Leukemia (2012) **26,** 1564 – 1575

www.nature.com/leu

Open

ORIGINAL ARTICLE

Quantification of subclonal distributions of recurrent genomic aberrations in paired pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia

SJL Knight[1,2], C Yau[3], R Clifford[4], AT Timbs[4], E Sadighi Akha[2], HM Dréau[4], A Burns[1,4], C Ciria[5], DG Oscier[6], AR Pettitt[7], S Dutton[5], CC Holmes[2,3], J Taylor[1,2], J-B Cazier[2] and A Schuh[1,4]

Genome-wide array approaches and sequencing analyses are powerful tools for identifying genetic aberrations in cancers, including leukemias and lymphomas. However, the clinical and biological significance of such aberrations and their subclonal distribution are poorly understood. Here, we present the first genome-wide array based study of pre-treatment and relapse samples from patients with B-cell chronic lymphocytic leukemia (B-CLL) that uses the computational statistical tool OncoSNP. We show that quantification of the proportion of copy number alterations (CNAs) and copy neutral loss of heterozygosity regions (cnLOHs) in each sample is feasible. Furthermore, we (i) reveal complex changes in the subclonal architecture of paired samples at relapse compared with pre-treatment, (ii) provide evidence supporting an association between increased genomic complexity and poor clinical outcome (iii) report previously undefined, recurrent CNA/cnLOH regions that expand or newly occur at relapse and therefore might harbor candidate driver genes of relapse and/or chemotherapy resistance. Our findings are likely to impact on future therapeutic strategies aimed towards selecting effective and individually tailored targeted therapies.

Leukemia (2012) **26,** 1564 – 1575; doi:10.1038/leu.2012.13

Keywords: B-CLL; clonal architecture; genome-wide arrays; OncoSNP; genome imbalance; copy neutral loss of heterozygosity

INTRODUCTION

Genome-wide array approaches and sequencing analyses are powerful tools for identifying genetic aberrations in cancers, including leukemias and lymphomas.[1-10] B-cell chronic lymphocytic leukemia (B-CLL) is the most common adult leukemia in the Western World.[11,12] It is characterized by a chronic relapsing course and the development of chemotherapy refractoriness, leading to death in a significant subset of patients. There is, therefore, a clinical need to develop novel treatment strategies that overcome chemotherapy resistance.

B-CLL can be subclassified according to immunoglobulin heavy chain variable gene sequence (*IGHV*) homology into 'mutated' or 'unmutated' cases, reflecting good or poor prognosis, respectively. B-CLL shares many of the molecular characteristics seen in other forms of cancer and recurrent copy number alteration (CNAs) with relevance to prognosis have been described, including trisomy of chromosome 12 (16%) and loss of chromosomal regions 17p13.1 (*TP53*; 7%),[13-17] 11q22.3 (*ATM*; 18%)[18-19] and 13q14.3 (*DLEU2* and/or miR-15a/16-1; 55%).[20-30] In addition, single-nucleotide polymorphism (SNP) and comparative genomic hybridization arrays have allowed novel recurrent genomic abnormalities to be identified[31-33] that have been linked to prognosis in B-CLL.[30,34-38]

Current models of cancer progression are based on the concept that tumors are subject to the Darwinian process of evolution and selection.[39,40] Recent studies in acute lymphoblastic leukemia have provided pivotal insights into the complex sequence of events during leukemogenesis, showing that the initiating mutation is followed by CNAs that drive the emergence of frank leukemia.[39-42] Together, these data imply that at least some CNAs/copy neutral loss of heterozygosity regions (cnLOHs) are likely to be involved in driving leukemia progression and therefore might contribute to relapse.

This led us to hypothesize that subclones containing driver CNAs/cnLOHs would newly occur or expand in relapse samples compared with samples taken before treatment and would be recurrent within our patient cohort. In contrast, random passenger mutations would remain unchanged or decrease/disappear in paired pre-treatment/relapse samples and would not be observed recurrently in different patients within the cohort. If correct, then we would anticipate that the identification of driver genes within recurrent or emerging/expanding regions of CNA/cnLOH might, in the longer term, have the potential to inform the design of novel therapies aimed at treating relapsed B-CLL.

In the present study we tested our hypothesis by systematically tracking the presence and subclonal distribution of CNAs/cnLOHs in patients before treatment and at subsequent relapse. To achieve this, we used high-resolution SNP array technology. We chose the newly developed computational statistical tool OncoSNP[10] that provides quantitative measures of cell admixture on a per-SNP level. OncoSNP was selected in preference to other dedicated cancer tools for SNP array analysis,[43-47] because our

[1]NIHR Biomedical Research Centre, Oxford, UK; [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [3]Department of Statistics, University of Oxford, Oxford, UK; [4]The Oxford Biomedical Research Centre Haemato-Molecular Diagnostic Service, Manor Way, John Radcliffe Hospital, Oxford, UK; [5]Oxford Clinical Trials Research Unit (OCTRU), Centre for Statistics in Medicine (CSM), Wolfson College Annexe, Oxford, UK; [6]Royal Bournemouth Hospital, Castle Lane East, Bournemouth, UK and [7]School of Cancer Studies, University of Liverpool, Liverpool, UK. Correspondence: Dr A Schuh, Oxford Cancer and Haematology Centre, Oxford University Hospitals NHS Trust, Churchill Hospital, Old Road, Oxford OX3 7LJ, UK.
E-mail: anna.schuh@ndcls.ox.ac.uk
Received 21 March 2011; revised 11 January 2012; accepted 12 January 2012; accepted article preview online 19 January 2012; advance online publication, 14 February 2012

data were generated from samples at two time points only and our study was specific to B-CLL, where spatial heterogeneity (for example, samples biopsied from different parts of a tumor, metastases, etc) is not applicable (see Supplementary Information, online). Furthermore, we were interested not only in the accumulation of genomic aberrations over time, but also particularly in the expansion of distinct subclones. Thus, OncoSNP best suited our study where (i) samples differed from solid tumors in being generally diploid with comparatively fewer CNAs (ii) normal cell contamination could be minimized and (iii) we had a specific interest in subclonal evolution.

PATIENTS AND METHODS

For a detailed account of the Materials, Patients and Methods used in this study please refer to the Supplementary Information online.

Patients and samples

All patients gave written informed consent in accordance with the Declaration of Helsinki. In all, 135 samples from 93 patients were included in our cohort. For 42 of the 93 patients, paired pre-treatment and relapse samples were available. Only samples from patients with lymphocytes contributing $>90\%$ of the total white blood cell count were included in the analysis. DNA was extracted from vital frozen cells using the QIAamp DNA Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol.

Fluorescence *in situ* hybridization (FISH)

Interphase cells from CLL cultures were analyzed by FISH using the Vysis CLL FISH panel probe set (Abbott Molecular, Illinois, IL, USA) according to manufacturer's instructions.

Arrays

Hybridization to Illumina Genome-wide SNP chips was performed according to manufacturer's protocols found on registration at http://www.illumina.com/products/human1m_duo_dna_analysis_beadchip_kits.ilmn and at http://www.illumina.com/support/array/array_kits/humanomni1-quad_beadchip_kit/documentation.ilmn. The data were processed using GenomeStudioV2009.2 (Illumina, Inc., San Diego, CA, USA) and then analyzed using OncoSNP v1.0 (see below). For visual comparisons, the data were processed also using Nexus 5 Discovery Edition (BioDiscovery, Inc., El Segundo, CA, USA).

OncoSNP analysis

Detailed methods have been described previously[10] and further details are given in the Supplementary Information online. To help identify and exclude germline CNVs of unlikely relevance from our data, we excluded SNP and monomorphic copy number probes in known germline CNV regions from the OncoSNP analysis and we also made use of the database of genomic variants[48] and the Wellcome Trust Case Control Consortium cohort data (see www.wtccc.org.uk).

TP53 mutation analysis

TP53 was screened for mutations using four high-resolution melting assays covering exons 5–8. Following high-resolution melting, results were analyzed on the high-resolution melting module of the Corbett Rotor-Gene 6000 software 1.7 (available for download from QIAGEN, http://www.qiagen.com/corbett/support/default.aspx). Positive PCR products were purified and then sequenced from both strands on the CEQ 8000 (Beckman Coulter, Fullerton, CA, USA).

IGHV mutation analysis

To identify clonal rearrangements of the *IGHV* gene and determine the somatic mutation status of the variable (V) gene sequence in patients we used the IGH Somatic Hypermutation Assay v2.0 (Invivoscribe, La Coutat, France) according to the manufacturer's instructions.

RESULTS

A flowchart outlining the patient numbers and overall study design is given in Supplementary Figure 1 online.

Our preliminary experiments demonstrated the importance of selecting an array platform capable of detecting CNAs and cnLOH (see Supplementary Information, Supplementary Table 1 and Figure 2 online) as we found regions of cnLOH signposting genes that might carry mutations relevant for treatment choice and that would otherwise have been missed using a non-SNP-based platform (for example, *TP53*, see Supplementary Information, Supplementary Table 2 and Figure 3 online). We also demonstrated in these experiments that OncoSNP provided accurate quantification of CNAs, based on the strong correlation between OncoSNP results and conventional FISH analysis (Supplementary Table 3 and Figure 4 online).

Genome-wide analysis of pre-treatment samples

For our main study, we focused initially on the analysis of pre-treatment samples from 93 patients undergoing regular follow-up at our institutions. This cohort was representative of CLL cohorts with a higher proportion of unmutated *IGHV* genes and/or treatment resistant cases that are referred to tertiary centers for treatment consideration (Table 1). A total of 80 patients had one or more CNAs/cnLOHs in well-established regions of prognostic importance - 17p13.1, trisomy 12, 13q14.3, 11q22.3 (Table 1, Supplementary Table 2, Figures 2, 3 and 5 online). CNAs involving chromosome arms 8p, 9p and 10q developed exclusively in patients with unmutated *IGHV*.

We identified 58 previously unreported recurrent CNAs/cnLOHs (Table 2) (see Supplementary Information online for recurrence criteria). We established the minimally deleted regions (MDRs)/minimally overlapping regions (MORs) of the recurrent CNAs/cnLOHs and revealed interesting candidate genes including those implicated in B-cell maturation (for example, *BLIMP1, NFkB2, TLR4,* and *CREBBP)*, tumor progression (for example, *RND3*, *RHOT1; RHDBL*; *RAB20,* and *TRAP1*), DNA damage response (for example, *RIF1, TP53, ATG4D,* and *ATG5),* tumor suppression *(*for example, *NMI, CNOT7*, *PDGFRL, FGF20*, *PI3 K*, and *FOG2)* and familial CLL *(*for example, *SP140L, SP100)* (Table 2). Figure 1 shows examples of four of these novel recurrent regions of interest that span only a few candidate genes, potentially facilitating the identification of driver mutations. Sequencing of the *TP53* gene, the known candidate underlying poor prognosis on 17p13.1, revealed mutations in 14 out of 93 patients and deletions in 11. However, exon sequencing of two of the emergent candidates, *BLIMP1* and *ATG5* revealed no mutations in 40 patients.

In order to establish the clinical significance of our findings, we correlated our array results with measures of clinical prognosis and risk, including *IGHV* mutation status and clinical risk scores defined by necessity for treatment intervention, progression free survival and chemotherapy resistance (see Supplementary Information, online). Genomic complexity was defined either by the presence of three or more CNAs $\geqslant 20$ kb and/or cnLOHs $\geqslant 2$ Mb in addition to, or other than, the known CNA regions assayed routinely by FISH, or by a total length of CNAs $\geqslant 5$ Mb. Within the 93 pre-treatment samples, patients with unmutated *IGHV* genes showed a statistically significant higher total number of CNAs/cnLOHs (50.0% $\geqslant 3$ CNAs/cnLOHs) than patients with mutated *IGHV* genes (16.0% $\geqslant 3$ CNAs/cnLOHs) (Cochran–Armitage trend exact test, P-value $= 0.0032$, Supplementary Table 4 online) and patients with del17p/*TP53* mutation showed a higher total length of CNAs/cnLOHs (90.9% >5 Mb CNAs/cnLOHs) (Cochran–Armitage trend exact test, P-value $= 0.0081$, Supplementary Table 5, online).

Table 1. Characteristics of patients included in the study

Pre-treatment analysis (n = 93)	
Patient Characteristics	*Patient numbers*
At time of blood sample:	
Mean age (range)	66.75 (36 - 87) years
Females	28
Males	65
del11q22.3	17
del13q14.3	47 (28 as sole abnormality)
Trisomy 12	14
del17p13.1	8
None of these	26
TP53 bi-allelic mutations	3
TP53 mono-allelic mutation	3
TP53 mutation/del17p	8
TP53 wt/del17p	1
TP53 mutation negative	68
TP53 mutation not done	10
Mutated *IGHV* genes	25
Unmutated *IGHV* genes	42
Not known	26
W and W (CRS = 0)	16
PFS > 1 year (CRS = 1)	22
PFS < 1 year (CRS = 2)	20
Refractory (CRS = 3)	18
FU < 1 year (CRS = N/A)	15
NK (CRS = N/A)	2
Treatments following blood sampling	
Chlorambucil	16
Chlorambucil + Rituximab or Chlorambucil + Ofatumumab	3
Bendamustine single agent	1
Purine analogue combination	12
Purine analogue combination + Rituximab	11
Ofatumumab single agent	1
Alemtuzumab	1
Chemotherapy refractory	18
NK	14
Number of CNAs/cnLOHs	
None	13
< 3	39
⩾ 3	41
Length of CNAs /cnLOHs	
< 1Mb	32
1 - 5Mb	11
> 5Mb	50

Paired (Pre-treatment and Relapse) analysis (n = 42)		
Patient Characteristics	*Patient numbers*	*Clonal evolution (n = 25)*
del11q22.3	5	4
del13q14.3	21 (13 as sole abnormality)	7
Trisomy 12	7	2
del17p13.1/*TP53* mutation	5	5
None of these	12	5
Mutated *IGHV* genes	13	7[a]
Unmutated *IGHV* genes	25	14[b]
NK	4	4
PFS > 1 year (CRS = 1)	19	10
PFS < 1 year (CRS = 2)	18	10
Chemotherapy refractory (CRS = 3)	5	5

Abbreviations: CNA, copy number alteration; cnLOH, copy neutral loss of heterozygosity; CRS, clinical risk score; FU, follow-up after treatment; *IGVH*, immunoglobulin variable heavy chain gene; NK, not known; PFS, progression free survival; W and W, Watch and Wait. [a]6/7 had % change only. [b]11/14 had newly emerging CNAs/cnLOHs.

In our small cohort, large Type II 13q deletions were not associated with a worse clinical outcome (Supplementary Tables 6 and 7 online).

Furthermore, patients with high clinical risk scores had a higher number of CNAs/cnLOHs than patients with low clinical risk scores (Kruskal–Wallis test, *P*-value = 0.0016) and showed a greater overall length of CNAs/cnLOHs than patients with low clinical risk scores (Kruskal–Wallis test, *P*-value = 0.0002) (Supplementary Table 8, online). Importantly, 16 patients had ⩾3 CNAs/cnLOH but no 17p13.1 or 11q22.3 loss and would not have been picked up by FISH as a poor risk group.

Genome-wide comparison of paired pre-treatment/relapse samples

Next, we focused on the analysis of the 42/93 patients for whom paired samples (that is, both pre-treatment and relapse samples) were available (Table 3). We hypothesized that in addition to recurrent CNAs, newly occurring or expanding genomic aberrations would also represent potential drivers of disease progression. The percentages of subclones carrying CNAs/cnLOH regions were calculated from the SNP data using OncoSNP.[10] Only one case with mutated *IGHV* genes (CLL086) had additional CNAs at relapse. The other changes were percentage changes of pre-existing CNAs. By contrast, 11 unmutated *IGHV* genes cases had complex evolution with emergence of one or more additional CNAs. When we looked at total length of CNAs/cnLOHs, patients without clonal evolution showed a statistically significant lower total length of CNAs/cnLOHs (58.8% < 1 Mb) than patients with clonal evolution (68.0% > 5 Mb)(Cochran–Armitage trend exact test, *P*-value = 0.0075; Supplementary Table 9, online).

All 13q14.3 deletions were present before treatment and of the 21 patients manifesting these; seven exhibited ⩾10% changed proportions of subclones carrying the deletion at relapse. In two cases, (CLL081 and CLL107), the proportion of subclones with the del13q14.3 appeared decreased at relapse and both exhibited additional CNAs elsewhere in the genome. For another two cases (CLL080 and CLL096), with deletions > 1 Mb, the proportion of subclones with the 13q14.3 deletion had increased over time; both had isolated chromosome 13 anomalies (Figure 2). Furthermore, in CLL071 and CLL080, there was extension of the 13q13.4 deletion in one allele (Supplementary Figure 6 online, Supplementary Table 5). Overall, the results suggest that extension of the deleted 13q14.3 locus may confer a clonal survival advantage in the minority of cases without additional CNAs/cnLOHs. By contrast, for patients with complex genomic aberrations, clones containing del13q14.3 are outcompeted.

Patients with del17p13.1/*TP53* mutations were most likely to acquire additional CNAs at relapse (5/5). In addition, three patients without del17p13.1 at diagnosis developed del17p and other CNAs at relapse (CLL084 and CLL108, CLL145). Patient CLL081 was a clear example of this, manifesting a 17p13.1 deletion and multiple other CNAs pre-treatment. At the time of relapse, five CNAs had altered in proportion (Figure 3) and two events (del2q33.1-q37.1 and del16p13) were newly identified. Furthermore, a number of the CNAs identified encompassed newly identified MDRs/MORs (Table 1) and included the genes *LDB-1* and *NFkB2 (10q24.32 MDR), SP140L* and *SP100* (2q37.1 MOR) and *BTBD12, DNASE1, TRAP1 and CREBBP (*16p13.3 MDR). Finally, the OncoSNP feature of defining proportions of subclones carrying CNAs/cnLOHs allowed us to use data from matched pre-treatment and relapse samples to infer likely subclonal populations and to map their evolutionary relationships. Schematic representations of the likely clonal architecture at different time-points for patients CLL081 and CLL106 are shown in Figure 4 with a further example for CLL092 in Supplementary Figure 7 online. These illustrate clearly a non-linear, branching subclonal hierarchy in B-CLL with multiple ancestral subclones, already present in pre-treatment samples and contributing to relapse. In conclusion, using the

Table 2. Newly defined regions showing recurrent/expanding CNAs/cnLOHs

Recurrent CNAs/cnLOH	*No. of pretreatment samples showing overlap*	*No. of new events observed at relapse*	*No. of events with increased percentage at relapse*	*MOR start (bp)*	*MOR end (bp)*	*Genes/microRNAs*
loss/gain/cnLOH 1p36.31p36.23	3	—	—	6 508 255	7 966 180	Many including *THAP3, CAMTA1, TNFRSF9/CD137*
loss/cnLOH 2q31.1	1	1	—	174 294 144	174 955 141	*SP3, OLA1/GTPBP9, SP9, CIR1/CBF1*
loss/cnLOH 2q36.2	3	1	—	225 091 358	225 626 330	*CUL3, DOCK10*
loss/cnLOH 2q37.1	4	1	—	230 940 929	231 244 281	*SP140L, SP100*
loss/gain 3p25.3	2[a]	—	—	9 610 471	10 363 762	Many including *CIDEC*
loss/gain 3p24.3	2[a]	—	1	21 592 730	22 039 016	*ZNF385D*
loss/gain 3p21.31	2[a]	—	—	46 827 053	47 426 525	Many including *PTPN23*
loss/gain 3p21.1	2[a]	1	—	52 038 099	52 553 933	*STAB1/CLEVER1/FEEL1, NT5DC2, C3orf78*
loss 4q13.1–q13.2	2	—	—	64 455 242	66 975 683	*TECRL, LOC401134, EPHA5, hsa-mir-1269, LOC100144602*
loss 4q13.2q13.3	2	—	—	69 847 111	71 783 789	Many including *IGJ, UTP3*
loss 5q21.3q22.1	3	—	—	106 355 299	110 412 993	*EFNA5, FBXL17, FER, PJA2, MAN2A1, LOc100289673, TMEM232, SLC25A46*
loss 5q22.2q22.3	3	—	—	112 371 443	113 463 403	*DCP1, MCC, TSSK1B, YTHDC2*
loss/gain 6p25.3–p25.2	2	2	—	3 138 439	4 294 807	*TUBB2B, PSMG4, SLC22A23, C6orf145, FAM50B, PRPF4B, PEC1*
loss 6q16.3	4	2	—	101 257 101	101 445 547	*ASCC3*
loss 6q21	4	—	—	106 208 162	106 917 920	*BLIMP-1, ATG5*
loss 6q21	3	1	—	109 405 342	109 444 628	*SESN1*
loss7q36	2	1	1	156 694 855	156 824 049	*UBE3C, KIAA0010, MRJ, DNAJB6*
loss 8p22p21.2	3	2	—	16 324 942	20 039 997	Many including *CNOT7, FGF20, MTUS1, PDGFRL*
loss 8p12	2	2	—	33 488 001	36 433 467	*c8orf41, RNF122, DUSP26, UNC5D*
loss 8p11.22p11.21	2	1	—	39 796 895	40 860 746	*ADAM2, IDO1, IDO2, C8orf4, ZMAT4*
loss 8p11.2p11.1	2	1	—	40 947 586	43 533 112	*SFRP1, GOLGA7, AGPAT6, NKX6-3, ANK1, MIR486, MYST3, AP3M2, PLAT, IKBKB, VDAC3, SLC20A2, C8orf40 CHRNB3, CHRNA6, THAP1, HOOK3, RNF170, FNTA, SGK196, POTEA*
loss/gain/cnLOH 8q23.1	3	1	1	106 323 747	106 959 523	*ZFPM2 (FOG2)*
loss/gain 9p22.3–p24.1	5	—	1	11 986 163	14 313 346	*TYRP1, MPDZ, NFIB*
loss/cnLOH 9q33.1	3	—	—	119 325 172	121 100 178	*TLR4, DBC1*
loss/cnLOH 9q33.1–q33.2	3	—	—	121 461 281	122 566 552	*MIR147, CDK5RAP2, MEGF9, FBXW2*
loss/cnLOH 9q34.13–q34.3	3	—	—	134 588 424	139 101 448	Many including *hsa-mir-126*
loss 10p12.33	2	—	—	17 900 269	18 398 113	*FAM23A, MRC1L1, MRC1, MIR511-1, SLC39A12*
loss/cnLOH 10q23.2	2	—	1	89 310 000	89 454 971	*PAPSS2/ATPSK2*
loss 10q23.33	—	2	—	95 437 455	95 462 005	*FRA10AC1*
loss 10q24.32	4	3	—	103 708 264	104 851 056	Many including *LDB1; NFkB2; SUFU; TRIM8*
loss /cnLOH 11q12.2q13.1	2	—	—	61 231 809	64 790 450	Many including *MTA2/MTA1L1/PID*
loss/cnLOH11q24.2q24.3	4	1	—	126 353 410	126 814 884	*KIRREL3, MIR3167*
loss/cnLOH11q24.3	4	1	—	127 852 266	128 613 424	*ETS1, FLI1, KCNJ1, c11orf45, TP53AIP1, ARHGAP32*
loss/cnLOH11q24.3	4	1	—	128 805 080	128 951 253	*BARX2*
loss/gain/cnLOH12p13.33	2[a]	—	—	977 044	1 959 491	*ERC1/ELKS/RAB6IP2, LOC100292680, FBXL14, WNT5B, ADIPOR2, CACNA2D4, LRTM2*
loss/gain/cnLOH12p13.2p13.1	2[a]	—	—	10 176 350	14 248 326	Many including *ETV6*
gain/cnLOH12q13.11q13.12	4[a]	—	—	47 384 207	48 029 106	Many including *MLL2*
loss/gain 12q14.3	2[a]	—	—	64 809 667	64 853 817	*LLPH, S1R, TMBIM4*
loss/gain/cnLOH 12q23.1	4[a]	—	—	99 175 388	99 224 488	*DEPDC4, SCYL2/CVAK104*
loss/gain/cnLOH 12q24.11	4[a]	—	—	109 411 836	109 658 160	*C12orf24, VPS29, RAD9B, PPTC7,(TAPP2C) TCTN1, HVCN1, PPP1CC*
gain/cnLOH12q24.31	4[a]	—	—	119 578 493	121 172 890	Many including *ORAI1 and BCL7A*
loss/cnLOH 13q14.11	4	—	1	39 760 000	39 910 000	*LOC646982/TTL-B1*
loss/cnLOH 13q14.11	6	—	2	40 389 170	40 761 341	*SUGT1L1, ELF1, WBP4, KBTBD6, MTRF1*
loss/cnLOH 13q14.11	6		2	41 687 912	41 826 016	*DGKH, AKAP11/AKAP220*
loss/gain/cnLOH 13q14.2	32	—	1	47 781 909	47 832 800	*RB1*
loss/gain/cnLOH 13q14.2	22	—	1	47 387 659	47 522 891	*SUCLA2, NUDT*
loss/cnLOH 13q33.3–q34	5	1	—	108 839 643	112 014 298	Many including *RAB20, ING1, ARHGEF7*
loss/gain 15q15.1	2	—	—	38 967 655	40 822 359	Many including *NDUFAF1*
loss 16p13.3	1	1	—	3 196 980	4 492 425	Many including *ADCY1, TFAP4, GLIS, DNAJA3, HMOX2, CDIP*
loss/gain 17q11.2	2	1	—	24 979 666	25 506 441	*SSH2, EFCAB5, CCDC55, MIR423, MIR3184*
loss/gain 17q11.2	2	1	—	27 442 088	27 692 902	*RHOT1, ARGFXP2, RHBDL3, C17orf75*
loss/gain 17q12	3	—	—	30 750 407	31 140 027	Many including *AP2B1, RASL10B, MMP28*
gain 17q12	3	—	—	34 817 937	34 948 998	*CDK12, CRKR5, MED1/TRAP220/PBP/PPARBP*
loss/cnLOH 18p11.32p31	4	1	1	1	6 572 104	Many including *USP14, THOC1/nuclear matrix protein p84/HPR1)*

Table 2 (*Continued*)

Recurrent CNAs/cnLOH	*No. of pretreatment samples showing overlap*	*No. of new events observed at relapse*	*No. of events with increased percentage at relapse*	*MOR start (bp)*	*MOR end (bp)*	*Genes/microRNAs*
gain 19p13.2	2[a]	1	—	10 071 486	10 926 409	Many including *ATG4D, hsa-mir-1181, hsa-mir-1238*
loss 19p13.11	1[a]	1	—	18 336 675	19 169 002	Many including *SSBP4, FKBP8, UBA52, CRTC1, BMP family (GDF1, LASS1), ELL*
loss/ LOH loss19q13.12q13.13	2[a]	—	—	41 440 201	43 149 438	Many including many *ZNF* genes
gain/cnLOH 20q13.13	4	—	1	47 292 495	47 500 772	*DDX27, ZNFX1, SNORD12C, SNORD12B, SNORD12, C20orf199, NCRNA00275, AK055386*
loss/cnLOH 20q11.23	4	—	1	34 645 412	34 974 353	*TGIF2, C20orf24, SLA2, NDRG3, DSN1, C20orf117, C20orf118, SAMHD1*
loss 20q11.23	5	—	1	35 010 463	35 020 688	*SAMHD1*
loss/gain 22q12.1	2	1	—	26 427 496	26 490 006	*MN1*
loss/gain 22q12.1q12.2	2	—	—	27 573 410	27 918 140	*ZNRF3, C22orf31, KREMEN1*
loss/gain/cnLOH Xq27.3	3	1	—	142 536 091	144 254 333	*SLITRK4, SPANXN2, SPANXN1*

Abbreviation: CNA, copy number alterations; cnLOH, copy neutral loss of heterozygosity regions; MOR, minimally overlapping region. All co-ordinates are hg36. Shaded blue results show the only genes/microRNAs mapping within the identified MORs. [a]Excluding full trisomies.

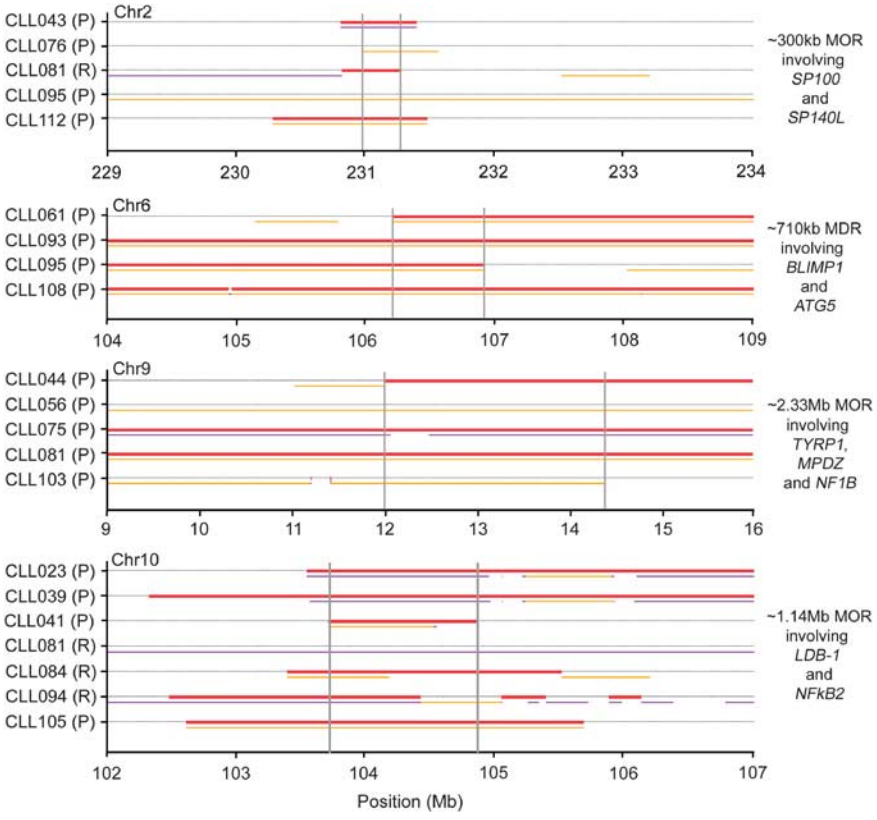

Figure 1. Four recurrent regions of interest observed in B-CLL patient samples. The boundaries of the MDRs/MORs, 2q37.1, 6q21, 9p22.3p24.1 and 10q24.32, are indicated by the vertical lines. Red lines show called losses, green lines show gains, purple lines show allelic imbalance and mustard lines show cnLOH. In each case the MDR/MOR was verified by visual inspection of Log ratio and B-allele frequency plots. P = Pre-treatment, R = Relapse.

combination of high-resolution SNP array and OncoSNP, we were able to detect clonal evolution in 60% of treated patients with B-CLL.

Interestingly, when comparing the regions affected by clonal expansion with recurrent regions we noticed a considerable overlap (Table 3). Many regions that were both recurrent and expanded or newly occurred in relapse samples included a very limited number of genes (Table 2, blue shading). These regions include genes that have a role in familial CLL (*SP140L*, *SP100*), in B-cell development and autophagy (*BLIMP1* and *ATG5*, respectively) and *FOG2*, a regulator of phosphatidylinositol 3-kinase, which is

Table 3. Comparison between CNA and cnLOH events noted in sequential pre-treatment versus relapse samples[a]

Patient no.	*Known CNAs*	*IGHV gene status*	*CRS*	*CNAs and cnLOH regions in pre-treatment sample*	*Differences at relapse*
CLL010	None	Unmutated	1	None noted	None noted
CLL023	None	Unmutated	1	46,XX, arr 4q24(107,162,437-107,237,623)x3[b], 40% 10p12.23(17,833,779-18,398,113)x1, 40% 10q24.32q25.2 (103,562,915-112,637,185)x1	46,XX, arr 10p12.23(17,833,779-18,398,113)x1 no longer detectable, ~10% 10q24.32q25.2 (103,562,915-112,637,185)x1
CLL035	Tri12	Mutated	1	47,XX, arr 3p24.3(21,592,730-22,039,016)x1[b], 12(0-132,146,663)x3	None noted
CLL050	del13q14.3	Unmutated	1	46, XX, arr 6q14.1 (83,165,917-83,265,032)x1[b], 7p21.1(18,576,853-18,602,061)x3[b], 11q12.3 (62,082,265-62,330,255)x1[b], 13q14.2q14.3 (48,608,987-51,326,824)x1	None noted
CLL068	del13q14.3	Mutated	1	46,XY, arr 90% 2q22.2q24.1(144,240,156-157,961,129)x1, 13q14.2q14.3(47,595,821-50,072,312)x1	46,XY, arr >90% 2q22.2q24.1(144,240,156-157,961,129)x1
CLL080	del13q14.3	Mutated	1	46,XX, arr 3q25.31(156,553,159-156,572,680)x1[b], 30% 13q14.11q14.3(40,674,362-49,342,922)x1[b], ~50% 13q14.3(49,342,923-50,332,192)x1[b]	46,XX, arr 80% 13q14.11q14.3(40,691,755-50,580,804)x1 (extended across region)
CLL083	None	Unmutated	1	46,XX, arr 4p14(39,001,471-39,049,059)x3[b], 11q25(131,123,910-131,186,533)x3[b], <10% 12q14.3(64,809,667-64,853,817)x1[b], 12q23.1(99,175,388-99,224,488)x1[b],10% 16p12.2(20,710,305-20,790,857)x3[b], 17q12(34,817,937-34,948,998)x3[b]	46,XY, arr 11q25(131,123,910-131,186,533)x3[b] no longer detectable. 12q14.3(64,809,667-64,853,817)x1[b] no longer detectable, 12q23.1(99,175,388-99,224,488)x1[b] no longer detectable, 16p12.2(20,710,305-20,790,857)x3[b] no longer detectable, 17q12(34,817,937-34,948,998)x3[b] no longer detectable
CLL085	del13q14.3	Unmutated	1	46,XY, arr 10q23.1q23.31(85,670,734-89,454,971)x2 hmz, 50% 13q14.11q21.31 (43,934,565-61,368,181)x1	46,XY, arr 10q23.33(95,437,455-95,462,005)x1[b]
CLL089	None	Unmutated	1	46, XY, arr 1q21.1(144,335,130-144,614,943)x3[b]; 1q25.2(174,495,962-174,547,429)x3[b]	None noted
CLL090	tri12 del13q14.3	Mutated	1	47,XX, arr 90% 7q33q34(134,397,353-139,656,577)x1, 90% 7q36 (156,694,855-156,824,049)x1[b], 50% 11q12.1qter(60,468,463-134,445,626)x2 hmz, 80% 12(0-132,146,663)x3, 80% 13q14.3(49,296,907-50,580,804)x1	47,XX, arr >90% 7q33q34(134,397,353-139,656,577)x1, >90% 7q36(156,694,855-156,824,049)x1[b], 90% 13q14.3(49,296,907-50,580,804)x1
CLL091	None	Mutated	1	46,XY, arr 14q24.3q31.1(77,490,883-80,987,659)x2 hmz[c]	None noted
CLL093	del13q14.3	Mutated	1	46,XY, arr 80% 6p25.3p25.2-(94,609-4,294,807)x3, 6q14.1-qter(83,755,489-170,753,209)x1, 70% 10q25.1(109,682,821-109,770,215)x3[b], 13q14.2-q14.3(47,556,552-50,610,384)x1, 70% 20q13.13(47,292,495-47,500,772)x3[b]	None noted
CLL096	del13q14.3	Mutated	1	46,XX arr 40% 13(0-114,121,252)x2 hmz, 13q14.3(49,425,974-49,522,141)x1, 13q14.3(49,522,141-50,380,713)x0	46,XX, arr 90% 13(0-114,121,252)x2 hmz. 13q14.3 (49,425,974-49,522,141)x1 no longer detectable
CLL097	del13q14.3	Mutated	1	46, XY, arr >90% 13q14.11q14.3(40,332,062-51,453,289)x1	46,XY, arr 90% 13q14.11q14.3(40,332,062-51,453,289)x1
CLL098	None	Unmutated	1	None noted	None noted
CLL099	None	Unmutated	1	46,XY, arr 70% 1q21.1(144,099,494-144,757,265)x3[b]	None noted
CLL101	de11q22.3 del13q14.3	Unmutated	1	46,XY, arr 11q14.1qter(77,080,067-134,445,626)x1, 60-70% 13q14.2q21.32(47,781,909-64,410,990)x1[b], 50% 13q21.32q33.2(64,350,990-103,707,225)x3[b], 70% 22q11.21qter(17,403,754-49,582,267)x3	None noted
CLL107	de11q22.3; del13q14.3	Unmutated	1	46,XY, arr 50% 3p21.31(46,827,053-50,133,778)x1, 50% 3p21.1(52,038,099-52,553,933)x1[b], >90% 11q14.3q24.3(91,864,846-127,805,084)x1, >90% 13q14.11(40,389,170-40,761,341)x1[b], >90% 13q14.11(41,687,912-41,826,016)x1[b], >90% 13q14.2q14.3(47,515,748-51,333,756)x1	46,XY, arr 10% 3p21.31(46,827,053-50,133,778)x1[b], 10% 3p21.1(52,038,099-52,553,933)x1 [b], 80-90% del11q14.3q24.3 (91,864,846-127,805,084)x1, 90% 13q14.11 (40,389,170-40,761,341)x1[b], 90% 13q14.11 (41,687,912-41,826,016)x1[b], 90% del13q14.2q14.3 (47,515,748-51,333,756)x1
CLL108	None	Unmutated	1	46,XX, arr 3q22.3(137,678,290-137,832,637)x1[b], 80% 6q14.1q22.33(81,463,977-129,051,286)x1, 90% 11q12.3qter(62,398,575-134,445,626)x2 hmz, 80% 12pterp12.2(0-20,356,457)x2 hmz, 80% 12p12.2q21.31(20,356,457-84,251,286)x3, 80% 12q21.31qter(84,251,286-132,146,663)x2 hmz	46,XX, arr 70% 6q14.1q22.33(81,463,977-129,051,286)x1, 70% 12(0-132,146,663)x2 hmz, 30% 17pterp13.1(0-8,033,419)x1, 12p12.2q21.31 (20,356,457-84,251,286)x3 no longer detectable
CLL003	de11q22.3 del13q14.3	Unmutated	2	46,XY, arr 11q22.3(104,158,306-109,733,134)x1, 11q23.1q23.2(110,556,535-114,089,910)x1,11q23.3(chr11:115,527,657-116,182,933)x1, 13q14.3(49,038,530-	46,XY, arr 5q12.1(58,925,836-58,976,604)x1

Table 3 (*Continued*)

Patient no.	*Known CNAs*	*IGHV gene status*	*CRS*	*CNAs and cnLOH regions in pre-treatment sample*	*Differences at relapse*
				50,906,239)x1, 17q21.33q22(45,720,750-49,017,380)x2 hmz[c]	
CLL006	tri12	Unmutated	2	47,XY, arr 90% 4q31.1(153,631,568-153,935,927)x1, 70% 12(0-132,146,663)x3, 17p12(13,118,349-13,171,201)x3[b]	None noted
CLL044	del13q14.3 de17p13.1	Mutated	2	46,XY, arr 46,XY, arr 2q36.2 (225,094,102-225,617,560)x1, ~20% 9p23p21.1(11,986,163-32,324,220)x1[b] , 50% 9q13qter(70,223,358-139,946,600)x3, 13q14.3(49,515,641-51,327,198)x0, 17pterp11.2(0-21,430,683)x1, 19q13.12q13.2(41,320,887-43,412,844)x2 hmz[c]	46,XY, arr ~30% 9p23p21.1(11,986,163-32,324,220)x1[b], <20% 9q13qter(70,223,358-139,946,600)x3[b]
CLL056	del13q14.3	NK	2	46,XY, arr 80% 5p13.3(37,408,772-37,784,135)x3[b], 80% 9pter9p11.2(0-43,254,687)x2 hmz[c], 70% 13q14.12q14.2(45,377,644-45,760,335)x3, 70% 13q14.2(46,403,477-47,574,445)x3, >90% 13q14.2(47,583,917-48,099,556)x1, 90% 13q14.2q14.3(48,719,245-50,523,975)x1, 70% 13q14.3qter(50,523,975-114,121,252)x3	46,XY, arr 30% 4pterp15.1(0-35,166,759)x1[b] , 90% 9pter9p11.2(0-43,254,687)x2 hmz[c], ~50% Xq26.2q28 (131,629,636-154,058,617)x2[b]
CLL063	de11q22.3 del13q14.3	Unmutated	2	46, XX, arr 11q14.1q24.1(77,652,729-122,488,138)x1, 13q14.11(39,758,892-39,909,885)x0[b], 13q14.3(49,425,974-50,389,210)x0[b], 13qcenqter(19,075,593-114,142,980)x2 hmz[c]	None noted
CLL079	None	Unmutated	2	None noted	46, XX, arr 6p25.3(746,618-789,781)x3[b], 6p22.3(19,723,828-19,765,153)x3[b], ~40% 6q16.3(101,257,101-101,307,223)x3[b], 6q21(109,405,342-109,444,628)x1[b]
CLL082	del13q14.3	Mutated	2	46, XX, arr 13q13.2-21.2(33,726,200-53,619,700)x1, 13q14.3(49,461,448-50,635,061)x0	None noted
CLL084	None	Unmutated	2	46,XY, arr 4q34.1(172,446,717-172,482,835)x1[b]	46,XY, arr <10% 6pterp22.3(0-17,028,648)x1[b],<10% 10q24.31qter (102,668,100-135,374,737)x1[b], <10% 11q14.1q23.2 (78,372,164-114,844,825)x1[b], <10% 17pterp11.2(0-18,877165)x1[b], <10% 18pterpcen (0-13,907,055)x1[b], <10% 18q21.1q21.2(41,130,378-50,155,862)x1[b], <10% 18q21.33q22.1 (59,798,490-62,065,418)x1[b], <10% del18q22.2qter (66,972,378-76,117,153)x1[b]
CLL086	None	Mutated	2	46,XY, arr 3p26.3p26.1(0-7,571,446)x1, >90% Xq27.1qter(139,443,330-154,582,606)x2	46,XY, arr 11q14.3 (89,597,760-89,690,254)x3[b], 12q24.23(118,811,967-118,875,247)x3[b], 22q12.1(26,427,496-26,490,006)x3[b]
CLL087	tri12	Unmutated	2	47,XY, arr 80% 12(0-132,146,663)x3, Xp11.3p11.4(41,074,455-43,983,367)x0	None noted
CLL088	del13q14.3	Mutated	2	46,XX, arr >90% 13q14.12q14.3(45,817,763-49,547,499)x1	None noted
CLL095	tri12	NK	2	47,XY, arr 80% 2q14.3qter(129,537,170-242,497,808)x2 hmz, 6q14.1q21(79,968,009-106,917,920)x1, 70% 12(0-132,146,663)x3, >90% 20q11.2qter (30,155,027-62,435,964)x2 hmz[c] , 22q11.23(35,010,463-35,020,688)x0[b]	46,XY, arr 2p14(chr2:67,974,461-68,005,924)x3[b]
CLL102	tri12	Unmutated	2	47,XY, arr 80% 12(0-132,146,663)x3	None noted
CLL103	tri12	Unmutated	2	47,XX, arr 1q25.3(183,360,633-183,397,169)x3[b], >90% 2q22.1q31.1(141,094,077-171,777,386)x2 hmz, >90% 9p24.1p22.3(9,000,336-14,313,346)x2 hmz, 10q26.3(135,089,371-135,374,737)x3[b], 80% 12(0-132,146,663)x3, >90% 13q33.3q34(108,839,643-112,014,298)x2 hmz, 70% 22q11.21q12(17,295,655-19,791,274)x3	None noted
CLL106	del17p13.1	Unmutated	2	46,XY, arr 90% 1p34.2(41,538,823-42,112,410)x1, 90% 6p25.3p25.1(1,376,135-6,173,214)x1, 90% 6p25.1p24.3(6,471,929-7,578,657)x1, 90% 6p24.3p24.1(10,155,458-11,223,334)x1, 90% 6p24.1p22.3(12,754,522-18,670,995)x1, 90% 6p21.2(38,889,727-39,544,026)x1, 90% 6p21.1p12.3(43,689,563-46,299,596)x1,90% 7p15.3(23,346,710-24,400,934)x1, 90% 9q33.3 (127,381,099-127,900,525)x1, 90% 12p13.33 (977,044-1,959,491)x1, 17pterpcen(0-22,125,162)x1, 18pterp11.31(0-6,572,104)x1, 18p11.31p11.23(6,802,509-7,333,763)x3,	46,XY, arr 80% 1p34.2(41,538,823-42,112,410)x1 , 60% 3p26.3p26.1(0-8,686,575)x1, 70% 3p25.1p22.3(12,214,891-34,481,950)x1, 70% 3p21.33p21.31(43,708,188-46,679,233)x1, 70% 3p21.1p14.2(52,515,813-61,534,811)x1, 70% 3p14.1p13(68,801,809-74,909,820)x1, 60% 8pterq11.21(0-49,228,092)x1, 60% 8q12.1q13.2 (57,011,940-70,263,279)x1, 60% 8q23.1 (106,323,747-106,959,523)x1, 90% 17pterpcen(0-22,125,162)x1, 90% 18pter18p11.31(0-6,572,104)x1, 90% 18p11.31p11.23(6,802,509-7,333,763)x3, 90% 18p11.22p11.21(8,926,107-

Table 3 (*Continued*)

Patient no.	*Known CNAs*	*IGHV gene status*	*CRS*	*CNAs and cnLOH regions in pre-treatment sample*	*Differences at relapse*
				18p11.22p11.21(8,926,107-13,281,880)x1, 18p11.21(13,536,937-15,348,275)x1, 90% 22q11.21(16,712,510-17,055,914)x1[b], 90% 22q11.23q12.1(23,409,925-27,331,789)x1, 90% 22q12.1q12.2(27,573,410-27,918,140)x1[b]	13,281,880)x1, 90% 18p11.21(13,536,937-15,348,275)x1, 80% 22q11.21(16,712,510-17,055,914)x1[b], 80% 22q11.23q12.1(23,409,925-27,331,789)x1, 80% 22q12.1q12.2(27,573,410-27,918,140)x1[b]
CLL145	de17p13.1	Unmutated	2	46,XY, arr <10% 3p25.1p14.3(12,516,876-55.709725)x1[b] , <10% 5q14.3q23.3(87,285,075-129,855,075)x1[b] , >90% 6p24.3p22.3(10,220,342-16,992,531)x2 hmz[b], 6q16.3(100,876,399-101,445,547)x1, 80% 8q22.3q24.21(106.058,337-128,286,818)x2 hmz, 80% 8q24.21(128,286,819-128,297,901)x3[b], 80% 8q24.21-q24.3(128,297,902-147,274,826)x2 hmz[b], 12q14.1q14.2(56,925,673-62,754,801)x2 hmz[b], 16q22.3q23.1(72,290,299-76,525,557)x2 hmz[b], <10%-50% 17p13.3p11.2(0-18,485,225)x2 hmz[b], ~30% 18q11q23(22,045,087-76,115,172)x2 hmz[b], 80% 22q11.23(,23,991,557-24,250,622)x1[b]	10-20% 3p25.1p14.3(12,516,876- 55.709725)x1[b] , 10-20% 5q14.3q23.3(87,285,075-129,855,075)x1[b] , 50% 8q22.3q24.21(106.058,337-128,286,818)x2 hmz, 50% 8q24.21(128,286,819-128,297,901)x3[b], 50% 8q24.21-q24.3 (128,297,902-147,274,826)x2 hmz[b], <10%-40% 17p13.3p11.2(0-18,485,225)x2 hmz[b], ~20% 18q11q23(22,045,087-76,115,172)x2 hmz[c],
CLL156	None	NK	2	46,XY, arr 80% 5q23.3 (124,282,026-124,531,381)x3[b], ~10% 19q13.12q13.13(41,440,201-43,149,438)x1 hmz[b]	None noted
ARR003	del13q14.3	Unmutated	2	46,XY, 13q14.2q14.3(49,983,042-51,331,787)x1[b]	46,XY, arr 50% 11q13.1q25(66,992,701-134,177,292)x2 hmz[b], 90% 13q14.2q14.3(49,970,598-51,333,659)x1[b],
CLL071	del13q14.3	Mutated	3	46,XY, arr <10% 13q13.3q34(36,483,665-114,123,125)x1[b], 70% 13q14.2q14.3(48,713,406-50,479,598)x1[b], 13q14.3(49,229,284-50,423,376)x0, 30-70% 17pterp11.2(0-21,100,678)x2 hmz	46,XY, ~20% 13q13.3q14.2(36,446,865-114,123,122)x1[b].
CLL081	del13q14.3 de17p13.1	Unmutated	3	46,XY, arr <10% 2pterp14(0-67,981,586)x3[b], 90% 2q11.1q11.2(95,597,687-97,676,314)x1, 90% 2q14.2q14.3(121,572,920-126,531,524)x1, 90% 2q14.3(127,036,406-128,569,251)x1, 90% 2q21.2q24.1(134,752,138-155,542,922)x1, 40% 8q22.2qter(100,925,155-146,263,890)x3, 90% 9pterp11.2(0-33,318,812)x1, 90% 9p13.3p11.2 (33,831,777-43,548,708)x1, >90% 9q21.13 (74,372,294-74,516,002)x1[b], <10% 10q23.2q23.31 (89,306,467-93,720,035)x1[b], 90% 11q12.2q13.1 (61,231,809-64,790,450)x1, 70% 12q13.11q13.12 (47,308,669-48,029,106)x3, 50% 13q11q13.1 (17,920,393-32,659,455)x1, 50% 13q14.3 (48,898,975-50,805,219)x1, 90% 17pterpcen(0-22,125,162)x1, 70% 17qcenqter(22,713,949-78,637,198)x3; 90% 19pterp13.2(0-10,071,486)x1; 70% 19p13.2p13.13(10,071,486-13,419,854)x3	46,XY, arr ~50% 2pterp14(0-67,981,586)x3[b], >90% 2q11.1q11.2(95,597,687-97,676,314)x1, >90% 2q14.2q14.3(121,572,920-126,531,524)x1, >90% 2q14.3(127,036,406-128,569,251)x1, >90% 2q21.2q24.1(134,752,138-155,542,922)x1, ~10% 2q31.1(174,294,144-174,955,141)x1[b], ~30% 2q33.1q37.1(204,070,804-231,244,281)x1[b], 90% 8q22.2qter(100,925,155-146,263,890)x3, >90% 9pterp11.2(0-33,318,812)x1, >90% 9p13.3p11.2(33,831,777-43,548,708)x1, 40% 10q23.2q23.31(87,419,694-107,982,339)x1, >90% 11q12.2q13.1(61,231,809-64,790,450)x1, ~10% 13q11q13.1(17,920,393-32,659,455)x1, <10% 13q14.3(48,898,975-50,805,219)x1[b], ~30% 16p13.3(3,407,330-4,492,425)x1[b] , >90% 17pterpcen(0-22,125,162)x1, >90%19pterp13.2(0-10,071,486)x1
CLL092	de11q22.3 del13q14.3	Unmutated	3	46,XY, arr 2p16.1(55,330,268-55,352,382)x1[b], 11q22.3q24.3(103,008,248-127,852,266)x1, 13q14.3(49,500,283-50,255,794)x1, ~20% 20q11.22qter(32,772,470-62,385,675)x2 hmz[c]	46,XY, ~10% 11q22.1(101,354,200-101,514,796)x1[b], arr 40% 20q11.22qter(32,772,470-62,385,675)x2 hmz
CLL094	de17p13.1	Unmutated	3	46,XY, arr 30% 2q33.2q36.3(206,690,791-230,780,953)x1, 40% 17pterp11.2(0-20,067,160) x1[b], 60% 17p11.2(20,067,161-21,471,049)x1[b]	46,XY, arr 20% 2pterp21(0-45,928,191)x1, 20% del2p16.32p16.1(48,617,246-59,276,760)x1, 20% 2p13.2p11.2(73,382,585-83,989,328)x1[b], 2q33.2q36.3(206,690,791-230,780,953)x1 (no longer detectable); 20% 6p24.3p25.2(3,138,439-9,391,113)x1, 20% 6p22.3(16,521,154-19,237,784)x1, 20% 6p21.31(35,154,526-36,820,525)x1[b], 20% 6p21.2(37,804,382-38,908,035)x1[b], 20% 6p21.1q12(48,685,550-67,685,569)x1[b], 20% 6q13q15(74,279,874-88,065,764)x1, 20% 6q21q22.1(111,800,880-116,278,896)x1[b], 20% 6q22.33(127,739,685-130,160,593)x1, 20% 7q21.11(80,379,709-85,591,711)x1, 20% 7q31.1q32.1(111,820,986-127,193,285)x1, 20% 7q32.3qter(130,945,628-157,796,605)x1, 20% 8pterp12(0-36,598,375)x1[b], 20% 10q24.1qter(98,893,384-135,085,880)x1, 50% 11q13.1q25 intermittent(63,515141-131,026,292)x3[b], 70% 17q11.2(24,979,666-25,506,441)x1, 70% 17q11.2(27,442,088-

Table 3 (*Continued*)

Patient no.	*Known CNAs*	*IGHV gene status*	*CRS*	*CNAs and cnLOH regions in pre-treatment sample*	*Differences at relapse*
					27,692,902)x1[b], 60% 19p13.3p13.11(5,676,493-18,095,588)x3, 70% 19p13.11p12(18,201,910-24,337,497)x1
CLL112	del13q14.3	NK	3	46,XYarr 90% 2q35(218,476,667-220,204,673)x1, 90% 2q36.3q37.1(230,239,060-231,446,513)x1, 8q21.11q21.13(78,181,345-80,471,188)x2 hmz[c], >90% 13q11.2qter(20,381,113-113,908,097)x2 hmz, 13q14.3(49,494,577-50,238,316)x0, 80% 16p12.1(21,860,290-22,350,261)x1	46,XYarr 90% 2q35(218,476,667-220,204,673)x1 no longer detectable, 90% 2q36.3q37.1 (230,239,060-231,446,513)x1 no longer detectable, >90% 13q11.2qter(20,381,113-113,908,097)x2 hmz no longer detectable, 13q14.3(49,494,577-50,238,316)x0 no longer detectable

Abbreviations: CNA, copy number alteration; cnLOH, copy neutral loss of heterozygosity; CRS, clinical risk score; hmz, homozygous; *IGVH*, immunoglobulin variable heavy chain gene; NK, not known. All co-ordinates are hg36 and determined using OncoSNP except [c] (see above). 'x2 hmz' indicates cnLOH. [a]Using size thresholds of 20 kb for CNAs (unless in a known region of interest) and 2 Mb for cnLOH. Percentages represent levels of mosaicism estimated from B-allele frequency values. Where a percentage is not noted, then all cells appear to carry the anomaly. [b]Co-ordinates determined using Nexus (because of variable CNA boundaries less well defined by OncoSNP or <500 kb in size or below the level of detection of OncoSNP). [c]Called by OncoSNP as possible germline variant and manually verified as probable somatic event. Co-ordinates determined using Nexus.

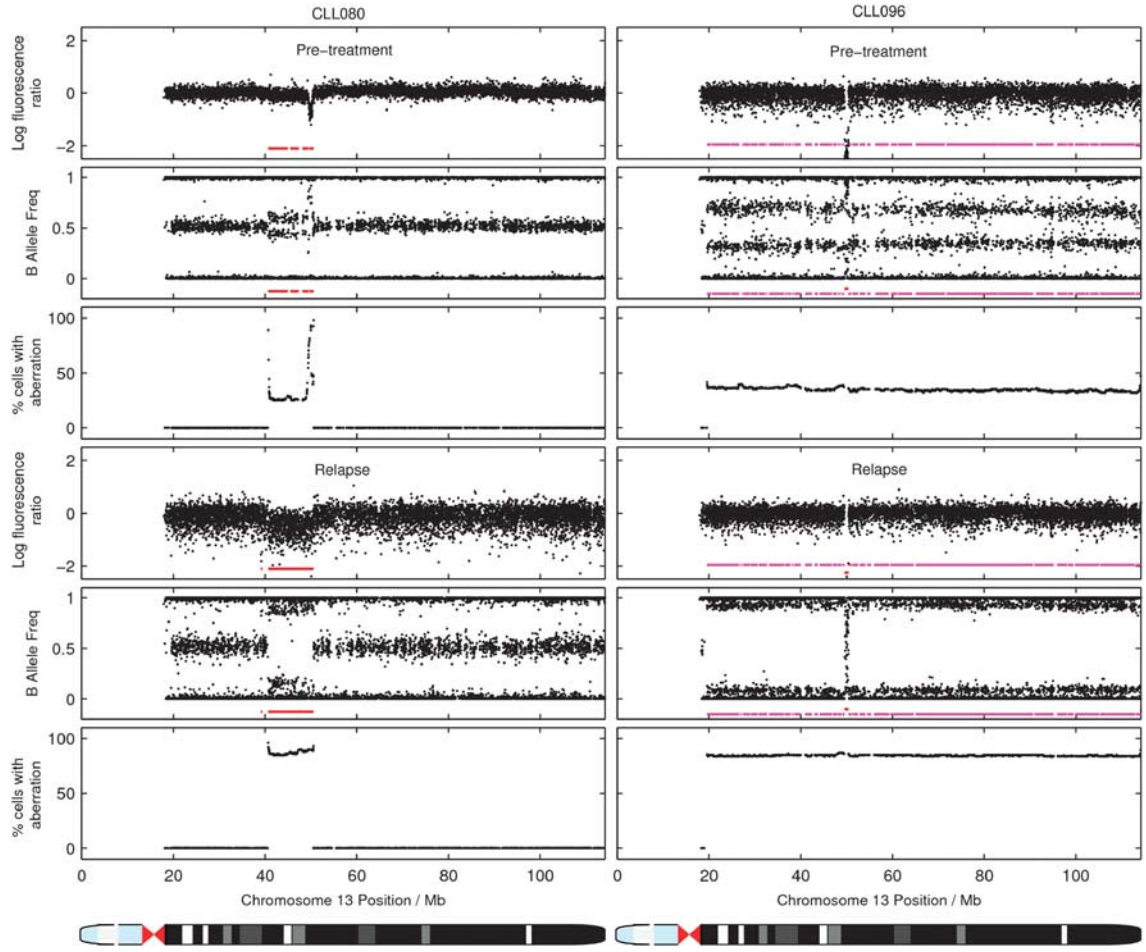

Figure 2. OncoSNP output showing clonal expansion events involving the chromosome 13q14.3 MDR and percentages of cells involved in pre-treatment and relapse samples from CLL080 and CLL096. The red lines indicate deletion events, whereas the magenta lines indicate allelic imbalance/cnLOH. In CLL080, 30% cells carry a large 13q14.11q14.3 deletion and an additional ~20% cells carry the smaller MDR deletion at pre-treatment. At relapse, ~90% cells carry the larger deletion. For CLL096, there is a homozygous deletion of 13q14.3, and cnLOH of the entire chromosome in ~40% cells at pre-treatment. This cnLOH expands to ~90% cells at relapse.

involved in cancer proliferation and survival. A fourth region affects *NFIB* involved in the *NFKB* pathway and two genes of no obvious CLL related annotation (*TYRP1*, *MPDZ*). A further region included DNA repair genes (*BTBD12*, *DNASE1*), *TRAP1*, a mitochondrial chaperone and regulator of apoptosis, as well as *CREBBP*, a frequently mutated gene in B-cell lymphomas.

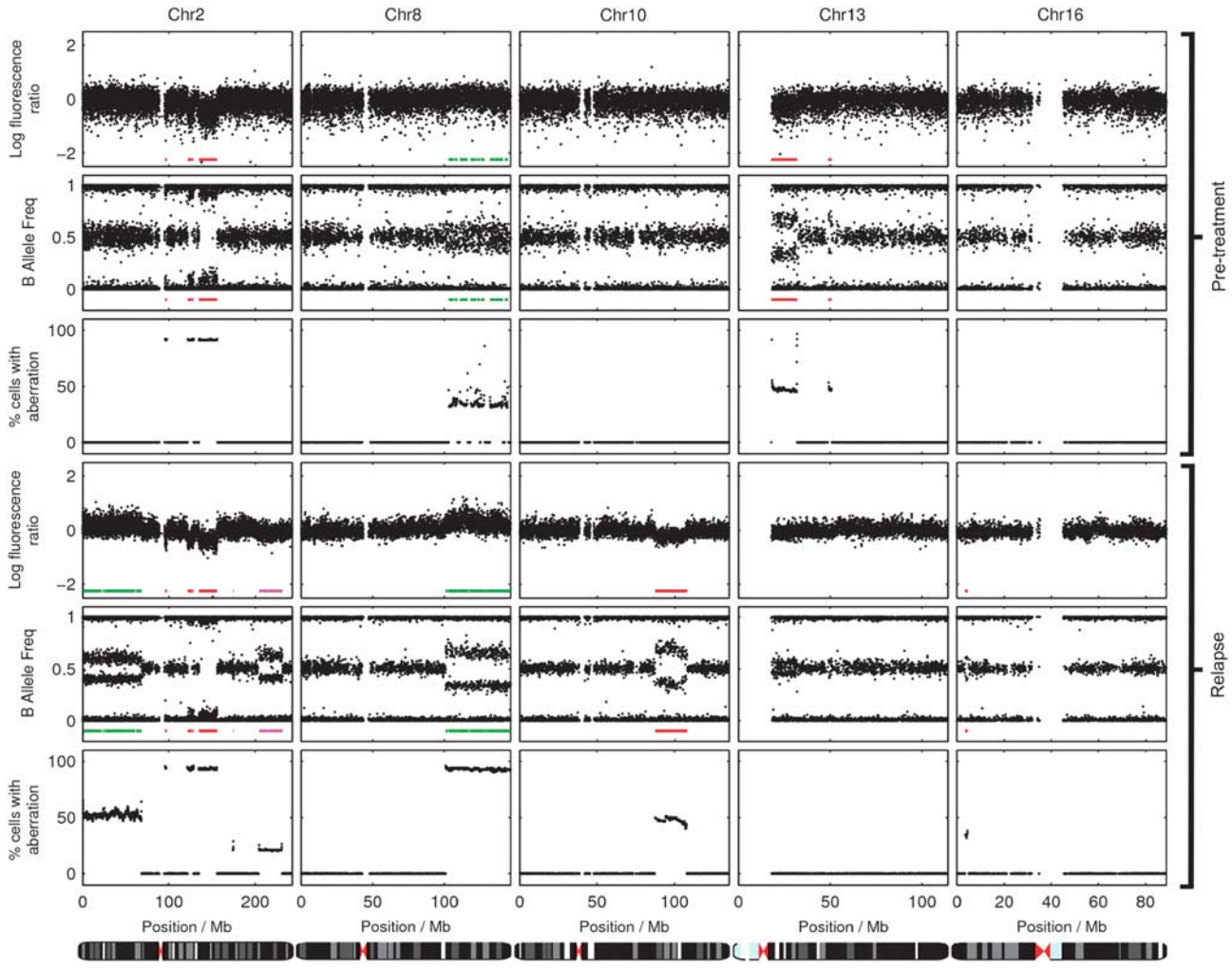

Figure 3. OncoSNP output showing clonal evolution in pre-treatment and relapse samples from CLL081. This patient shows a reduction in the percentage of cells carrying the 13q abnormalities whilst other CNAs expanded (8q22.2qter and 10q23.2q23.3) or were newly identified at relapse (2q33.1q37.1 and 16p13.3).

When we performed pathway analysis of the 546 genes in the MDRs/MORs affected by recurrent and emerging/expanding CNAs, we identified 17 over-represented, statistically significant and independent pathways of which eight contained 55 cancer-related genes (phagosome *P*-value = 0.0024; apoptosis *P*-value = 0.0079; small-cell lung cancer *P*-value = 0.0232; prostate cancer *P*-value = 0.0284; pancreatic cancer *P*-value = 0.0355; chronic myeloid leukemia *P*-value = 0.0414; wnt signaling *P*-value = 0.0436; cell cycle *P*-value = 0.0497) (see Supplementary Table 10, online).

DISCUSSION

Our study is the first genome-wide array based analysis of CNAs/cnLOHs that both characterizes and quantifies the proportion of subclones carrying genomic changes before treatment and at subsequent relapse in cancer. The results demonstrate that (i) many CNAs/cnLOHs in CLL are recurrent and therefore non-random events that expand over time owing to Darwinian selective pressure, (ii) selected genes identified in both well-recognized and newly defined MDRs/MORs present plausible candidates for driving disease progression and include those involved in mature B-cell development, DNA damage response (DNA repair, apoptosis and autophagy), tumor progression and familial B-CLL; pathway analysis was consistent with this. The identification of genes in this way will be informative for focused sequencing strategies. (iii) In our cohort there is strong supportive evidence that genomic complexity is associated with poor risk disease. This adds strength to previous studies that indicated a link between genomic complexity or clonal evolution to survival in CLL.[30,36,38] (iv) OncoSNP is a powerful computational statistical tool that alters fundamentally our ability to investigate clonal architectures and tease apart complex clonal dynamics. Furthermore, our results demonstrate the importance of developing comparable algorithms for application to next generation sequencing data. Although it is likely that in the longer term array technology will be replaced by whole genome sequencing, this study contributes to our understanding of the significance of CNAs/cnLOH in cancer progression. Finally, our results do not support the existence of a single leukemia propagating cell population but instead, shed light on the considerable genomic heterogeneity of cells driving disease progression.

Clinically, the existence of multiple, genetically distinct, subpopulations that escape therapeutic intervention presents formidable challenges for the development of effective treatments for patients with relapsed refractory B-CLL. Therefore, detailed characterization of the molecular basis of the condition

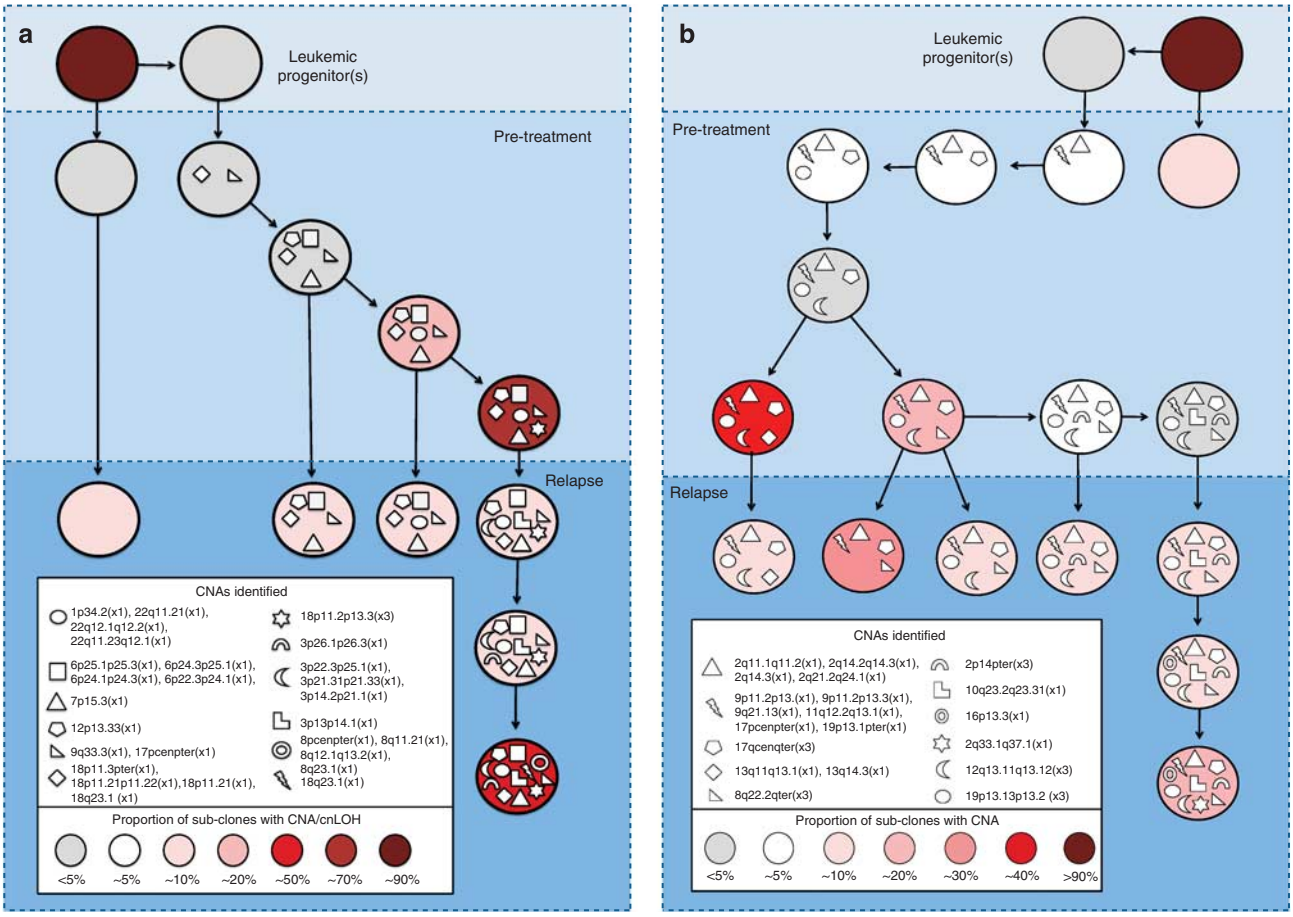

Figure 4. Schematic representation of the possible clonal architecture pre-treatment and at relapse (pale blue and mid-blue shaded areas, respectively) for patients (**a**) CLL106 and (**b**) CLL081. Proportions below 10% and above 90% and cell populations without identifiable CNAs/cnLOHs are inferred. The precision of the percentage contribution of subclones with CNAs/cnLOH is ±10%. Note that for CLL081, the proportion of subclones with the 13q14.3 loss decreases at relapse and the 12q13.11q13.12 and 19p13.13p13.2 gains are no longer observed in one subclone.

and elucidation of the underlying mechanisms of clonal diversity will be essential for future targeted selection of effective therapeutic agents.

CONFLICT OF INTEREST

We would like to declare that OncoSNP is freely available for use in academic or non-profit research activities upon request. Isis Innovation Ltd may request a fee for the use of the software in commercial or other profit generating activities. It is the policy of Isis Innovation Ltd that a share of the revenue generated from such licensing arrangements would be distributed via the University of Oxford to CY and CCH.

ACKNOWLEDGEMENTS

This work was supported by the NIHR Biomedical Research Centre, Oxford with funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. In addition, the publication presents independent research commissioned by the Health Innovation Challenge Fund (HICF-1009-026), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health. This study makes use of data generated by the Wellcome Trust Case - Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. The project was also supported by the Wellcome Trust (090532/Z/09/Z), (076113), (085475) and by Illumina, Inc. CY is funded by a UK Medical Research Council Specialist Training Fellowship in Biomedical Informatics (Ref No. G0701810). We would also like to acknowledge Dr David Bentley (Illumina Inc) and Prof Ian Tomlinson (WTCHG) for their role in supporting this work.

AUTHOR CONTRIBUTIONS

SK designed and performed the research, analyzed the data and wrote the manuscript; RC, ATT, HMD, AB and ESA helped design the research and performed experiments; ARP and DGO provided samples; CY, CCH, JB, CC and SD analyzed data and contributed to the manuscript, JT was responsible for project management and co-ordination and helped write the manuscript, AS provided samples, designed the research, analyzed the data, and wrote and edited the manuscript.

REFERENCES

1 Lindblad-Toh K, Tanenbaum DM, Daly MJ, Winchester E, Lui WO, Villapakkam A *et al.* Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol* 2000; **18**: 1001 - 1005.
2 Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008; **40**: 722 - 729.
3 Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009; **458**: 719 - 724.
4 Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* 2008; **322**: 1377 - 1380.
5 Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* 2010; **463**: 899 - 905.
6 Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011; **476**: 298 - 303.

7 Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2011; **475**: 101 - 105.
8 Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* 2011; **471**: 235 - 239.
9 Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* 2011; **471**: 189 - 195.
10 Yau C, Mouradov D, Jorissen RN, Colella S, Mirza G, Steers G *et al.* A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biol* 2010; **11**: R92.
11 Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA, Edwards BK (eds). *SEER Cancer Statistics Review*. National Cancer Institute: Bethesda, MD, 1975-2008. http://seer.cancer.gov/csr/1975_2008/, based on November 2010 SEER data submission, posted to the SEER web site, 2011.
12 CRUK. http://info.cancerresearchuk.org/cancerstats/types/leukaemia/.
13 Gaidano G, Ballerini P, Gong JZ, Inghirami G, Neri A, Newcomb EW *et al.* p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* 1991; **88**: 5413 - 5417.
14 Zenz T, Mohr J, Edelmann J, Sarno A, Hoth P, Heuberger M *et al.* Treatment resistance in chronic lymphocytic leukemia: the role of the p53 pathway. *Leuk Lymphoma* 2009; **50**: 510 - 513.
15 Stilgenbauer S, Sander S, Bullinger L, Benner A, Leupolt E, Winkler D *et al.* Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. *Haematologica* 2007; **92**: 1242 - 1245.
16 Zenz T, Krober A, Scherer K, Habe S, Buhler A, Benner A *et al.* Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. *Blood* 2008; **112**: 3322 - 3329.
17 Malcikova J, Smardova J, Rocnova L, Tichy B, Kuglik P, Vranova V *et al.* Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage. *Blood* 2009; **114**: 5307 - 5314.
18 Stankovic T, Stewart GS, Fegan C, Biggs P, Last J, Byrd PJ *et al.* Ataxia telangiectasia mutated-deficient B-cell chronic lymphocytic leukemia occurs in pregerminal center cells and results in defective damage response and unrepaired chromosome damage. *Blood* 2002; **99**: 300 - 309.
19 Hallek M. First-line treatment with fludarabine (F), cyclophosphamide (C), and rituximab (R) (FCR) improves overall survival (OS) in previously untreated patients (pts) with advanced chronic lymphocytic leukemia (CLL): results of a randomized phase III trial on behalf of an international group of investigators and the German CLL study group. *Blood (ASH Ann Meet Abstr)* 2009; **114**: 535.
20 Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 2000; **343**: 1910 - 1916.
21 Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE *et al.* A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 2005; **353**: 1793 - 1801.
22 Bandi N, Zbinden S, Gugger M, Arnold M, Kocher V, Hasan L *et al.* miR-15a and miR-16 are implicated in cell cycle regulation in a Rb-dependent manner and are frequently deleted or down-regulated in non-small cell lung cancer. *Cancer Res* 2009; **69**: 5553 - 5559.
23 Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci USA* 2005; **102**: 13944 - 13949.
24 Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR *et al.* Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* 2007; **27**: 2240 - 2252.
25 Raveche ES, Salerno E, Scaglione BJ, Manohar V, Abbasi F, Lin YC *et al.* Abnormal microRNA-16 locus with synteny to human 13q14 linked to CLL in NZB mice. *Blood* 2007; **109**: 5079 - 5086.
26 Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L *et al.* The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nat Med* 2008; **14**: 1271 - 1277.
27 Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T *et al.* The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell* 2010; **17**: 28 - 40.
28 Ouillette P, Erba H, Kujawski L, Kaminski M, Shedden K, Malek SN. Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. *Cancer Res* 2008; **68**: 1012 - 1021.
29 Parker H, Rose-Zerilli MJ, Parker A, Chaplin T, Wade R, Gardiner A *et al.* 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* 2011; **25**: 489 - 497.
30 Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K *et al.* The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. *Clin Cancer Res* 2011; **17**: 6778 - 6790.
31 Sellick GS, Wade R, Richards S, Oscier DG, Catovsky D, Houlston RS. Scan of 977 nonsynonymous SNPs in CLL4 trial patients for the identification of genetic variants influencing prognosis. *Blood* 2008; **111**: 1625 - 1633.
32 Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM *et al.* Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. *Blood* 2007; **109**: 1202 - 1210.
33 Gunn SR, Mohammed MS, Gorre ME, Cotter PD, Kim J, Bahler DW *et al.* Whole-genome scanning by array comparative genomic hybridization as a clinical tool for risk assessment in chronic lymphocytic leukemia. *J Mol Diagn* 2008; **10**: 442 - 451.
34 Kujawski L, Ouillette P, Erba H, Saddler C, Jakubowiak A, Kaminski M *et al.* Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. *Blood* 2008; **112**: 1993 - 2003.
35 Kay NE, Eckel-Passow JE, Braggio E, Vanwier S, Shanafelt TD, Van Dyke DL *et al.* Progressive but previously untreated CLL patients with greater array CGH complexity exhibit a less durable response to chemoimmunotherapy. *Cancer Genet Cytogenet* 2010; **203**: 161 - 168.
36 Gunnarsson R, Isaksson A, Mansouri M, Goransson H, Jansson M, Cahill N *et al.* Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. *Leukemia* 2010; **24**: 211 - 215.
37 Gunnarsson R, Mansouri L, Isaksson A, Goransson H, Cahill N, Jansson M *et al.* Array-based genomic screening at diagnosis and follow-up in chronic lymphocytic leukemia. *Haematologica* 2011; **96**: 1161 - 1169.
38 Ouillette P, Collins R, Shakhan S, Li J, Peres E, Kujawski L *et al.* Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood* 2011; **118**: 3051 - 3061.
39 Hong D, Gupta R, Ancliff P, Atzberger A, Brown J, Soneji S *et al.* Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* 2008; **319**: 336 - 339.
40 Bateman CM, Colman SM, Chaplin T, Young BD, Eden TO, Bhakta M *et al.* Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* 2010; **115**: 3553 - 3558.
41 Notta F, Mullighan CG, Wang JC, Poeppl A, Doulatov S, Phillips LA *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* 2011; **469**: 362 - 367.
42 Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, Colman SM *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011; **469**: 356 - 361.
43 Sun W, Wright FA, Tang Z, Nordgard SH, Van Loo P, Yu T *et al.* Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res* 2009; **37**: 5365 - 5377.
44 Greenman CD, Bignell G, Butler A, Edkins S, Hinton J, Beare D *et al.* PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. *Biostatistics* 2010; **11**: 164 - 175.
45 Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA* 2010; **107**: 16910 - 16915.
46 Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res* 2010; **20**: 68 - 80.
47 Letouze E, Allory Y, Bollet MA, Radvanyi F, Guyon F. Analysis of the copy number profiles of several tumor samples from the same patient reveals the successive steps in tumorigenesis. *Genome Biol* 2010; **11**: R76.
48 Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949 - 951.

Supplementary Information accompanies the paper on the Leukemia website (http://www.nature.com/leu)