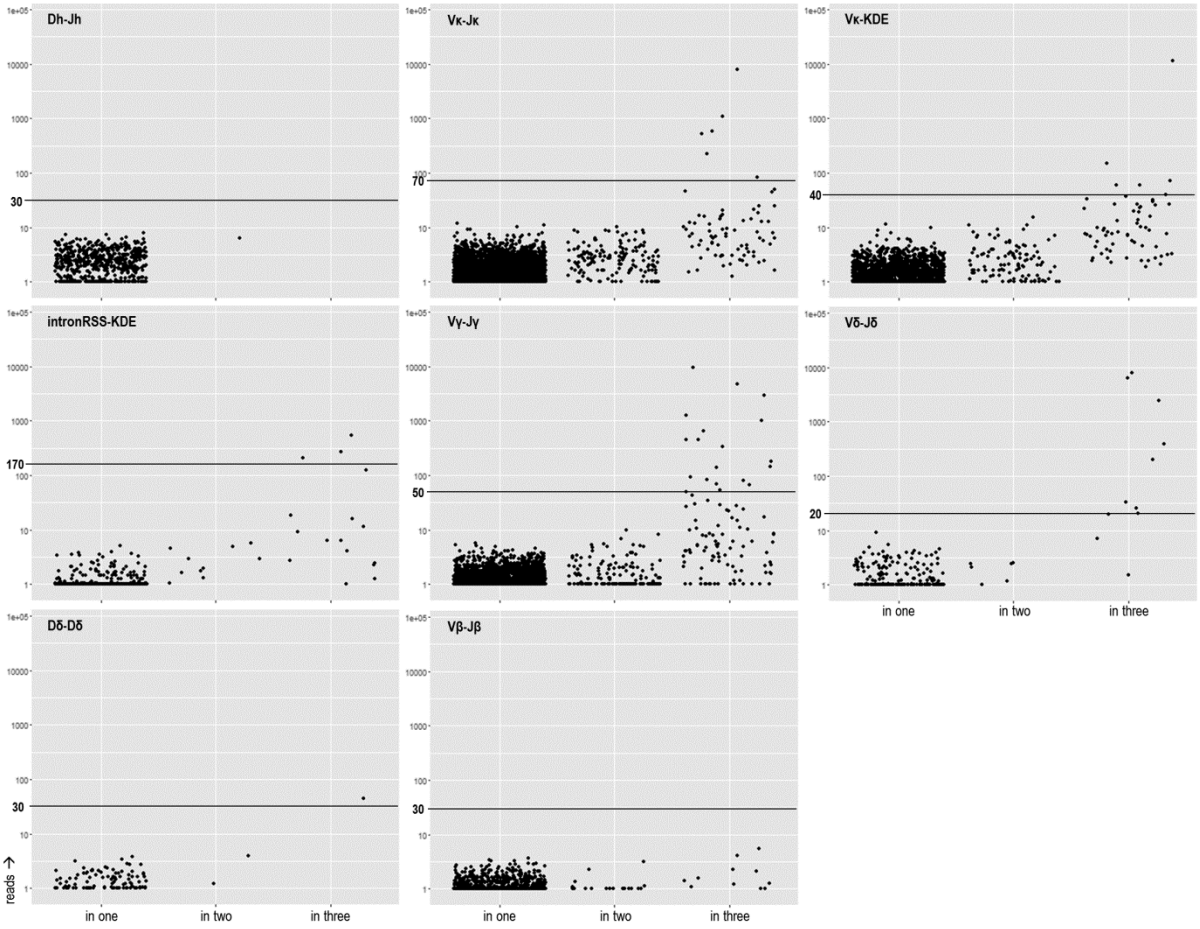
Ctrl= control; ctrl7.1, 7.2 and 7.3 are samples from the same control subject, amplified and sequences in three independent runs.

SUPPLEMENTARY FIGURES

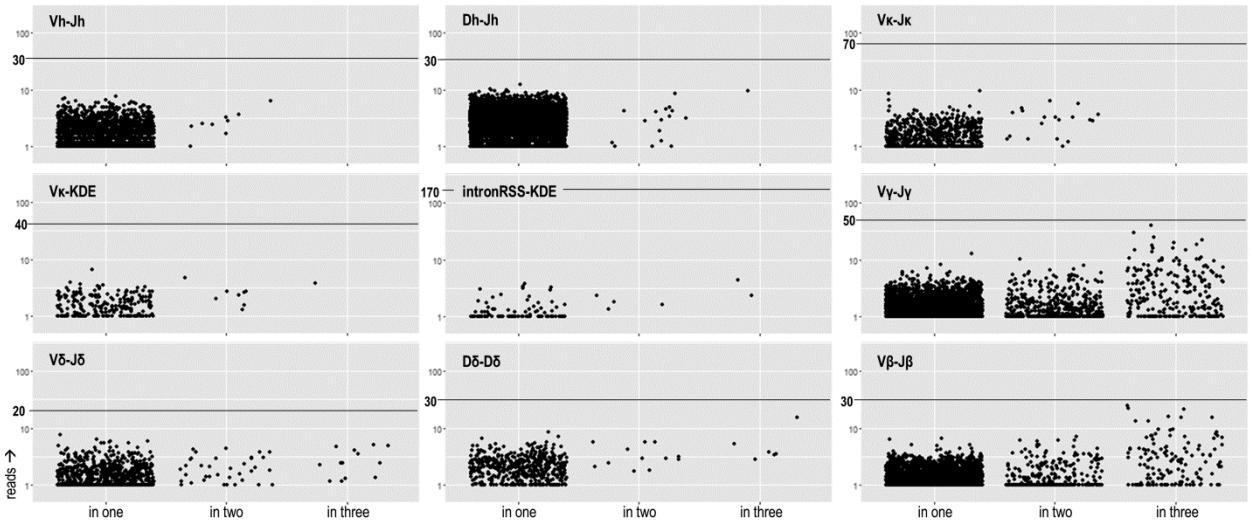


Supplementary Figure 1 (see Legend on next page)

Supplementary Figure 1. Read counts of Dh-Jh, Vκ-Jκ, Vκ-KDE, intronRSS-KDE, Vγ-Jγ, Vδ-Dδ, Dδ-Dδ and Vβ-Jβ rearrangements in BM from a healthy child, BM from a representative T-ALL patient at a therapyinterval (regenerating BM; n=5) and BM from a representative BCP-ALL patient at one year after end of therapy (n=2). Rearrangements with a read count of <1 (caused by amplification-correction) were displayed as rearrangements with a read count of 1. The black line indicates the threshold for the involved type of rearrangement.

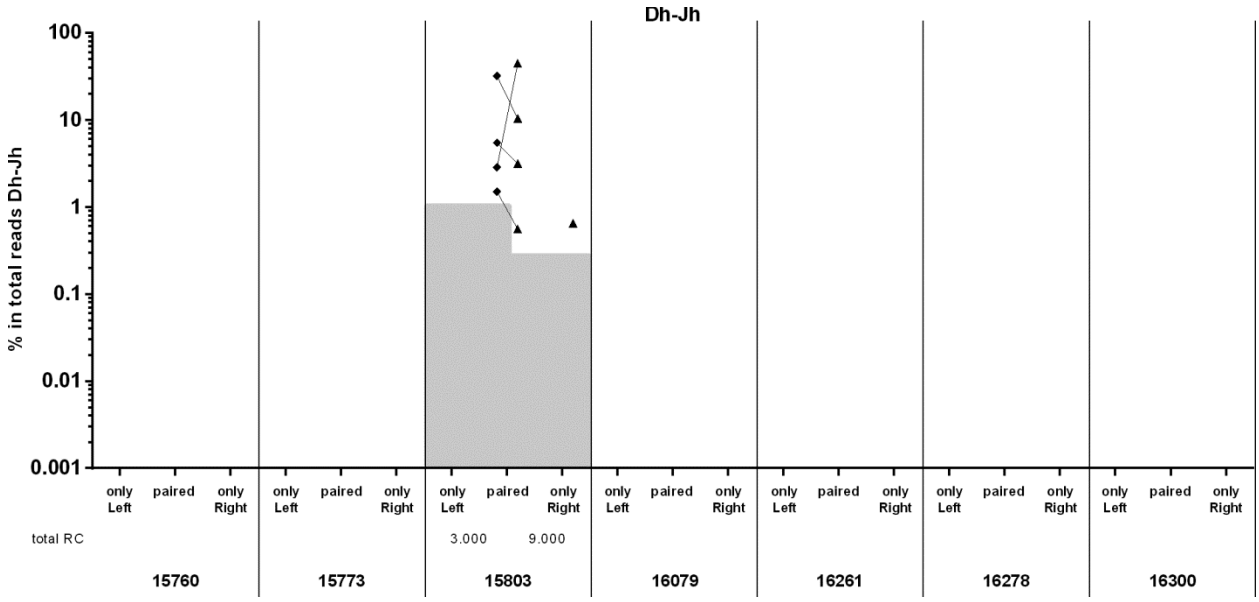


Supplementary Figure 2. Read counts of Dh-Jh, Vκ-Jκ, Vκ-KDE, intronRSS-KDE, Vγ-Jγ, Vδ-Dδ, Dδ-Dδ and Vβ-Jβ rearrangements in BM from a BCP-ALL patient at diagnosis. This BM was sequenced in triplicate, resulting in rearrangements that were found in one, in two or in all three of the replicates. Rearrangements with a read count of <1 (caused by amplification-correction) were displayed as rearrangements with a read count of 1. The black line indicates the previously defined threshold for the involved type of rearrangement.

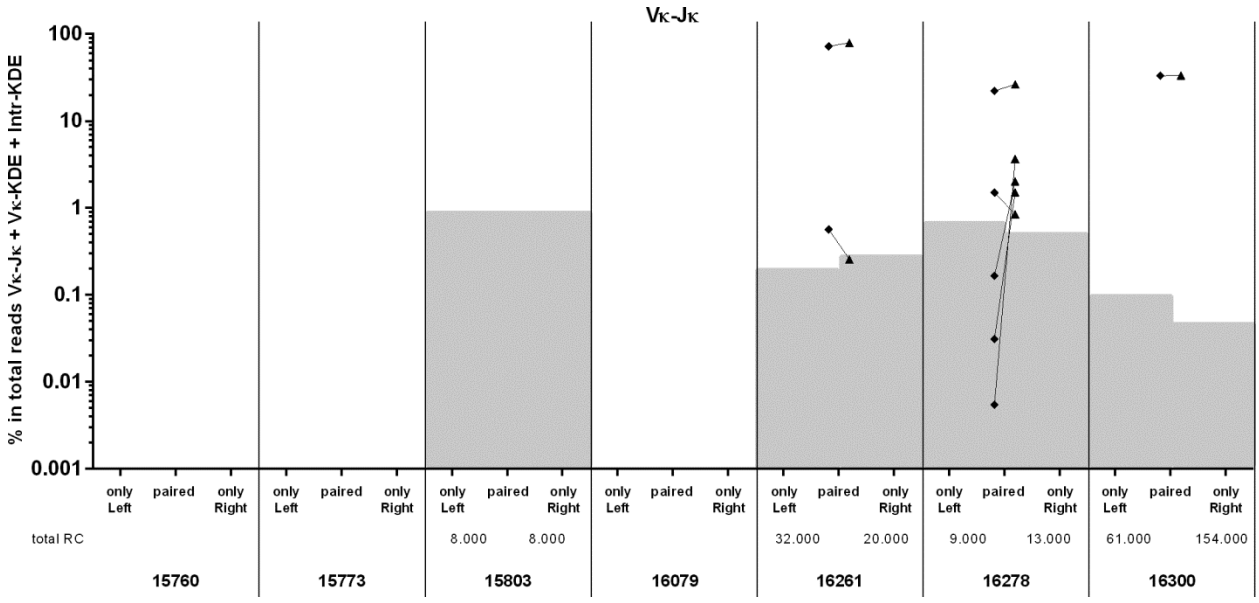


Supplementary Figure 3. Read counts of Vh-DJh, Dh-Jh, Vκ-Jκ, Vκ-KDE, intronRSS-KDE, Vγ-Jγ, Vδ-Dδ, Dδ-Dδ and Vβ-Jβ rearrangements in regenerating BM from a T-ALL patient at a therapy interval. This BM was sequenced in triplicate, resulting in rearrangements that were found in one, in two or in all three of the replicates. Rearrangements with a read count of <1 (caused by amplification-correction), were displayed as rearrangements with a read count of 1. The black line indicates the previously defined threshold for the involved type of rearrangement.

A.

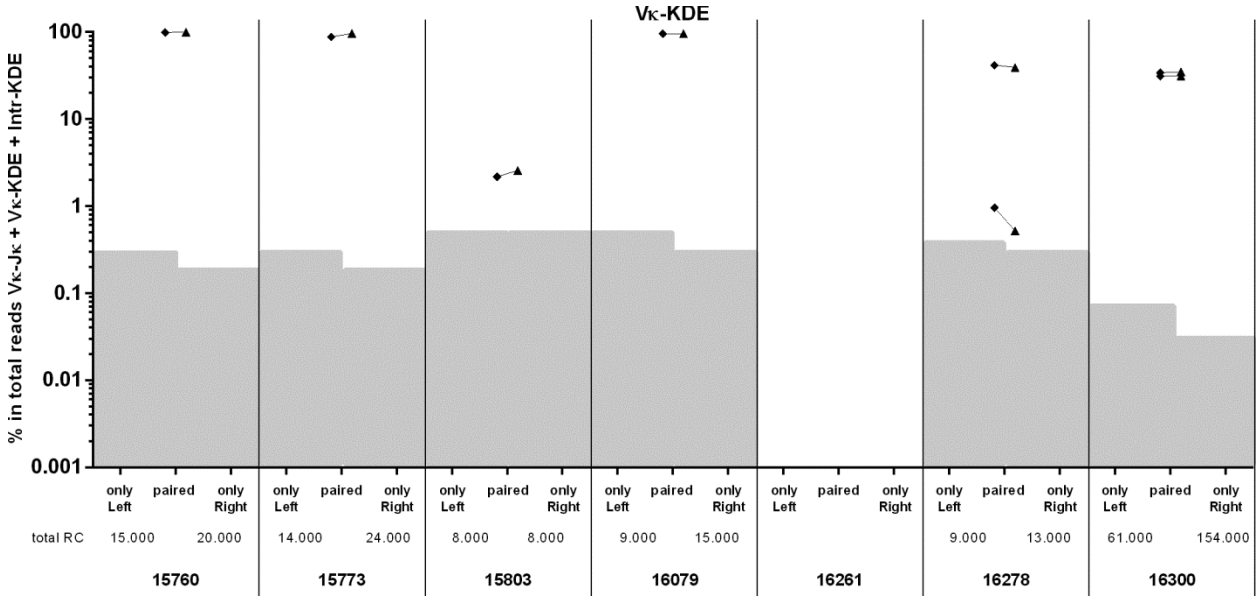


B.

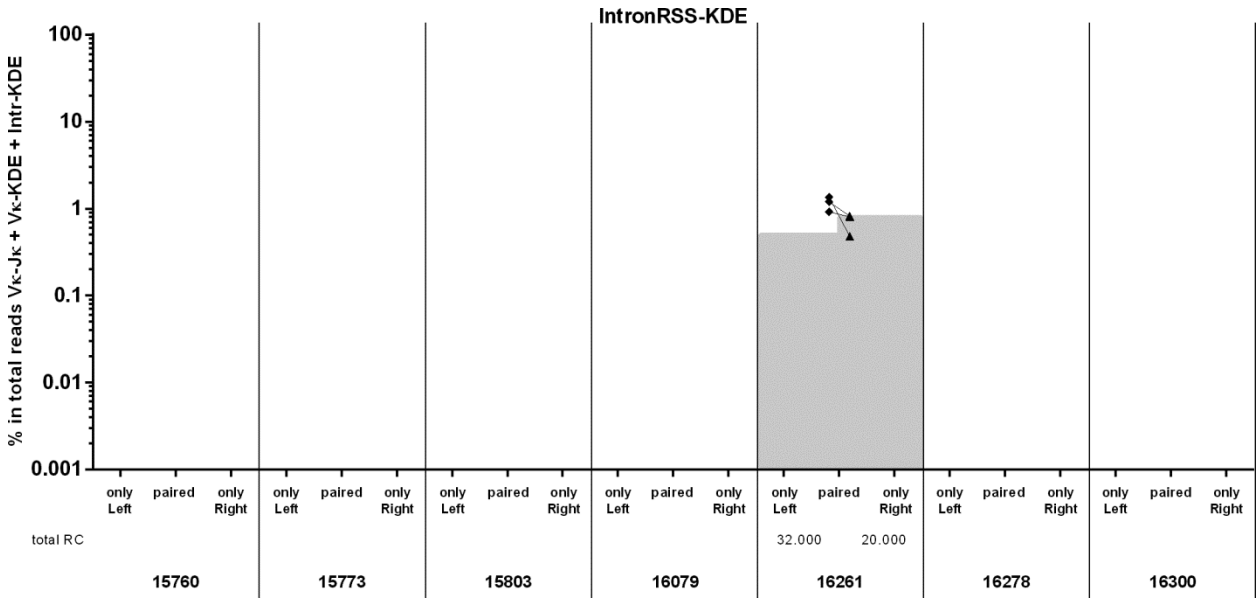


Supplementary Figure 4 (see legend on page 12)

C.

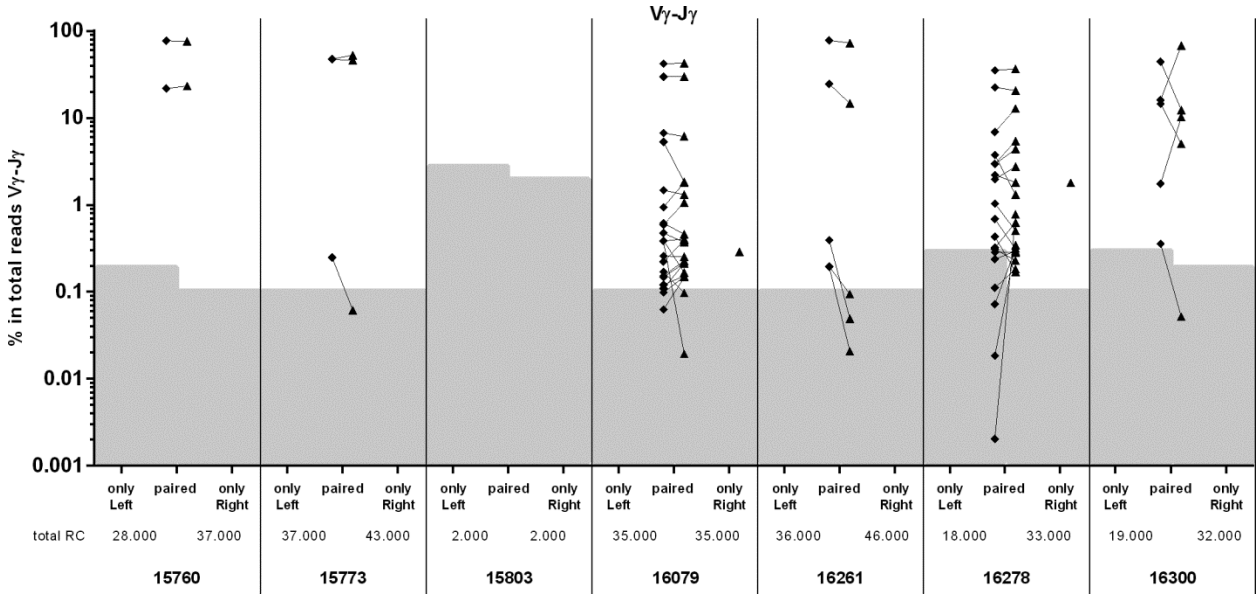


D.

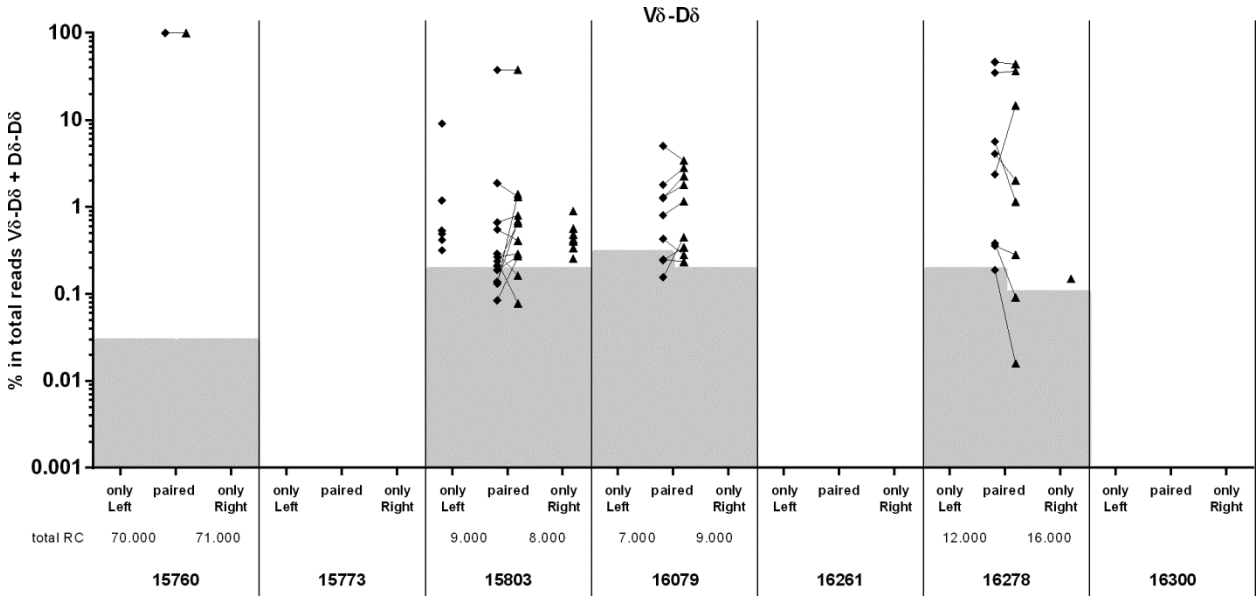


Supplementary Figure 4 (see legend on page 12)

E.

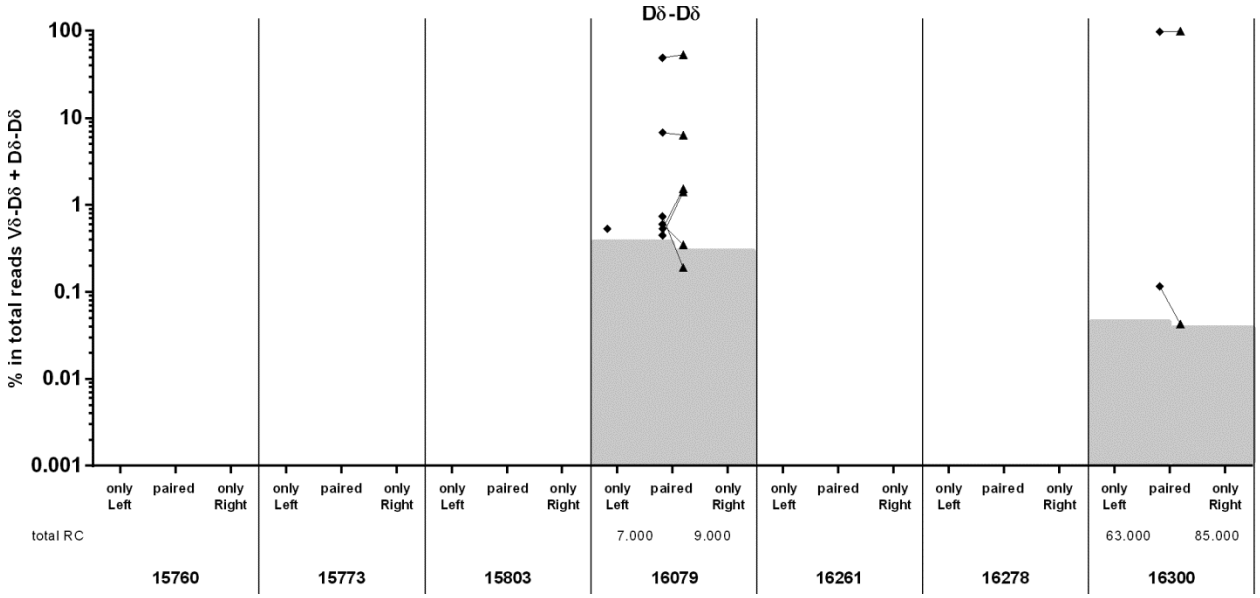


F.

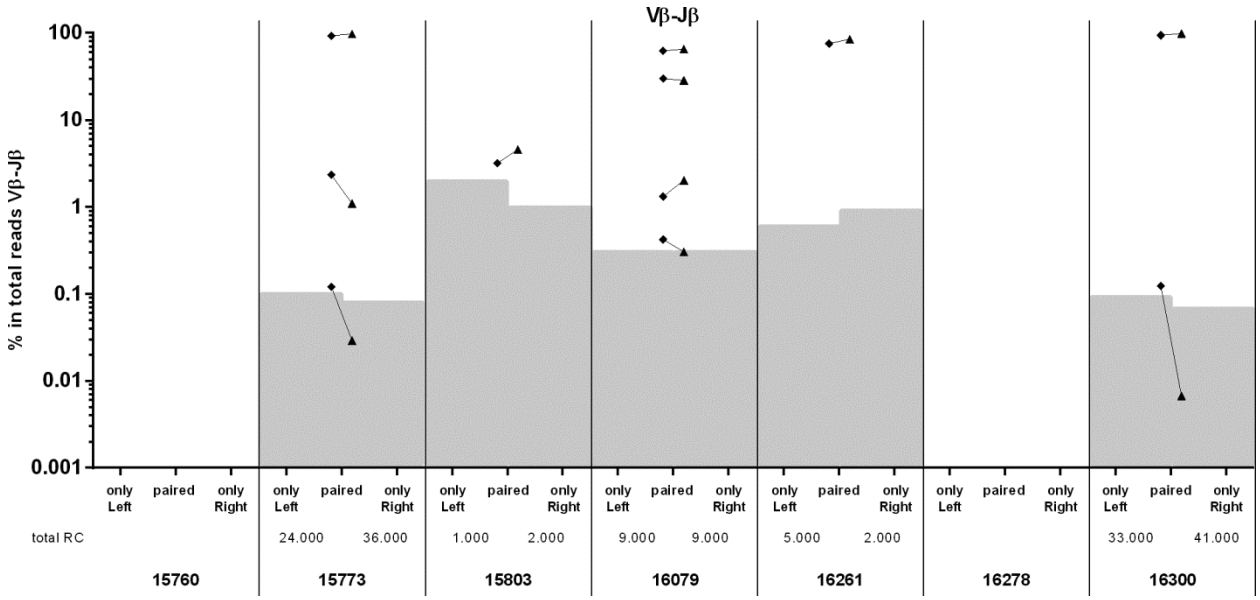


Supplementary Figure 4 (see legend on page 12)

G.

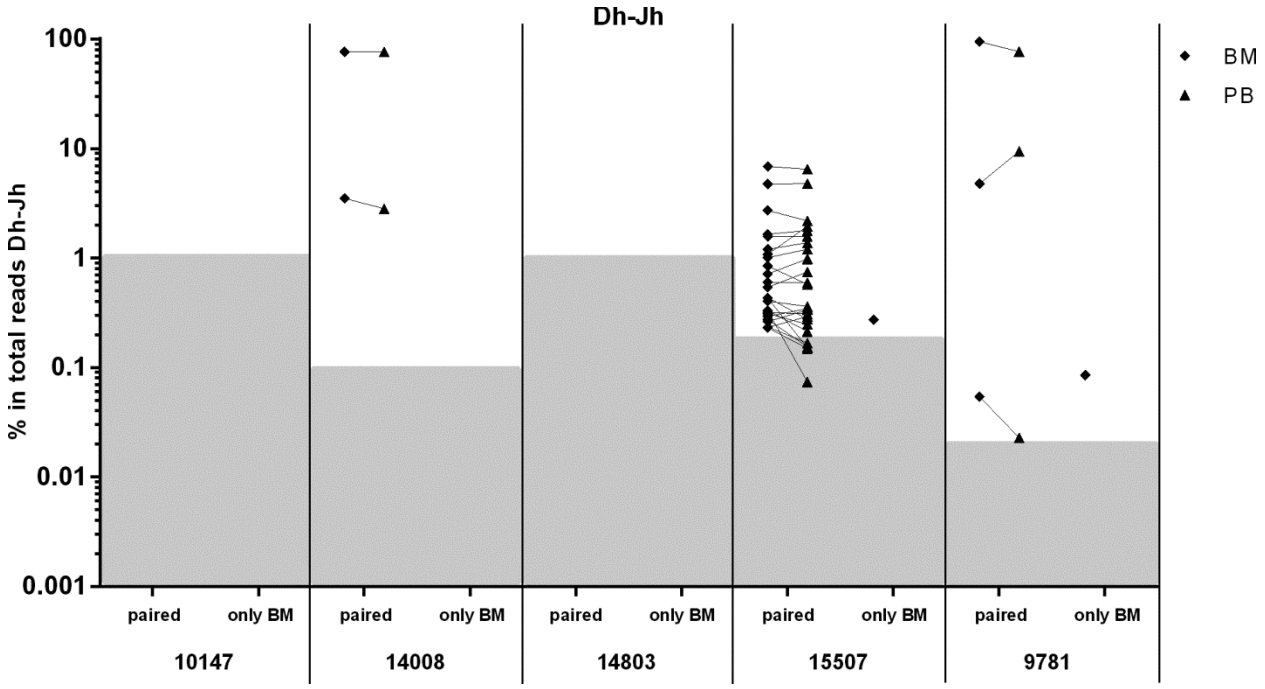


H.

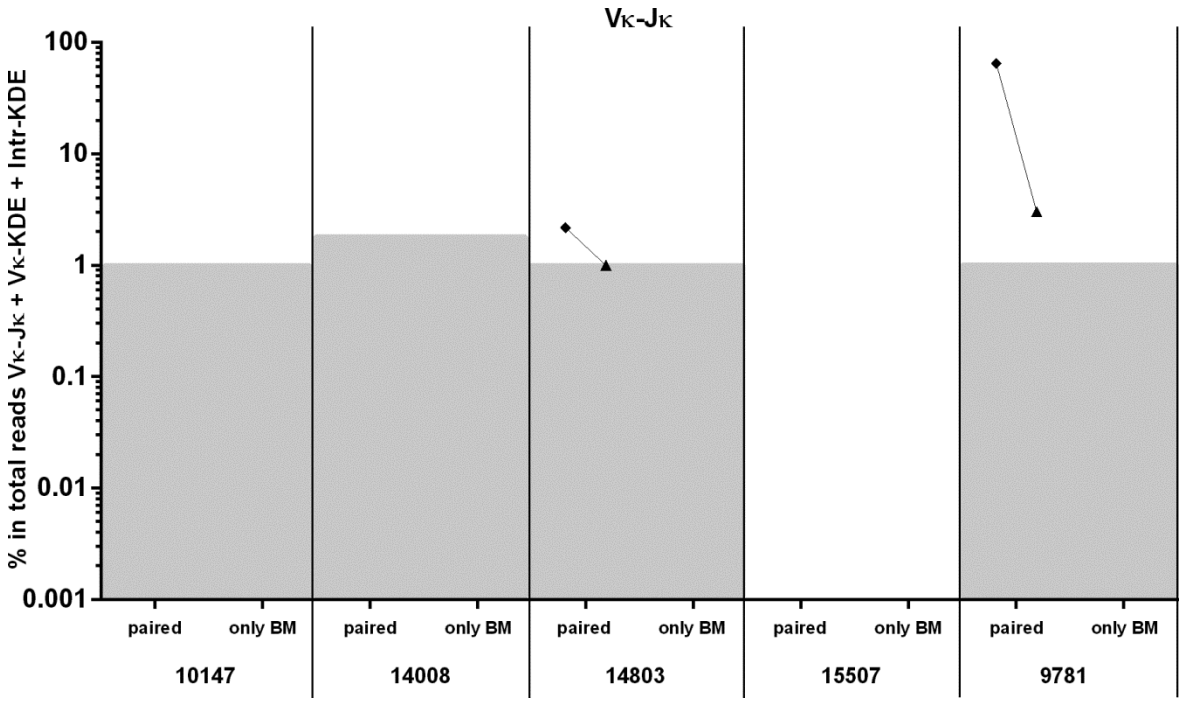


Supplementary Figure 4. Frequencies of leukemic Dh-Jh (A) , Vκ-Jκ (B), Vκ-KDE (C), intronRSS-KDE (D), Vγ-Jγ (E), Vδ-Dδ (F), Dδ-Dδ (G) and Vβ-Jβ (H) rearrangements in paired BM samples (left and right pelvic bone) from BCP-ALL patients at diagnosis. For each leukemic rearrangement, the presence in both BM samples (paired) or in only one of the two BM samples is indicated. The background-area which also contains IG/TR gene rearrangements derived from normal B-cell and T-cell clones is indicated for each sample (grey).

A.

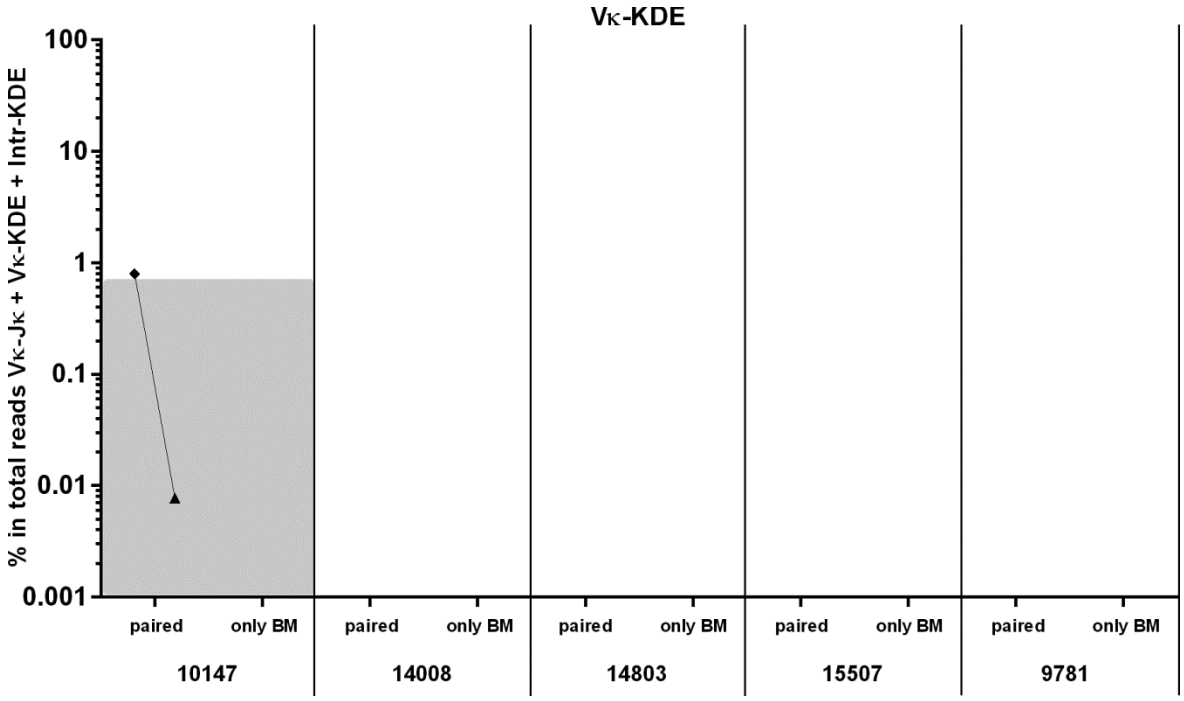


B.

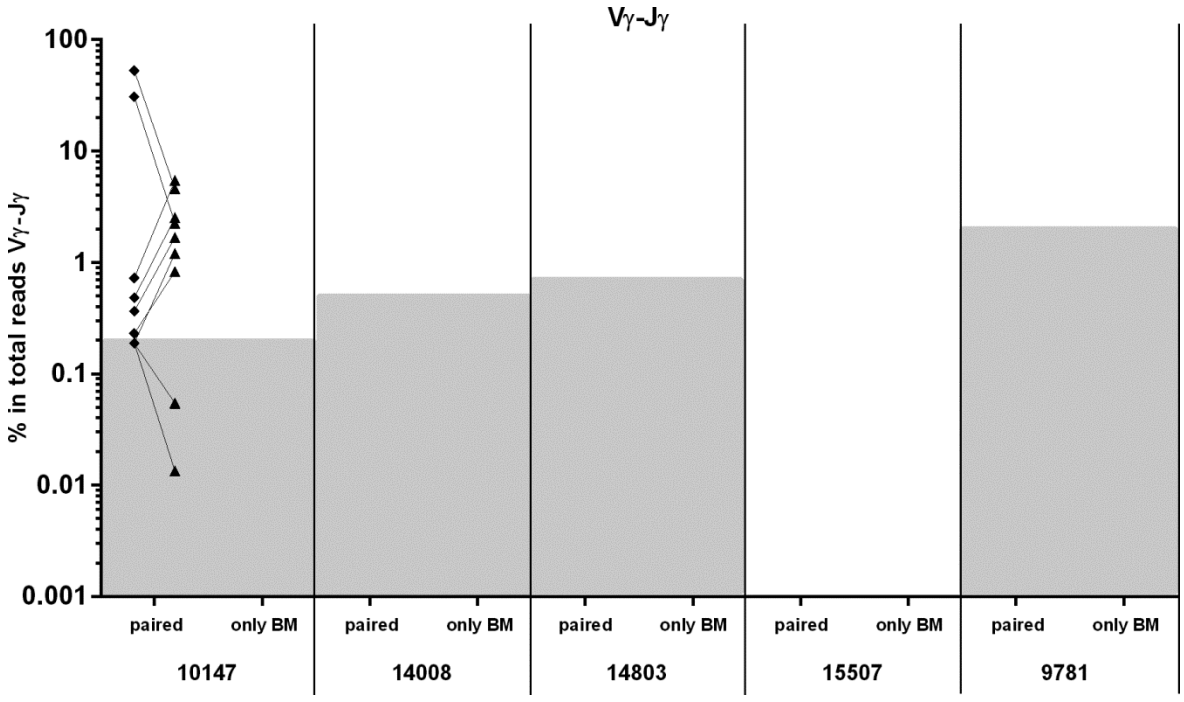


Supplementary Figure 5 (see legend on page 16)

C.

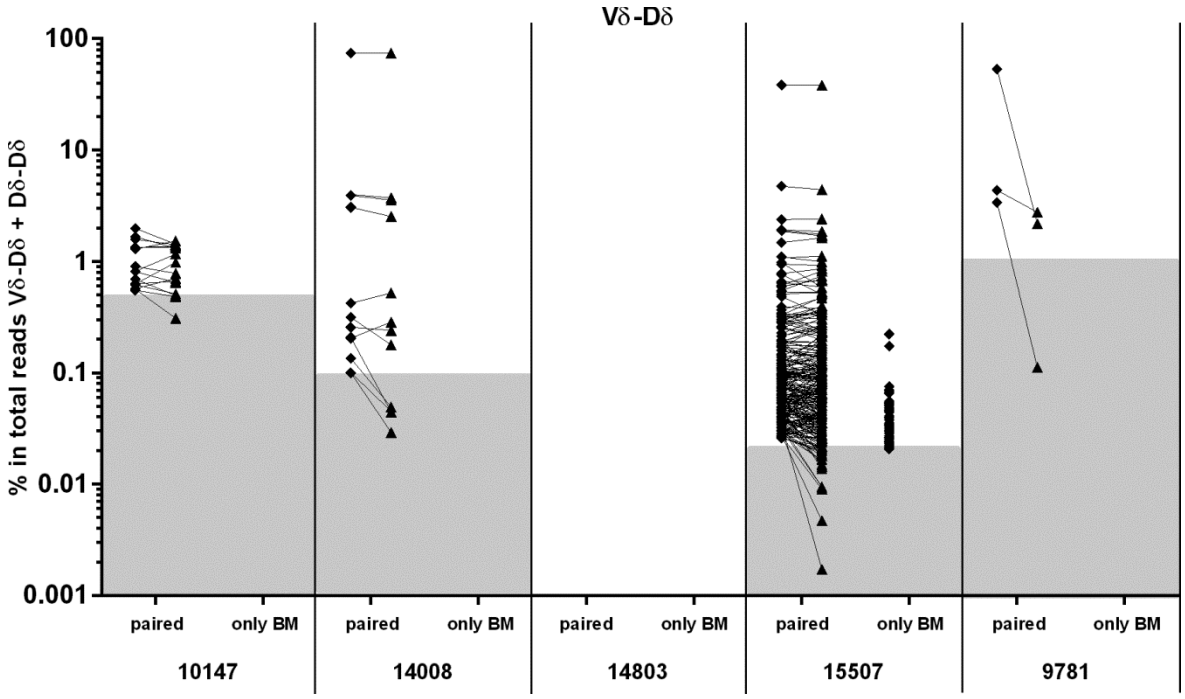


D.

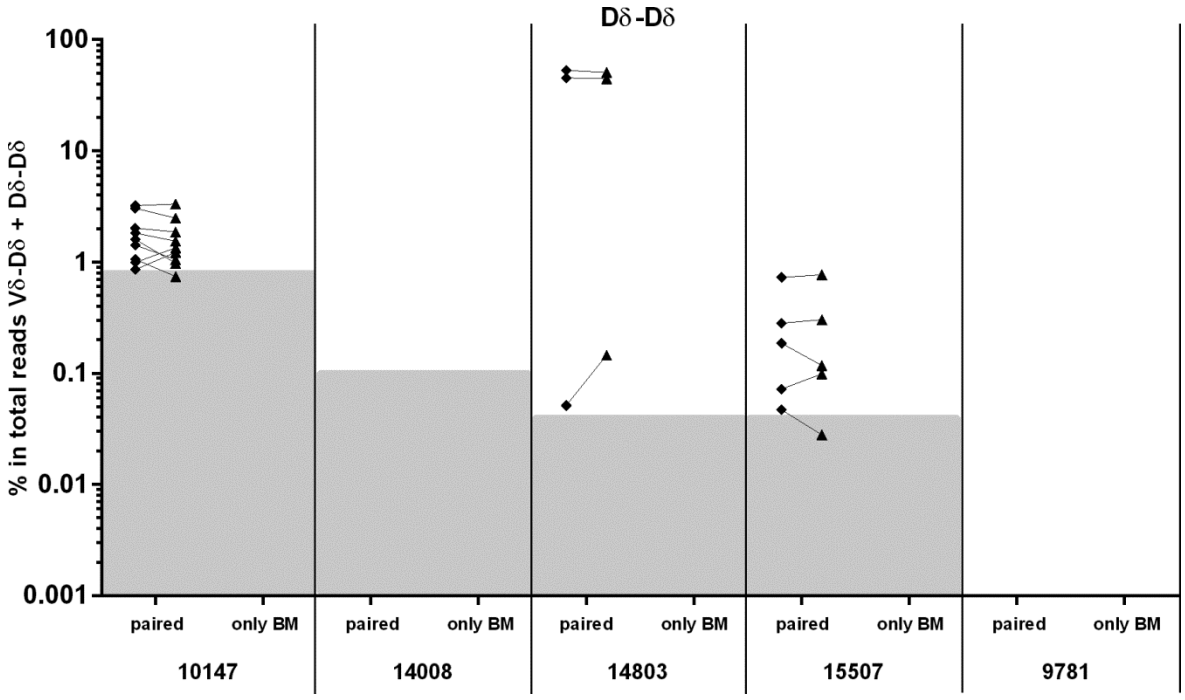


Supplementary Figure 5 (see legend on page 16)

E.

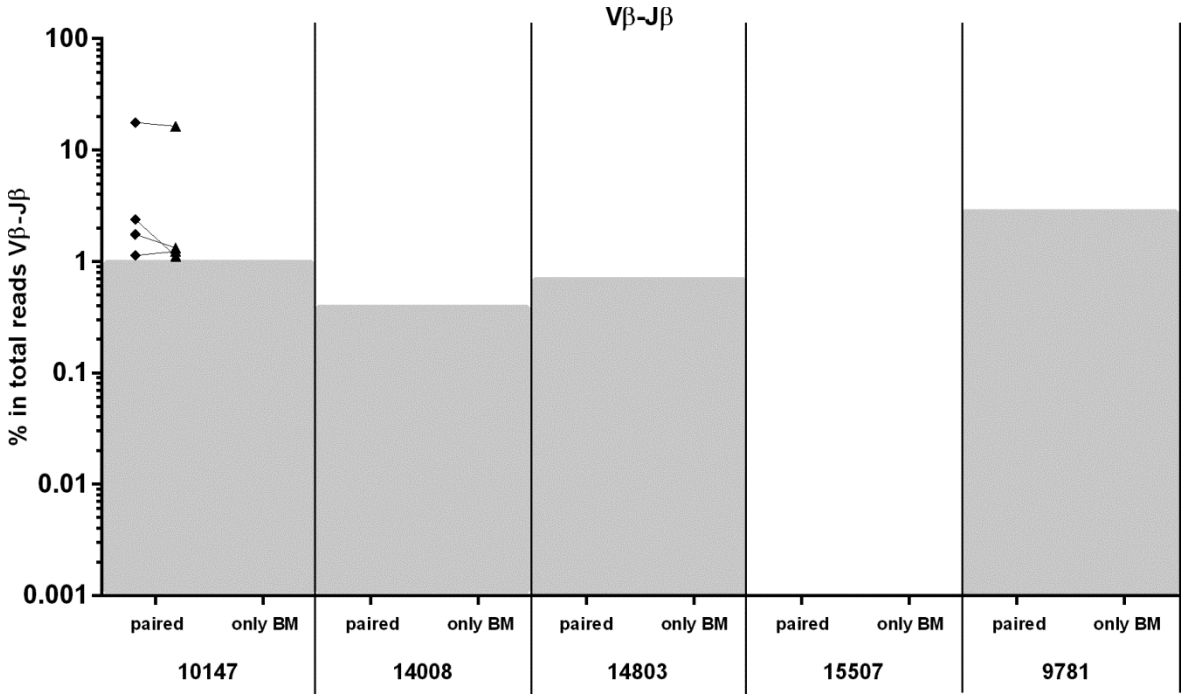


F.



Supplementary Figure 5 (see legend on page 16)

G.



Supplementary Figure 5. Frequencies of leukemic Dh-Jh, Vκ-Jκ, Vκ-KDE, Vγ-Jγ, Vδ-Dδ, Dδ-Dδ and Vβ-Jβ rearrangements in paired BM-PB samples from BCP-ALL patients at diagnosis. For each leukemic rearrangement in the BM sample (i.e. with a frequency above the BM threshold), the presence in both samples (paired) or in only the BM sample is indicated. Diamonds represent rearrangements found in BM, triangles represent rearrangements found in PB. The BM background-area which also contains IG/TR gene rearrangements derived from normal B-cell and T-cell clones is indicated for each sample (grey). IntronRSS-KDE rearrangements with a frequency above the threshold were absent in all patients.

REFERENCES

1. Faham, M., et al., *Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia.* Blood, 2012. **120**(26): p. 5173-80.
2. Blankenberg, D., et al., *Dissemination of scientific software with Galaxy ToolShed.* Genome Biol, 2014. **15**(2): p. 403.