Supporting Information

Logan et al. 10.1073/pnas.1118357109

SI Materials and Methods

CLL Cell Isolation, Flow Cytometry, and DNA Isolation. Peripheral blood samples were acquired from patients at indicated time points following HCT, and the mononuclear cell fraction (PBMC) was cryopreserved in 10% dimethyl sulfoxide (DMSO) after separation on a Ficoll-Hypaque (Sigma-Aldrich) density gradient. Flow cytometric assessment of MRD was achieved by quantifying live lymphocytes [Live/Dead Aqua (Invitrogen) negative] bearing the CLL immunophenotype (low forward/side scatter, $CD45^{+}CD3^{neg}CD56^{neg}CD14^{neg}CD19^{+}CD5^{+}CD23^{+}$) with a median 500,000 events for live cells collected (range 150,000–500,000) on a LSRII flow cytometer (Becton Dickinson). For our dilution series experiment, we purified CLL cells with the above immunophenotype from one patient to >98% purity using an InFlux flow cytometer (Becton Dickinson). These purified CLL cells were diluted into healthy donor PBMC from 1:10 down to 1:100,000. DNA was harvested after washing with buffered saline (pH 7.4), cellular disruption in lysis buffer (50 mM Tris-HCl, 50 mM EDTA, and 1% SDS; pH 7.4) containing proteinase K, followed by phenol extraction, ethanol precipitation, and resuspension in Tris-EDTA buffer (pH 7.4). When comparing MRD quantification by IGH-HTS and FC, we converted the flow cytometric quantification from CLL phenotype cells per microliter of blood to CLL cells per microgram of genomic DNA based on a diploid human genome mass of 6.49 pg per cell.

ASO-PCR. Quantitative real-time PCR was performed by using patient ASO primers as described (1). With this technique, an IGH V-region consensus probe was used with patient-specific primers, one of which anneals to the clonal complementarity determining region 3 (CDR3). When the consensus probe yielded insufficient sensitivity, patient allele-specific probes spanning the CDR3 region were used. Q-PCR reactions were performed on an ABI 7900 thermocycler (Applied Biosystems) by using AmpliTaq Gold DNA polymerase (Applied Biosystems).

Data Analysis. Sequence reads were mapped to germ-line V and J reference sequences downloaded from the IMGT Web site (www.imgt.org) (2) with a method described by Wang et al. (3). Briefly, the IRmap program, a modification of Smith–Waterman algorithm (4), was used to search for the germ-line V and J gene segments for best matches while taking the sequence read quality score at each base into account. Each sequence read was then assigned with the best-matched V and J segments. Reads mapped with <20 V bases or 15 J bases—most, if not all of which were primer-dimers or other PCR artifacts—were eliminated. We analyzed reads through 200 nt to ensure averaged sequence quality score was >35 at every position (Fig. S1).

For MRD quantification, reads with identical V and J segment use and with <20 nucleotide differences from the dominant CLL clone, previously determined for every CLL patient from a traditional Sanger sequencing read performed for CLL prognostication, were retained for further analysis. 454 pyrosequencing is error-prone at homopolymeric regions and produces primarily insertion and deletion (indel) errors (4). To avoid undercounting the cancer clone, a rescue procedure was designed to salvage reads differing at homopolymeric regions. After aligning each read with the dominant CLL clone, each homopolymer region with three or more of the same nucleotide in succession were identified in pairwise alignment. Insertions, deletions, or substitutions of one or two nucleotides around each homopolymer were adjusted according to the dominant clone sequence. Multiple sequence alignments of clones and potentially related subclones were then performed by using the ClustalW2 algorithm (www.ebi.ac.uk/Tools/clustalw2/index.html) with default parameters (5).

Alignment of sequencing reads corrected for homopolymeric indels and substitutions led to discovery of random single-nucleotide indels at low frequencies, which likely also resulted from pyrosequencing errors (6). These indels were removed according to the dominant CLL clonotype for each patient. We then analyzed the resulting sequence alignments for MRD quantification and phylogenetic analysis. Sequencing reads that differed by up to one nucleotide substitution or gap were aggregated into the final CLL MRD count. Remaining reads with two or more nucleotide substitutions or gaps were analyzed for their relationship to the dominant CLL clone if their clonal frequency was two or greater and they were observed in at least two samples from each patient. By using these stringent criteria, artifactual clonotypes were minimized in the phylogenetic analysis. Neighbor-joining phylogenetic trees were constructed by using unrooted methods implemented in Phylip (evolution.gs.washington.edu/phylip.html) with default parameters (7) and were plotted with the APE R package (ape.mpl.ird.fr) (8).

1. Ladetto M, et al. (2000) Real-time polymerase chain reaction of immunoglobulin rearrangements for quantitative evaluation of minimal residual disease in multiple myeloma. *Biol Blood Marrow Transplant* 6:241–253.
2. Giudicelli V, Chaume D, Lefranc MP (2005) IMGT/GENE-DB: A comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* 33 (Database issue):D256–D261.
3. Wang C, et al. (2010) High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci USA* 107:1518–1523.
4. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. *Genome Res* 17:1195–1201.
5. Chenna R, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.
6. Nguyen P, et al. (2011) Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire. *BMC Genomics* 12:106.
7. Retief JD (2000) Phylogenetic analysis using PHYLIP. *Methods Mol Biol* 132:243–258.
8. Paradis E, Claude J, Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.

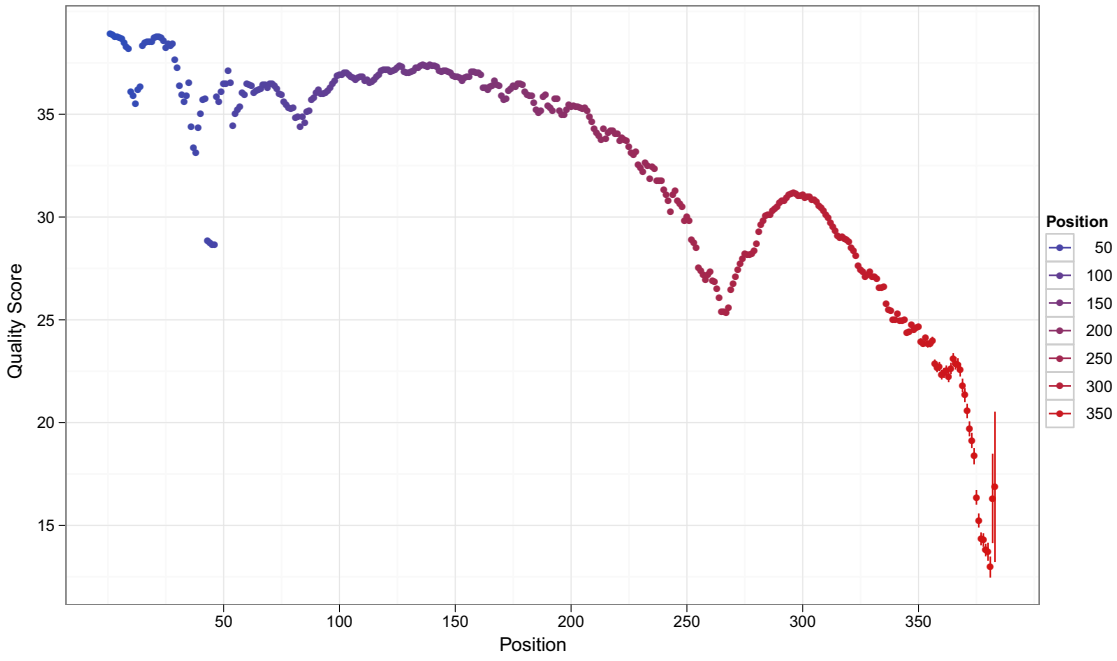

Fig. S1. Sequence quality scores by nucleotide position. The median sequence quality score at each position across the entire 454 pyrosequencing dataset demonstrate that sequence quality degraded after position 200.

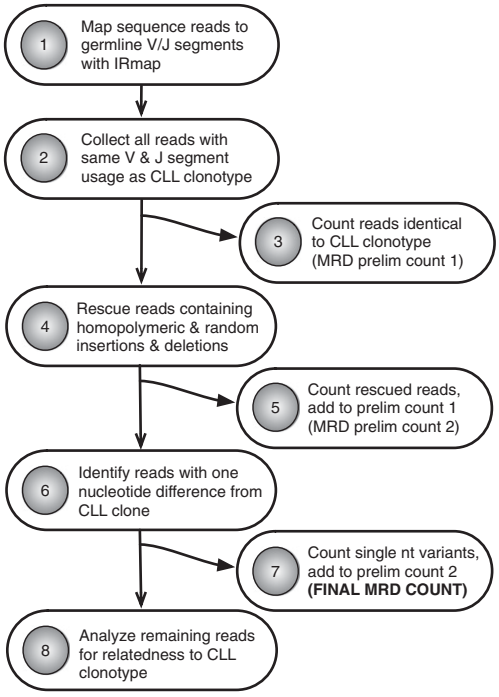

Fig. S2. 454 pyrosequencing error handling algorithm. An algorithm for systematically handling 454 pyrosequencing errors was developed. Final MRD counts were an aggregation of reads identical to the dominant CLL clonotypes (step 3), reads corrected for homopolymeric and random single-nucleotide (nt) indels (step 5), and reads containing up to one nt substitution or gap (step 7). Remaining reads were analyzed for subclonal phylogenetic relationships.

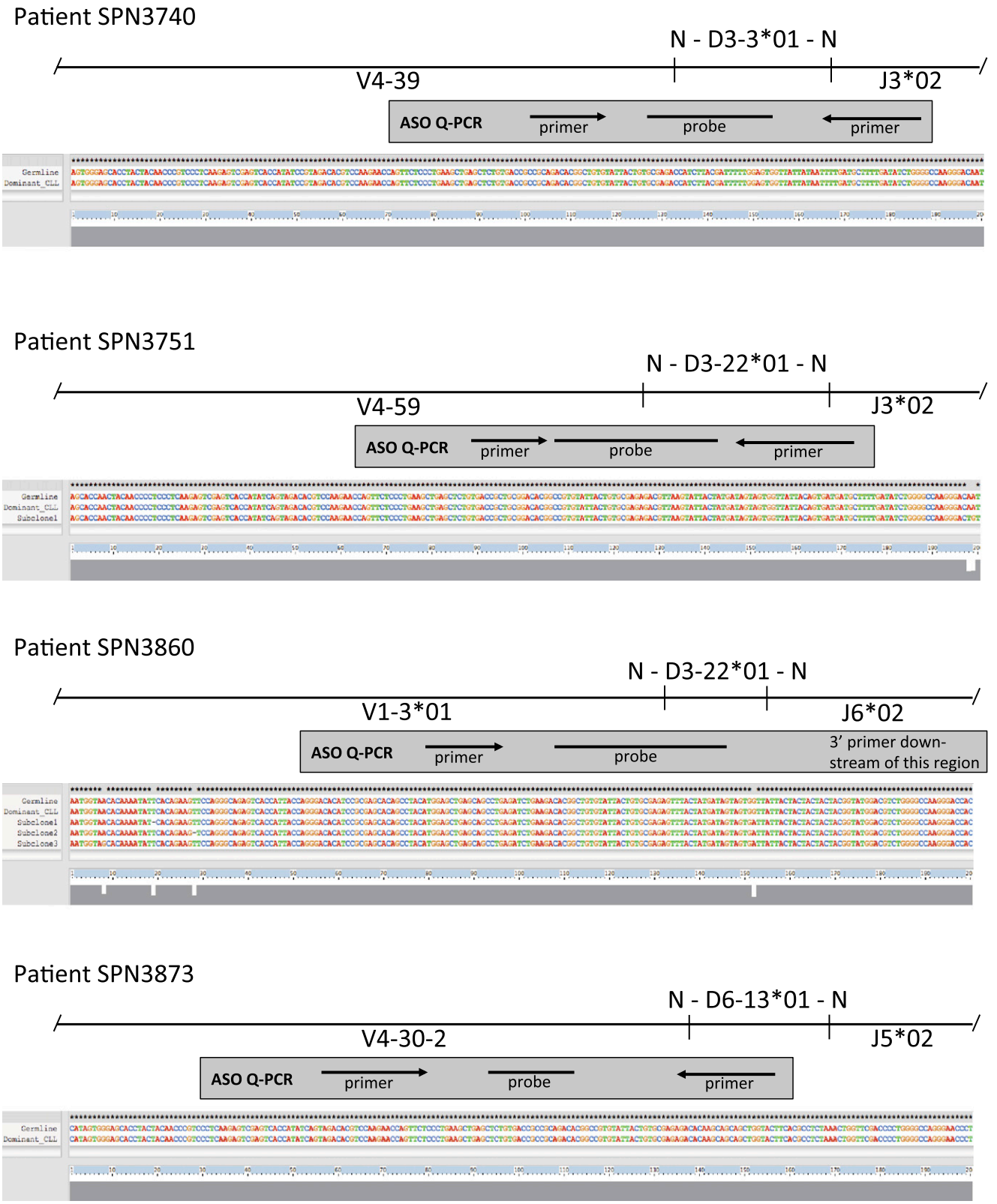

Fig. S3. (Continued)

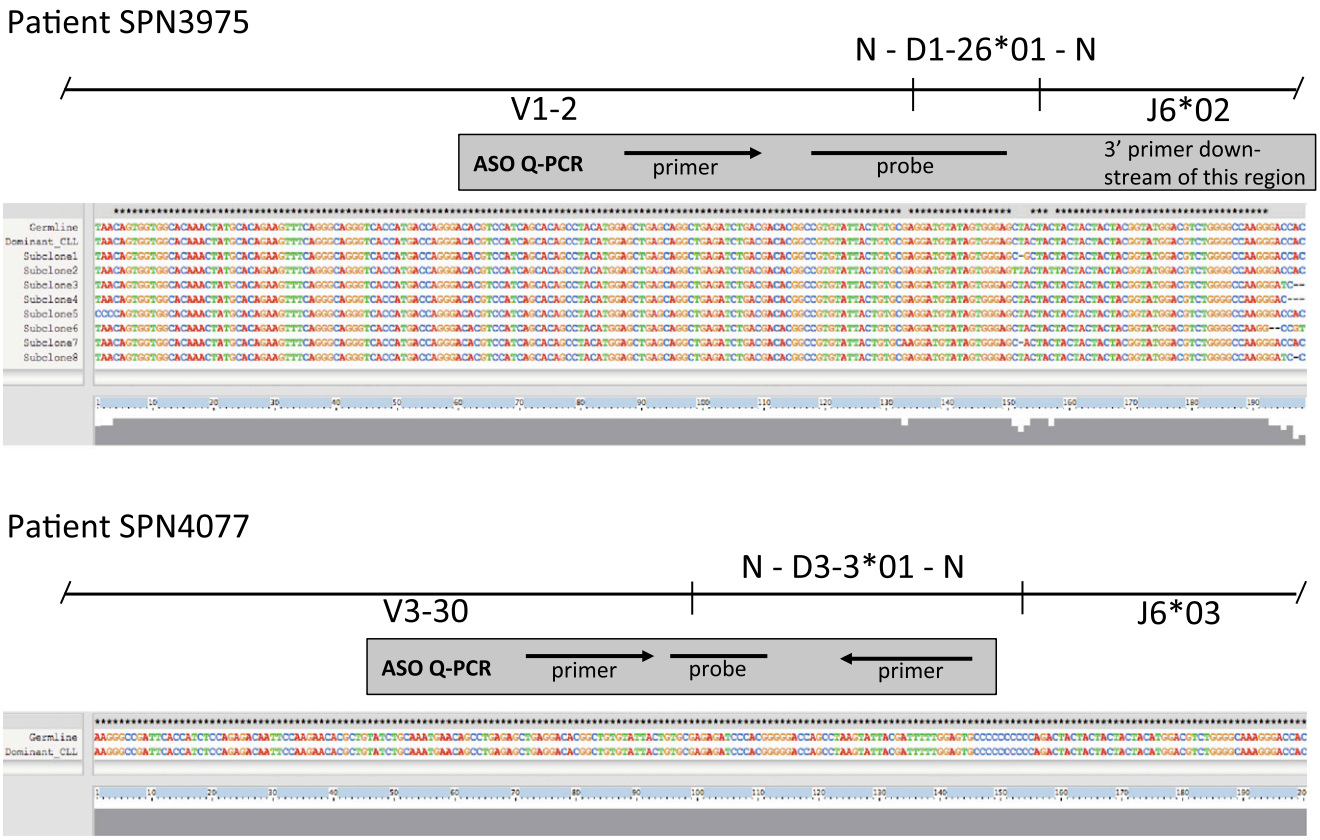

Fig. S3. Subclone sequence alignments. Sequence alignments of the dominant CLL clonotypes (Dominant_CLL) for each patient with the concordant IMGT consensus germ-line sequences are shown in ClustalX output. For those patients with legitimate subclones, their sequence alignment is also shown. Missing asterisks above the nucleotide sequence indicate positions at which nucleotide substitutions or gaps were found. The inverted histogram beneath the sequence position markers indicates the relative number of mutated clonotypes found at each position. The annealing sites for each patient's ASO-PCR primers and probes are depicted to demonstrated their position relative to mutation sites in subclones.

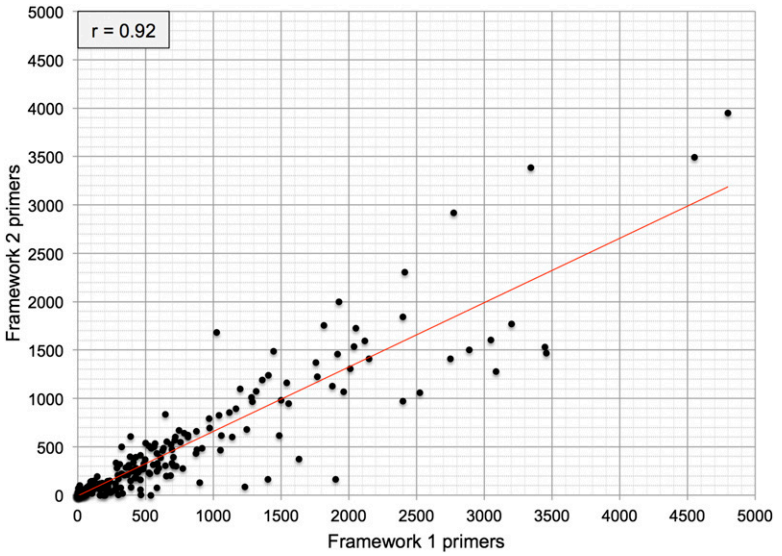

Fig. S4. Concordance of IGH repertoire coverage using FR1-J and FR2-J amplimers. Each position on the graph represents the number of times a specific clonotype was seen with each primer set. The Pearson *r* correlation coefficient was 0.92.

Patient SPN3740

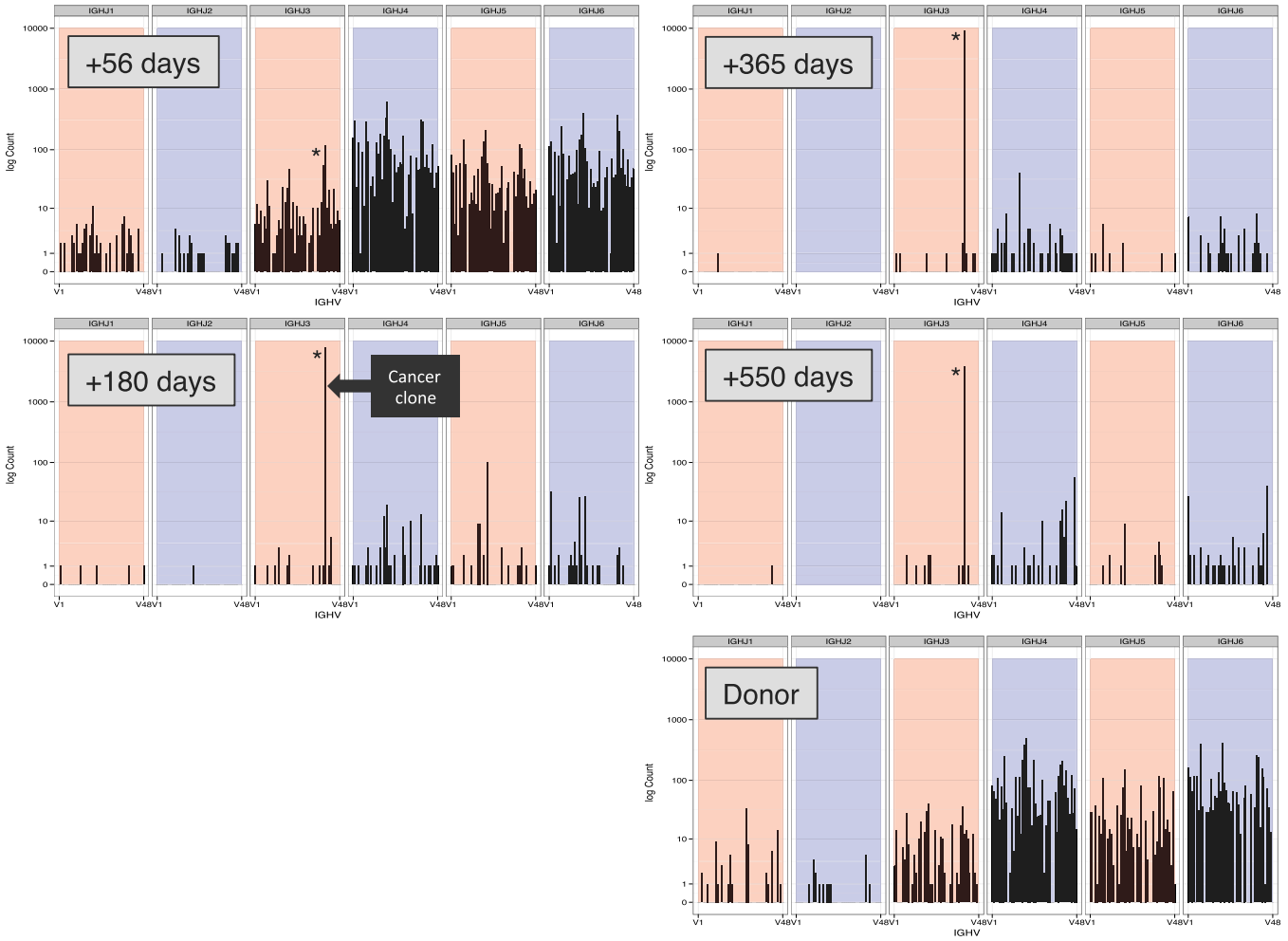

Fig. S5. (Continued)

Patient SPN3751

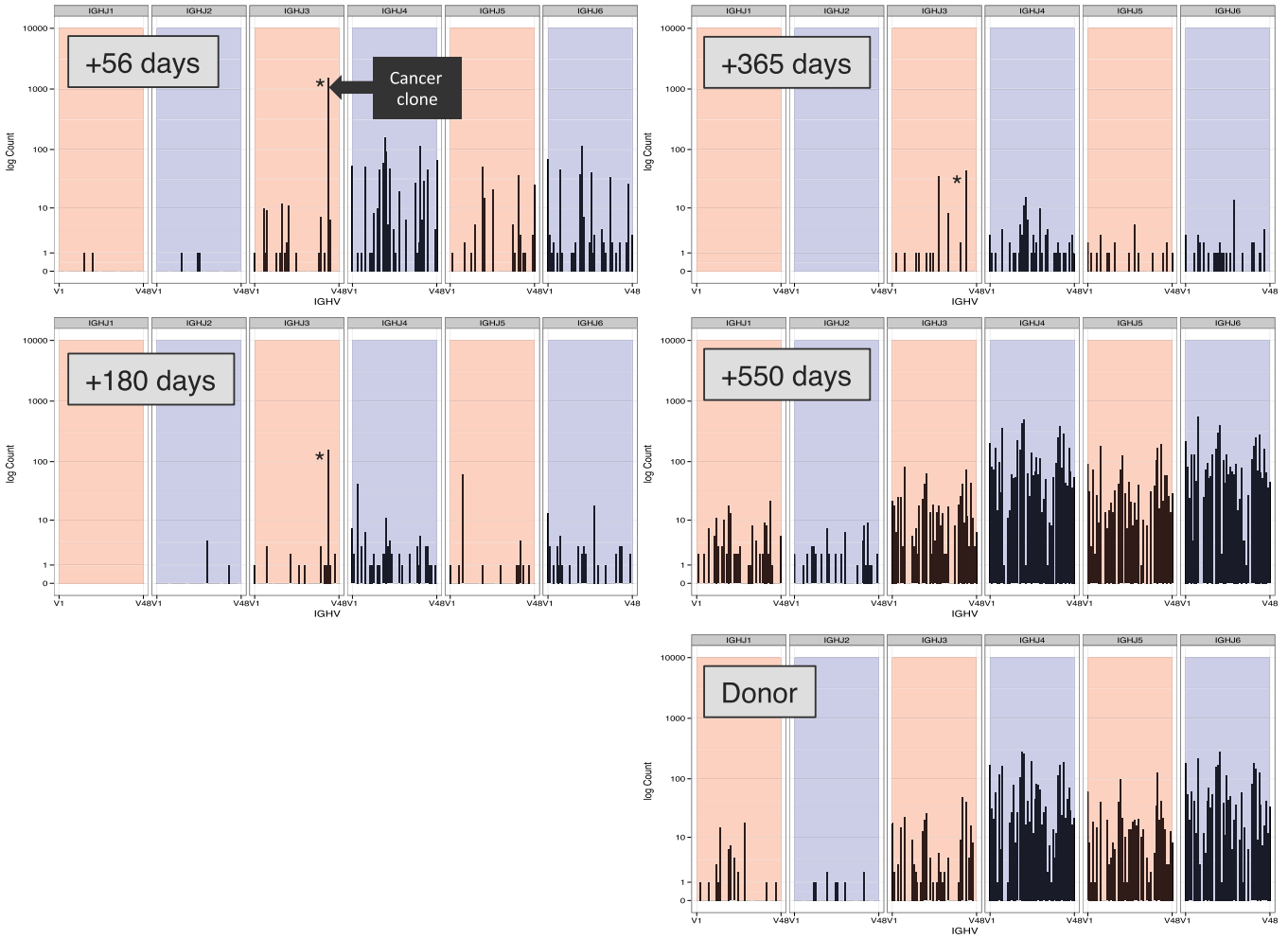

Fig. S5. (Continued)

Patient SPN3860

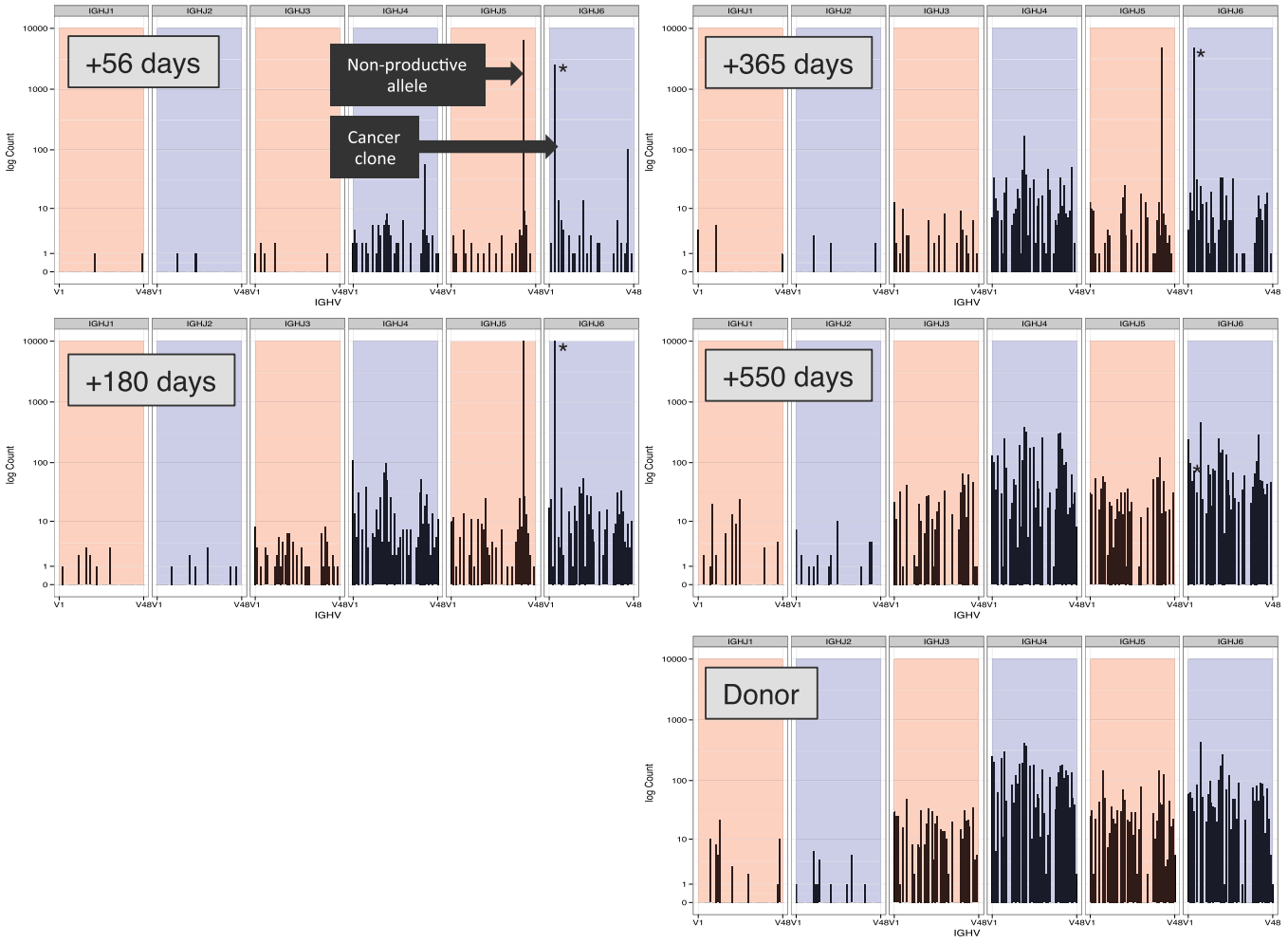

Fig. S5. (Continued)

Patient SPN3873

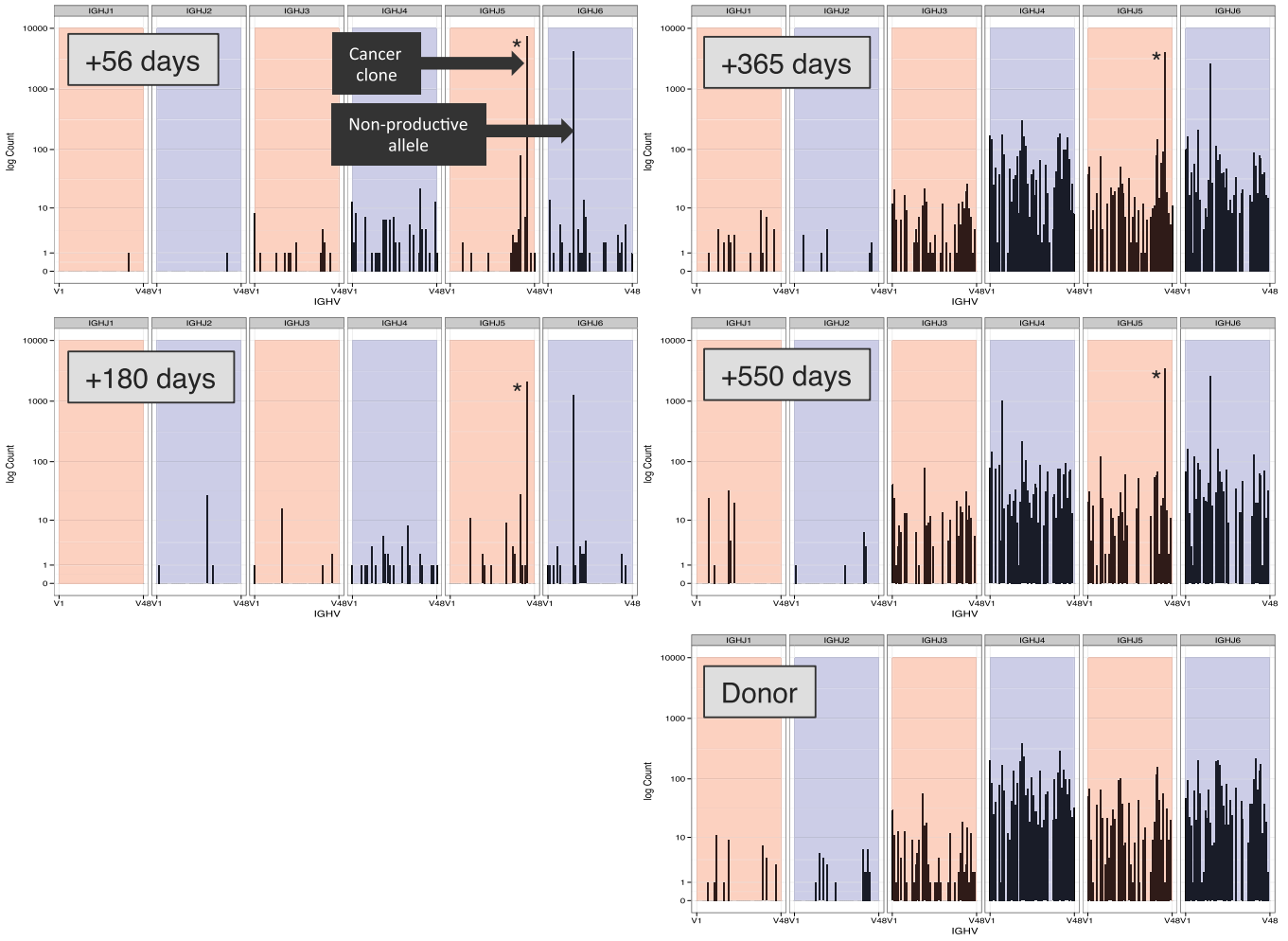

Fig. S5. (Continued)

Patient SPN3975

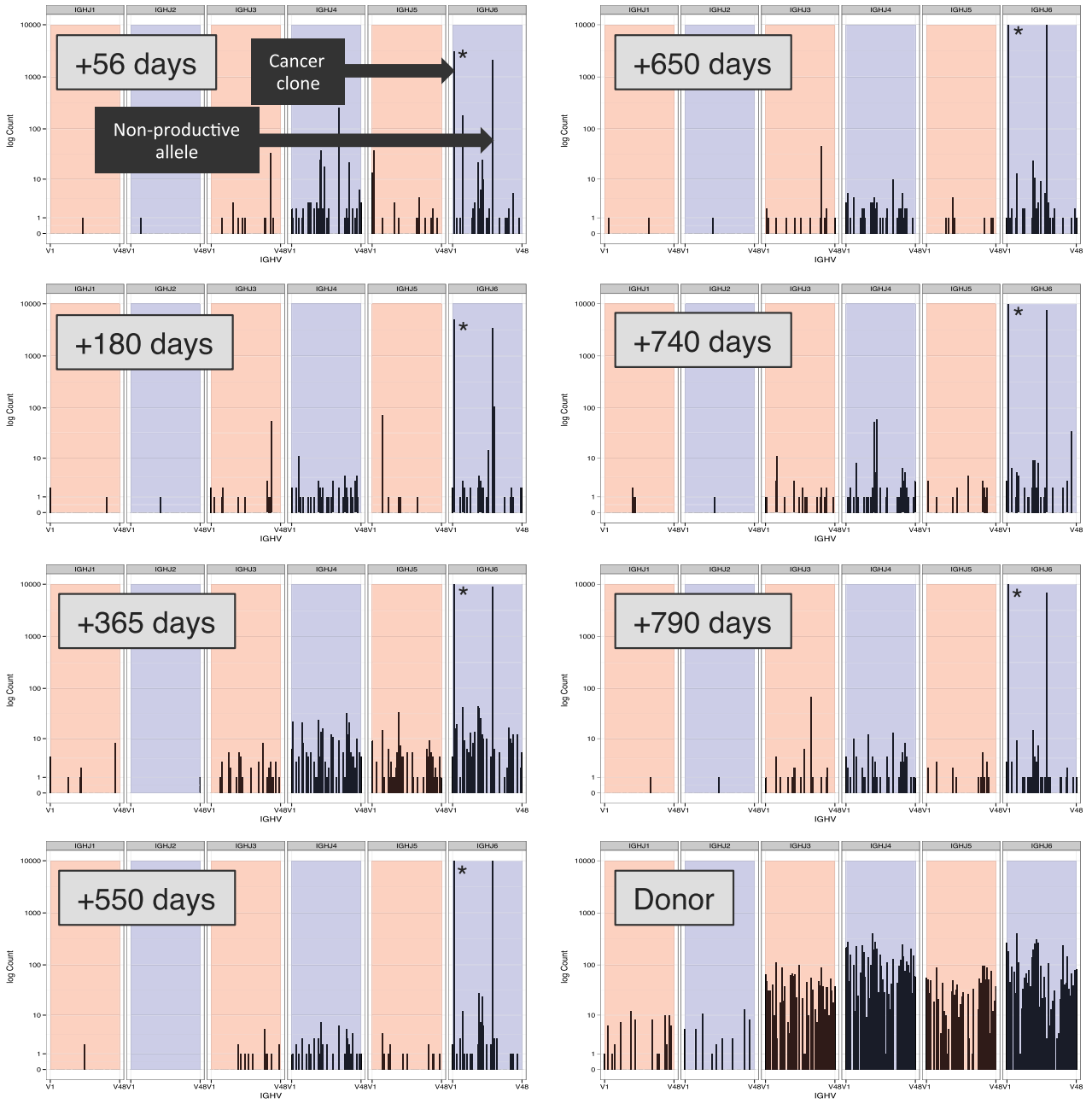

Fig. S5. IGH-HTS reveals kinetics of IGH repertoire reconstitution following allo-HCT for CLL. All V-J recombinations detected in peripheral blood from patients at days +56, +180, +365, and +550 following allo-HCT are demonstrated. The *x* axis at the bottom of each section represents the IgH V segments 1–49 which combined with IgH J segments 1–6 as defined along the *x* axis at the top of each section. The *y* axis represents the number of total reads for that recombination pair. The repertoire of each patient's donor is demonstrated for comparison. The CLL clonotype is demarcated by an asterisk.

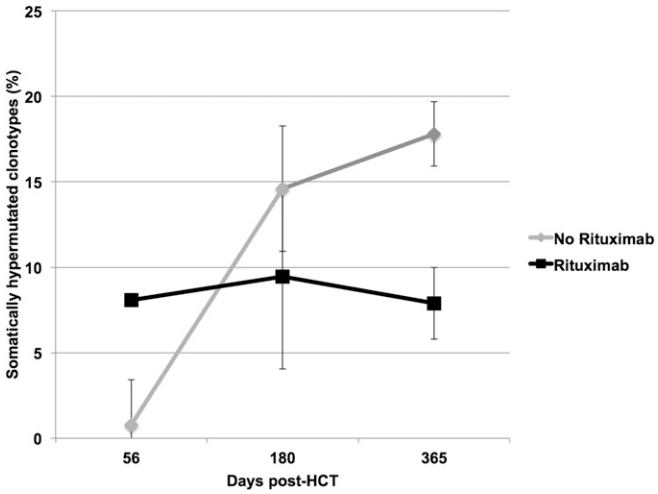

Fig. S6. IGH-HTS reveals effects of posttransplant anti-B-cell therapy on IGH repertoire reconstitution. The degree of somatic hypermutation across the entire IGH repertoire in patients who did ($n = 4$) or did not ($n = 2$) receive posttransplant Rituximab for GVHD prophylaxis is demonstrated at days +56, +180, and +365 following allotransplant.

Table S1. Patient characteristics

Patient characteristics			CLL disease features								CLL status at transplant				
SPN	Age at CLL dx	Sex	Cyto by FISH	Flud refract*	CLL IGH	VH	JH	Max Rai stage	No. of prior regimens	Alemtuzamab before HCT	CLL status at HCT	WBC at HCT	Bone marrow CLL as % CD45	LN >5 cm before HCT	Time from dx to HCT, mo
3740	48	F	17p-	Yes	Unmut	4–39	3	IV	5	No	PR3	4.0	80	Yes	128
3751	53	M	17p-	Yes	Unmut	4–59	3	IV	6	Yes	PR2	7.0	1	No	43
3860	55	M	nl	Yes	Unmut	1–3	6	IV	6	Yes	PR4	6.0	22	No	84
3873	50	F	13q-	No	Unmut	4–22	5	III	4	Yes	PR4	3.3	50	No	82
3975	43	M	17p-	Yes	Unmut	1–2	6	III	3	Yes	CR3	1.6	0	No	60
4077	51	M	11q-	No	Unmut	3–30-3	6	II	2	No	CR2	6.8	0	No	72

CR complete response; Cyto by FISH, cytogenetics determined by FISH; dx, diagnosis; F, female; Flud refract, fludarabine refractory; IGH, heavy chain Ig; JH, junctional heavy chain; max, maximum; LN, lymph node; M, male; PR, partial response; SPN, Stanford patient number; VH, variable heavy chain; Unmut, unmutated IGH CLL clone (CLL IGH sequence varies by <2% from germ line); WBC, white blood cell.
*Defined as failure to achieve a partial response or complete response to at least one fludarabine-containing regimen, disease progression while on fludarabine treatment, or disease progression within 6 mo of the last dose of fludarabine.

Table S2. Hematopoietic cell transplant clinical outcomes

Pt characteristics			Donor characteristics		Transplant characteristics		GVHD		Donor CD3 T-cell chimerism, %		CLL disease response			
SPN	Age at HCT	Sex	Sex	Relation	No. of CD34+ cells per rec weight kg	Post-HCT Rituximab	Acute (grade)	Chronic	Day 30	Day 90	Time to relapse, d	DLI day	CLL status 3 y post-HCT	Current status, d post-HCT
3740	58	F	M	Sibling	12.5×10 (6)	Yes	0	None	93	84	182	206	Died of tCLL	Died, 646
3751	56	M	M	Unrel.	13.3×10 (6)	Yes	0	Exten.	92	99	1,120	1,180	PD	Alive, 1,587
3860	62	M	M	Unrel.	9.2×10 (6)	No	1	Exten.	32	64	None	None	CR	Alive, 1,426
3873	59	F	M	Sibling	5.7×10 (6)	Yes	0	None	86	88	798	840	CR	Alive, 1,398
3975	48	M	F	Sibling	7.4×10 (6)	Yes	0	Exten.	91	60	210	301	PD	Alive, 1,237
4077	57	M	M	Unrel.	13.3×10 (6)	No	0	None	97	99	910	1,090	PD	Alive, 1,150

CR, complete response; DLI, donor lymphocyte infusions; Exten., extensive; F, female; GVHD, graft versus host disease; M, male; PD, progressive disease; PR, partial response; Pt, patient; SPN, Stanford patient number; tCLL, transformed CLL; Unrel., unrelated.

Table S3. IGH-HTS results and error corrections

		HPI				RSI				SBS					
Day	Dominant clone—identical reads [A]	Total HPI aggregated [B]	MRD subtotal with HPI	[B] as % of [A+B+C+D]	Unique HPI	Total RSI aggregated [C]	MRD subtotal with RSI	[C] as % of [A+B+C+D]	Unique RSI	Total SBS aggregated [D]	MRD final with SBS	[D] as % of [A+B+C+D]	Unique SBS	[B+C+D] as % of [A+B+C+D]	[A] as % of [A+B+C+D]
SPN3740															
56	73	14	87	13.1	6	14	101	13.1	6	6	107	5.6	6	31.8	68.2
180	5,474	717	6,191	9.4	109	951	7,142	12.4	92	509	7,651	6.7	136	28.5	71.5
365	5,951	1,421	7,372	15.8	145	1,179	8,551	13.1	106	431	8,982	4.8	128	33.7	66.3
550	2,570	352	2,922	9.3	63	581	3,503	15.4	65	281	3,784	7.4	98	32.1	67.9
SPN3751															
56	1,219	48	1,267	3.1	36	106	1,373	6.9	30	167	1,540	7.7	77	20.8	79.2
180	117	4	121	2.6	3	14	135	9.1	7	19	154	10.6	17	24.0	76.0
365	37	0	37	0.0	0	1	38	2.3	1	5	43	8.3	4	14.0	86.0
550	2	2	4	25.0	0	0	4	0.0	0	4	8	12.0	1	75.0	25.0
SPN3860															
56	2,082	54	2,136	2.1	33	167	2,303	6.6	48	238	2,541	9.4	76	18.1	81.9
180	7,719	386	8,105	3.7	105	685	8,790	6.5	102	1,674	10,464	16.0	181	26.2	73.8
365	3,504	112	3,616	2.4	67	322	3,938	6.8	54	808	4,746	17.0	132	26.2	73.8
550	26	1	27	1.5	3	6	33	9.2	4	32	65	49.2	6	60.0	40.0
SPN3873															
56	6,301	389	6,690	5.1	95	402	7,092	5.3	101	521	7,613	6.8	125	17.2	82.8
180	1,633	93	1,726	4.6	47	127	1,853	6.3	46	174	2,027	8.6	79	19.4	80.6
365	3,265	183	3,448	4.5	68	303	3,751	7.5	65	286	4,037	7.1	99	19.1	80.9
550	2,658	185	2,843	5.4	78	228	3,071	6.7	60	325	3,396	9.6	122	21.7	78.3
SPN3975															
56	2,046	370	2,416	12.2	101	297	2,713	9.8	65	317	3,030	10.5	98	32.5	67.5
180	3,399	478	3,877	9.9	99	296	4,173	6.1	56	644	4,817	13.4	135	29.4	70.6
365	10,727	1,061	11,788	7.5	180	1,078	12,866	7.6	133	1,245	14,111	8.8	180	24.0	76.0
550	19,848	1,376	21,224	5.5	200	1,852	23,076	7.4	247	1,851	24,927	7.4	216	20.4	79.6
650	18,108	1,563	19,671	6.7	217	1,774	21,445	7.6	244	1,938	23,383	8.3	233	22.6	77.4
739	7,958	981	8,939	9.1	165	852	9,791	7.9	136	974	10,765	9.0	183	26.1	73.9
790	8,283	824	9,107	7.5	152	933	10,040	8.5	140	919	10,959	8.4	169	24.4	75.6
SPN4077															
DX	2,640	664	3,304	11.3	133	2,314	5,618	39.3	46	273	5,891	7.7	108	55.2	44.8
56	0	1	1	100.0	0	0	1	0.0	0	0	1	10.6	0	100.0	0.0
180	1	0	1	0.0	0	0	1	0.0	0	0	1	8.3	0	0.0	100.0
365	0	0	0	0.0	0	0	0	0.0	0	0	0	0.0	0	n/a	n/a
550	4	3	7	18.8	3	9	16	56.3	6	0	16	5.9	0	75.0	25.0

HPI, homopolymeric indels; RSI, random single-nucleotide indels; SBS, single base substitution; SPN, Stanford patient number.