UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO DI PEDIATRIA "SALUS PUERI"
Direttore: Ch.mo Prof. Giuseppe Basso

SCUOLA DI DOTTORATO DI RICERCA IN
MEDICINA DELLO SVILUPPO E SCIENZE DELLA PROGRAMMAZIONE
INDIRIZZO EMATO-ONCOLOGIA E IMMUNOLGIA-CICLO XXIII

Genome-based technologies provide novel insights into the pathogenesis and the clinical response of lymphoid malignancies

Direttore della Scuola: Ch.mo Prof. Giuseppe Basso
Supervisore: Ch.mo Prof. Giuseppe Basso

Dottorando: Giulia Fabbri

January 31, 2011

Index

Summary

After obtaining the M.D. Degree from the University of Padua with a thesis entitled "Evaluation of Minimal Residual Disease in ALL relapsed patients", I started in the same University the Ph.D. School in Developmental Medicine and Programming Sciences, Immunology and Onco-haematology, under the supervision of Professor Giuseppe Basso. Since then, the research projects in which I was directly involved were all focused in hematology-oncology related topics.

During the first two years of Ph.D. training I applied molecular biology technologies to monitor the clinical course of patients affected by acute lymphoblastic leukemia (ALL), which is the most frequent pediatric hematologic malignancy. In particular, I was involved in research projects aimed at evaluating the prognostic value of Minimal Residual Disease (MRD) analysis in pediatric ALL patients.

Despite the major advances in ALL treatment achieved in the last decades, recurrence of this disease remains the leading cause of treatment failure, occurring in 20% of patients. The early response to therapy is the most important prognostic factor in ALL. Around 95% of ALL patients achieve the Complete Remission (CR), which is judged by the absence of morphologically identifiable leukemic blasts in bone marrow smears. However, such finding can correspond to a still substantial burden of leukemic cells, and indeed many patients, despite achieving the CR, subsequently relapse. The persistence of leukemic cells not detectable by common cyto-morphological techniques is defined Minimal Residual Disease. Combined MRD status at the end of induction treatment and at the beginning of therapy consolidation is recognized as the most important prognostic factor, and it is now routinely implemented in stratifying ALL patients in different classes of risk. This stratification is expected to contribute to improve the overall outcome of children affected by ALL. Among the different alternative methods available for MRD monitoring, the detection of immune gene (Ig/TCR) rearrangements by real-time quantitative PCR (RQ-PCR) is the most widely used as it is feasible in the vast majority of the patients.

In a study performed in Professor Basso's Lab, described in Chapter 1.1, we analyzed the role of MRD after reinduction therapy on the outcome of 60 'high-risk' relapsed ALL patients (those with an early BM recurrence or with T-ALL marrow relapse), enrolled in the AIEOP LAL REC 2003 study protocol. We performed MRD monitoring based on the

detection of Ig/TCR rearrangements by RQ-PCR at different time-points during therapy: i) time point 1 (TP1), after induction; ii) time point 2 (TP2), after re-induction, and iii) time point 3 (TP3), after consolidation. The overall 3-year event-free survival for patients with MRD levels at the TP1 $\geq 10^{-4}$ was significantly lower (19%), as compared to patients with MRD levels negative or below 10^{-4} (73 and 45%, respectively). The statistical significance of MRD predictive prognostic value was confirmed by multivariate analysis. This study demonstrated that early MRD evaluation in 'high-risk' ALL relapses allows the identification of a subset of patients who seem not to benefit from conventional treatment, including chemotherapy and hematopoietic stem cell transplantation, and who could therefore need alternative, novel, experimental therapies.

Medullary ALL relapse is currently still defined employing morphologic criteria, i.e. presence of a blast percentage exceeding 25% in a bone marrow sample after complete remission. To date there are only a few studies suggesting that it is possible to predict impending morphologic relapse through MRD monitoring in hematologic malignancies. In another study performed in Professor Basso's Lab (described in Chapter 1.2) we evaluated the clinical significance of MRD monitoring in ALL pediatric patients as an indicator of impending hematological disease recurrence. Towards this aim we monitored MRD during and after treatment in 113 patients enrolled in the AIEOP-BFM ALL 2000 clinical trial and we performed a case-control study of 37 relapsed ALL children matched with 37 ALL controls in continuous complete remission. The cumulative incidence of relapse in patients with positive quantifiable results was significantly higher (85.7%) than that of children with MRD results either negative or positive below the quantitative range. Relapse of ALL could have been predicted by MRD monitoring in 82% of patients who were evaluated within 3 months before relapse occurrence. This study suggested that detection of positive and quantifiable MRD during follow-up is highly predictive of subsequent hematologic relapse of pediatric ALL. Accuracy of such a prediction would improve by performing monitoring every 3 months. In addition, it could be helpful to design and administer personalized salvage therapy before overt hematologic relapse, namely when the tumor burden is lower. Additional clinical studies are needed to prove whether therapeutic intervention based on MRD results can improve the final outcome of pediatric ALL patients

After finishing the first two years of my Ph.D. program, I wanted to get more involved in basic research, aimed at understanding the pathogenetic mechanisms of hematologic malignancies. I therefore continued my training in Professor Riccardo Dalla-Favera's

Laboratory (Columbia University, New York), under the supervision of Professor Laura Pasqualucci, whose main focus is the study of the pathogenesis of B-cell derived lymphomas. Cancer is associated with the accumulation of various genetic alterations, whose spectrum can provide important insights in understanding its pathogenesis and rational bases for innovative targeted therapeutic strategies. We applied novel genome-wide technologies including next generation whole-exome sequencing and genome-wide high-density single nucleotide polymorphism (SNP) array analysis to an extended panel of B cell malignancies, including diffuse large B cell lymphoma (DLBCL) and chronic lymphocytic leukemia (CLL), in order to determine their full spectrum of genetic lesions. The ultimate goal of these projects is to identify novel target genes involved in the pathogenesis of these diseases, assess their physiological role in normal B cell development and their pathological role in lymphomagenesis.

The whole exome sequencing analysis of 7 DLBCL cases integrated with SNP array analysis of 72 DLBCL cases led to the identification of >450 loci affected by somatic point mutations and/or by recurrent, focal gene copy number (CN) aberrations. Among those that have been so far independently validated, we found frequent genomic deletions and/or somatic mutations inactivating the genes that encode CREBBP and, in a smaller fraction of cases, EP300, two highly related histone and non-histone acetyltranferases (HAT) that act as transcriptional coactivators in multiple signaling pathways (overall frequency ~39%). The analysis of these genes in various types of mature B-Non-Hodgkin lymphomas, including Follicular lymphoma (FL), Burkitt Lymphoma (BL), Marginal-Zone Lymphoma (MZL) and Chronic Lymphocytic Leukemia (CLL), revealed the presence of genomic lesions affecting these genes also in ~41% of FL cases. The vast majority of the identified lesions were removing or inactivating the HAT coding domain of these two genes. Interestingly, *CREBBP* and *EP300* lesions were commonly affecting one allele, suggesting that reduction in HAT dosage is important for lymphomagenesis. By studying the functional consequences of these lesions, we demonstrated specific defects in acetylation-mediated inactivation of the BCL6 onco-protein and activation of the p53 tumor-suppressor. These results identified *CREBBP/EP300* mutations as a major pathogenetic mechanism shared by common forms of B-NHL, and have direct implications for the use of drugs targeting acetylation/deacetylation mechanisms. The results of this study are described in Chapter 2.1.

The whole exome sequencing analysis of 5 CLL cases revealed a total of 38 somatic non-silent mutations, corresponding to 38 distinct genes at an average of 7.6 mutations/case (range, 5-10/case). The somatically mutated genes were belonging to a variety of functional

classes, including chromatin remodeling genes, members of pathways such as NF-κB, TGF-β and Wnt, and genes involved in cytoskeleton organization and cell motion. The majority of the identified events were represented by single base-pair substitutions (N=36, of which 32 missense mutations and 4 nonsense mutations), while frameshift insertions/deletions were rare (N=2 deletions of 4 and 32 base pairs, respectively). Analysis of the mutation features showed a prevalence of transitions versus transversions (69% vs 31%) and an elevated G+C over A+T ratio (69% vs 31%), analogous to the mutation spectrum observed in the genome of epithelial tumors such as colorectal, pancreatic and brain cancer.

The SNP array analysis of the same five cases used for the whole exome sequencing analysis identified a relatively low number of gene CN aberrations (N=14), corresponding to an average of 2.8 lesions/case (range, 0-5/case), mostly represented by deletions (N=11/14, 80%).

The genes newly identified through the whole exome sequencing approach were tested for recurrence by PCR amplification and direct sequencing of their entire coding region and by SNP array analysis in an independent panel of 48 CLL cases. This analysis revealed that the majority of these genes were not recurrently altered in CLL, while 5 genes showed either recurrent mutations or CN aberrations at a relatively low frequency (≤ 6% of CLL cases).

Overall, our integrated approach revealed that the coding genome of CLL appears to be relatively stable, harboring on average 10 potentially relevant genetic alterations per case. The low frequency of alterations in individual genes may reflect a high recurrence of lesions affecting commom pathways important for the pathogenesis of the disease. The results of this study are described in Chapter 2.2.

Compendio

Dopo aver conseguito la Laurea in Medicina e Chirurgia presso l'Università degli Studi di Padova con una tesi dal titolo "Valutatione della Malattia Residua Minima in pazienti LAL ricaduti", ho iniziato, presso lo stesso Ateneo, la Scuola di Dottorato in Medicina dello Sviluppo e Scienze della Programmazione con indirizzo in Emato-oncologia sotto la supervisione del Professor Giuseppe Basso. Nel corso della Scuola di Dottorato ho partecipato a progetti di ricerca focalizzati sullo studio di malattie emato-oncologiche.

Nel corso dei primi due anni di Dottorato mi sono avvalsa di tecniche di biologia molecolare per monitorare il decorso clinico di pazienti affetti da leucemia acuta linfoblastica (LAL), la più frequente neoplasia ematologica dell'età pediatrica. In particolare, mi sono occupata di studi mirati alla valutazione del valore prognostico del monitoraggio della Malattia Residua Minima (MRM) in pazienti pediatrici LAL.

Nonostante i notevoli progressi compiuti nelle ultime decadi nel trattamento della LAL pediatrica, la recidiva di malattia, che colpisce il 20% dei pazienti circa, resta la causa più frequente di fallimento terapeutico. La risposta precoce alla terapia è il fattore prognostico più importante nella LAL. Il 95% dei pazienti raggiunge la remissione completa (RC), che è definita secondo criteri morfologici come assenza di blasti leucemici nell'aspirato midollare. Tale reperto può tuttavia corrispondere ad una massa tumorale ancora considerevole. Pertanto molti pazienti, nonostante raggiungano la remissione completa, vanno incontro a recidiva. La persistenza di cellule leucemiche non identificabili con le comuni tecniche citomorfologiche è definita Malattia Residua Minima (MRM). Lo status della MRM al termine della terapia di induzione e all'inizio del consolidamento è considerato il fattore prognostico più importante nella LAL pediatrica, tanto da essere diventato un parametro usato routinariamente per stratificare i pazienti LAL in diverse classi di rischio che vengono trattate con distinti regimi terapeutici.
La Real-Time quantitative PCR (RQ-PCR) per i riarrangiamenti genici dei recettori degli antigeni (Ig/TCR) è il metodo attualmente più usato per il monitoraggio della MRM essendo applicabile nella grande maggioranza dei pazienti.

In uno studio condotto nel Laboratorio del Professor Basso, e descritto nel Capitolo 1.1, abbiamo analizzato il ruolo della MRM post-terapia di reinduzione sulla prognosi di 60

pazienti LAL ricaduti ad alto rischio, affetti da recidiva midollare precoce o ad immunofenotipo T, in accordo con i criteri del protocollo terapeutico AIEOP REC 2003. Abbiamo monitorato la MRM mediante uso di RQ-PCR per i riarrangiamenti genici Ig/TCR in diverse fasi della terapia: i) dopo l'induzione (TP1); ii) dopo la reinduzione (TP2) e iii) dopo il consolidamento post-ricaduta (TP3).

L'Event Free Survival (EFS) a tre anni dalla ricaduta è risultata significativamente inferiore in pazienti con MRM con valori$\geq 10^{-4}$ al TP1 (19%), rispetto a pazienti con MRM negativa o con valori $<10^{-4}$ (73 e 45%). Il valore prognostico predittivo della MRM è risultato essere statisticamente significativo anche mediante analisi multivariata.

I risultati di questo studio dimostrano che una valutazione precoce del valore della MRM in pazienti LAL recidivati ad alto rischio permette di identicarne il subset che sembra non beneficiare dei regimi terapeutici convenzionali, quali la chemioterapia ed il trapianto allogeneico, e che potrebbe necessitare di innovative terapie sperimentali.

La ricaduta midollare di LAL è attualmente definita secondo criteri morfologici come conta di blasti $\geq$25% in un prelievo midollare successivo a quello della remissione completa. Ad oggi è stato pubblicato un numero relativamente basso di studi che suggeriscono la possibilità di prevedere la ricaduta morfologica di LAL mediante monitoraggio della MRM in neoplasie ematologiche.

In uno studio svoltosi presso il Laboratorio del Professor Basso, descritto nel Capitolo 1.2, abbiamo valutato il significato del monitoraggio MRM come indicatore di successiva ricorrenza ematologica di LAL. A questo scopo abbiamo monitorato i livelli di MRM nel corso e dopo il termine della terapia in 113 pazienti arruolati nel protocollo terapeutico AIEOP-BFM LAL 2000 ed abbiamo condotto uno studio caso-controllo su 37 pazienti LAL ricaduti e 37 pazienti LAL in RC continua. L'incidenza cumulativa di ricaduta in pazienti con riscontro di uno o più prelievi con MRM positiva quantificabile nel corso del follow-up è risultata significativamente più elevata (86%) rispetto ai casi con MRM negativa o al di sotto del *quantitative range* della tecnica. Inoltre, in pazienti monitorati per MRM entro 3 mesi dalla ricaduta, la ricorrenza di malattia sarebbe risultata prevedibile nell'82% dei casi. Questo studio ha dimostrato che il riscontro di un prelievo positivo per MRM nel corso del follow-up è altamente predittivo di successiva ricaduta ematologica in pazienti pediatrici LAL. L'accuratezza di tale predizione sembra essere più elevata in caso di monitoraggio ogni 3 mesi. Inoltre, questi dati suggeriscono che una somministrazione precoce della terapia di recupero prima del manifestarsi della ricaduta ematologica, in presenza di una massa tumorale più limitata, potrebbe migliorare la risposta al trattamento ed il decorso clinico dei

pazienti. Ulteriori studi clinici saranno necessari per dimostrare in maniera conclusiva se un intervento terapeutico basato sulla MRM può migliorare la prognosi dei pazienti pediatrici LAL.

Al termine dei primi due anni di Dottorato ho deciso di dedicarmi in prima persona non solo a progetti di ricerca clinici, ma anche a progetti di ricerca di base, con lo scopo di capire e studiare i meccanismi fisiopatologici delle neoplasie ematologiche. Ho quindi continuato la mia formazione presso il Laboratorio del Professor Riccardo Dalla Favera, sotto la supervisione della Professoressa Laura Pasqualucci, il cui principale campo di interesse è lo studio della patogenesi dei linfomi a cellule B. Il cancro è una patologia associata con il progressivo accumulo di alterazioni genetiche, il cui spettro è molto importante non solo per comprenderne la patogenesi, ma anche per lo sviluppo di nuove strategie terapeutiche mirate. Abbiamo usato tecnologie *genome-wide* quali il *next-generation whole exome sequencing* e la analisi di *SNP array* in un pannello di neoplasie ematologiche costitutito da linfomi diffusi a grandi cellule (DLBCL) e da leucemie linfatiche croniche (CLL). Lo scopo di questi progetti è l'identificazione di nuovi geni coinvolti nella patogenesi di queste malattie, lo studio del loro ruolo fisiologico nello sviluppo dei linfociti B e patologico nella linfomagenesi.

Il sequenziamento dell'esoma di 7 casi di DLBCL integrato con l'analisi di *SNP Array* di 72 casi di DLBCL ha portato all'identificazione di più di 450 loci colpiti da mutazioni somatiche e/o da aberrazioni del *copy number* (CNAs) focali e ricorrenti. Tra i geni validati fino ad ora, abbiamo trovato frequenti delezioni e/o mutazioni somatiche inattivanti i geni che codificano per CREBBP e, in una percentuale più bassa di casi, EP300 (frequenza totale, 39% dei DLBCL). CREBBP e EP300 sono due acetil-transferasi istoniche e non altamente correlate, che agiscono come coattivatori trascrizionali in numerose cascate di trasduzione del segnale. L'analisi di questi geni in altri tipi di linfomi a cellule B ha portato all'identificazione di lesioni genetiche a carico di *CREBBP* e *EP300* anche nel 41% dei casi di linfoma follicolare (FL). La grande maggioranza di queste lesioni provoca la rimozione o l'inattivazione dell'HAT domain di questi due enzimi. Inoltre, le alterazioni di *CREBBP* e EP300 colpiscono frequentemente un singolo allele, suggerendo che la riduzione dell'*HAT dosage* ha un ruolo importante nella patogenesi dei DLBCL e dei FL. Lo studio funzionale delle consequenze di queste lesioni ha dimostrato l'evidenza di difetti nell'inattivazione dell'onco-proteina BCL6 e nell'attivazione dell'onco-soppressore p53 mediate da acetilazione.

Questi dati hanno dunque dimostrato il ruolo patogenetico di lesioni geniche a carico di *EP300/CREBBP* in due comuni tipi di linfoma a cellule B. Inoltre i risultati di questo studio hanno un'implicazione clinica diretta, suggerendo il possibile utilizzo di farmaci operanti su meccanismi di acetilazione e deacetilazione nella terapia dei linfomi. I risultati di questo lavoro sono descritti nel Capitolo 2.1.

Il sequenziamento dell'esoma di 5 casi di CLL ha rivelato 38 mutazioni somatiche a carico di 38 distinti geni, con una media di 7.6 mutazioni per caso (range, 5-10/caso). I geni mutati appartengono a diverse classi funzionali, essendo coinvolti in processi di rimodellamento cromatinico, in *pathways* quali NF-κB, TGF-β e Wnt, e in meccanismi di organizzazione del citoscheletro e di motilità cellulare.

La maggioranza delle mutazioni somatiche identificate è risultata essere costituita da sostituzioni di una singola base (N=36, di cui 32 mutazioni missenso e 4 nonsenso), mentre una minoranza da inserzioni/delezioni causanti alterazioni del frame (N=2 delezioni di 4 e 32 basi, rispettivamente). L'analisi dello spettro delle mutazioni puntiformi ha mostrato una prevalenza di transizioni (69% degli eventi) e un'elevata ratio G+C/A+T (69%/31%). Tale pattern è risultato essere analogo allo spettro di mutazioni riportato nell'esoma di neoplasie epiteliali, tra cui tumori a carico del colon e del pancreas, e tumori del sistema nervoso centrale, tra cui neuroblastomi e medulloblastomi.

L'analisi di *SNP array* degli stessi casi usati per il sequenziamento dell'esoma ha rivelato un numero relativamente basso di CNAs (N=14), corrispondente ad una media di 2.8 lesioni per caso (range, 0-5 CNAs/caso), rappresentate soprattuto da delezioni (N=11/14, 80%).

L'analisi dei nuovi geni identificati tramite il sequenziamento dell'esoma in un pannello indipendente di 48 casi di CLL ha rivelato che la maggioranza di questi geni non è alterata in modo ricorrente nella CLL. In 5 geni sono state riscontrate mutazioni e/o CNAs con una frequenza relativamente bassa (≤ 6% dei casi).

In conclusione, questi dati dimostrano che il genoma codificante della CLL contiene in media 10 alterazioni geniche per caso. Lo screening di questi geni per mutazioni e CNAs in un pannello esteso di casi ha rivelato che i geni identificati tramite questo approccio sono alterati a bassa frequenza nella CLL. Tale bassa frequenza di alterazioni a carico del singolo gene può suggerire la presenza di alterazioni ricorrenti a carico di altri geni coinvolti in pathways comuni che potrebbero avere un ruolo importante nella patogenesi della CLL. I risultati di questo studio sono descritti nel Capitolo 2.2.

Chapter 1. The role of MRD in pediatric Acute Lymphoblastic Leukemia

Acute Lymphoblastic Leukemia (ALL), a tumor of lymphoid progenitors that may be of B- or T-lymphoid lineage, is the most common pediatric cancer and it also represents an important cause of morbidity and mortality from hematopoietic malignancies in the adult[1].

The outcome of pediatric ALL has remarkably improved in recent decades, with cure rates nowadays exceeding 80%[2], but up to one-quarter of patients eventually suffer relapse. Despite substantial second remission rates and wide availability of haematopoietic stem cell transplant, most relapsed children die[3]. Therefore novel therapeutic strategies are needed, not only in salvage regimens, but also in frontline protocols, especially for those patients who are at high risk of relapse.

ALL is frequently associated with recurrent genetic alterations, including whole chromosomal gains and losses and translocations resulting in the expression of chimeric fusion genes (i.e. *ETV6-RUNX1*, *TCF3-PBX1*, *BCR-ABL1* and rearrangements of the gene *MLL*) or in dysregulation of genes by juxtaposition to antigen receptor gene loci[4]. Some of these lesions, including hypodiploidy and *MLL* rearrangements, confer a high risk of treatment failure and relapse. However, the frequencies of these gross chromosomal rearrangements are insufficient to fully explain treatment failure in ALL.

In recent years genome-wide approaches have been widely used to identify the full spectrum of structural genetic lesions, especially at the submicroscopic level, present in ALL (reviewed in Mullighan CG[5]). The majority of ALL cases harbor a relatively low number of alterations, mostly focal and affecting genes involved in lymphoid development and leukemogenesis[6]. The frequency of such aberrations is varying among different ALL subtypes: *MLL*-rearranged cases have less than one Copy Number (CN) aberration per case, suggesting that this potent oncogene requires very few cooperating mutations to induce leukemic transformation, while *ETV6-RUNX1* and *BCR-ABL1* rearranged-cases have more than 6 lesions per case. More than two third of ALL cases carry inactivating deletions and/or mutations affecting transcription factors involved in early stages of B-cell development, most commonly *PAX5* and *IKZF1*[6-7]. In the majority of the cases these lesions affect only one allele, suggesting haploinsufficiency. While *PAX5* lesions have not been shown to be associated with ALL outcome, *IKZF1* aberrations confer a very poor prognosis in ALL patients. *IKZF1* deletions or sequence mutations are not only important determinants of the lineage and progression of *BCR-ABL1* (Ph+) positive ALL, where they are found in up to

80% of cases[8-9], but are also associated with a poor survival in Ph- childhood ALL. Therefore *IKZF1* inactivating events represent an attractive candidate to be used to stratify ALL patients at diagnosis[7,10]. Recent studies identified additional targets of recurrent genetic lesions in 'high-risk' ALL patients, such as rearrangements and, less frequently, somatic mutations affecting *CRLF2*[11]. Most of the cases carrying lesions affecting this gene, also harbor somatic activating mutations affecting *JAK1* and *JAK2*[12-13]. The alterations involving *CRLF2* are present in up to 50% of ALL cases associated with Down Syndrome and in up to 7% of B-ALL cases, where they associate with *JAK* mutations and *IKZF1* alterations and confer a very poor outcome[12,14].

T-ALL is less frequent than B-ALL, and it is associated with a worse prognosis[15]. Recurrent genetic abnormalities occurring in this disease include dysregulation of *LMO2*[16], amplifications of *MYB*[17-18], amplifications associated with the *NUP214-ABL1*[19] chromosomal rearrangement, and deletions and inactivating mutations of *PTEN* and *WT1*[20-21]. In addition, oncogenic mutations of *NOTCH1* and inactivating mutations of *FBXW7*, which are observed in up to 50% of the cases, have been largely associated with a favorable prognosis[22-23].

Moreover, two recent studies, based on second-generation deep sequencing of X chromosome, revealed recurrent inactivating mutations of the gene *PHF6* in pediatric T-ALL[24].

Our current knowledge of the biologic determinants of treatment failure and relapse in ALL is still limited. Major insights regarding the genetic basis of relapse were provided by a study based on genome-wide DNA copy number analyses on matched diagnosis and relapse samples from ALL pediatric patients[25]. The diagnosis and relapse samples typically showed different patterns of genomic CNAs, with the CNAs acquired at relapse preferentially affecting genes implicated in cell cycle regulation and B cell development. Most relapse samples lacked some of the CNAs present at diagnosis, which suggests that the cells responsible for relapse are ancestral to the primary leukemia cells. Backtracking analyses revealed that cells corresponding to the relapse clone were often present as minor subpopulations at diagnosis. These data suggest that genomic abnormalities contributing to ALL relapse are selected for during treatment, and they point to new targets for therapeutic intervention.

So far the most important prognostic factor in ALL is represented by the early response to induction treatment[26]. The complete remission after the initial phases of therapy is judged by the absence of morphologically identifiable leukemic blasts in bone marrow smears. However, complete remission does not mean that all leukemic cells are completely

eradicated, but only that their level is below the sensitivity of the cytomorphologic methods (1-5%). At this time, up to 10^{10} tumoral cells can still remain in samples apparently devoid of ALL cells. The persistence of cells not identifiable through cytomorphologic techniques is termed Minimal Residual Disease (MRD).

In the last two decades a lot of effort has been applied to develop novel sensitive methodologies to measure more accurately the kinetics of treatment response by monitoring MRD in ALL patients.

Currently MRD can be measured in ALL patients by flow-cytometry, PCR amplification of fusion transcripts and PCR amplification of Immunoglobulin (Ig) and T-cell receptor (TCR) gene rearrangements.

The targets of flow cytometric studies are represented by leukemia-associated phenotypes, which are not expressed by normal hematopoietic cells. The most widely used targets are represented by markers normally expressed during lympho-hematopoiesis but found in abnormal combinations in leukemia cells[27]. This approach allows MRD monitoring in up to 95% of B-ALL patients[28].

ALL cells carry genetic abnormalities (see above) resulting in the overexpression of aberrant mRNA transcripts which can be used for MRD detection. The fusion-transcripts most widely used are *BCR-ABL1, MLL-AFF1, TCF3-PBX1* and *ETV6-RUNX*[29]. Such recurrent abnormalities suitable for MRD evaluation are present in approximately 40% of pediatric ALL patients. Among the abnormalities recently identified in ALL, fusions involving *CRLF2* could be used for MRD monitoring[11,30].

The second category of PCR targets for MRD studies in ALL is represented by the junctional regions deriving by Ig/TCR gene rearrangements which are unique to the leukemic clone[31]. The most common approach includes a PCR-based screening of diagnostic samples with primers matching the V and J regions of the various antigen-receptor genes to determine if rearrangements are present. The identified rearrangements are tested for clonality by heteroduplex analysis[32], and the sequences obtained for each specific junctional region are used to design allele-specific oligonucleotides. Quantitation of MRD levels is performed by using real-time quantitative PCR, testing in parallel mixtures of diagnostic leukemic and normal DNAs to assess the sensitivity of the assay. This approach can be performed in the vast majority of ALL patients[33-34].

The prognostic value of MRD measurement to evaluate early response to therapy was first convincingly demonstrated in 3 studies published in the late nineties by the EORTC, St Jude and BFM groups[35-37]. Since then MRD assays have been incorporated into several

clinical trials, becoming routinely used to classify patients into different risk classes and to guide their subsequent therapeutic regimen[33,38]. MRD levels after first phases of chemotherapy nullify the prognostic value of presenting clinical and biological parameters, including leukocyte count, age, early prednisone response and genetic subtype[38]. Within different clinical protocols a poor clearance of the disease, as assessed by high MRD levels at specific time points, was associated with a significantly higher level of relapses (i.e. days +15, +33 and +78 in the AIEOP protocol[33,39]).

MRD is a strong prognostic factor also for patients who have suffered a first relapse and have achieved a second remission[40-42] or patients with isolated extramedullary relapse[43]. Moreover, detection of MRD before allogeneic hematopoietic stem cell transplantation is associated with an increase risk of relapse post-transplantation[44-48].

In the Ph.D. research activity herein described we further investigated the role of MRD status in pediatric ALL, not only as a prognostic factor in high-risk relapsed patients, but also as a useful tool to recognize leukemia relapse before it is morphologically overt.

References

1. Pui, C.H., Robison, L.L. & Look, A.T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030-1043 (2008).
2. Pui, C.H., *et al.* Long-term results of St Jude Total Therapy Studies 11, 12, 13A, 13B, and 14 for childhood acute lymphoblastic leukemia. *Leukemia* **24**, 371-382 (2010).
3. Nguyen, K., *et al.* Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. *Leukemia* **22**, 2142-2150 (2008).
4. Harrison, C.J. Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia. *Br J Haematol* **144**, 147-156 (2009).
5. Mullighan, C.G. & Downing, J.R. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. *Leukemia* **23**, 1209-1218 (2009).
6. Mullighan, C.G., *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
7. Mullighan, C.G., *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* **360**, 470-480 (2009).
8. Mullighan, C.G., *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453**, 110-114 (2008).
9. Martinelli, G., *et al.* IKZF1 (Ikaros) deletions in BCR-ABL1-positive acute lymphoblastic leukemia are associated with short disease-free survival and high rate of cumulative incidence of relapse: a GIMEMA AL WP report. *J Clin Oncol* **27**, 5202-5207 (2009).
10. Kuiper, R.P., *et al.* IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leukemia* **24**, 1258-1264 (2010).

11. Mullighan, C.G., *et al.* Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. *Nat Genet* **41**, 1243-1246 (2009).
12. Harvey, R.C., *et al.* Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood* **115**, 5312-5321 (2010).
13. Mullighan, C.G., *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **106**, 9414-9418 (2009).
14. Cario, G., *et al.* Presence of the P2RY8-CRLF2 rearrangement is associated with a poor prognosis in non-high-risk precursor B-cell acute lymphoblastic leukemia in children treated according to the ALL-BFM 2000 protocol. *Blood* **115**, 5393-5397 (2010).
15. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat Rev Immunol* **8**, 380-390 (2008).
16. Van Vlierberghe, P., *et al.* The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood* **108**, 3520-3529 (2006).
17. Lahortiga, I., *et al.* Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nat Genet* **39**, 593-595 (2007).
18. Clappier, E., *et al.* The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood* **110**, 1251-1261 (2007).
19. Graux, C., *et al.* Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet* **36**, 1084-1089 (2004).
20. Palomero, T., *et al.* Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nat Med* **13**, 1203-1210 (2007).
21. Tosello, V., *et al.* WT1 mutations in T-ALL. *Blood* **114**, 1038-1045 (2009).
22. Kox, C., *et al.* The favorable effect of activating NOTCH1 receptor mutations on long-term outcome in T-ALL patients treated on the ALL-BFM 2000 protocol can be separated from FBXW7 loss of function. *Leukemia* **24**, 2005-2013 (2010).
23. Breit, S., *et al.* Activating NOTCH1 mutations predict favorable early treatment response and long-term outcome in childhood precursor T-cell lymphoblastic leukemia. *Blood* **108**, 1151-1157 (2006).
24. Van Vlierberghe, P., *et al.* PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet* **42**, 338-342 (2010).
25. Mullighan, C.G., *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377-1380 (2008).
26. Pui, C.-H. & Evans, W.E. Treatment of acute lymphoblastic leukemia. *N Engl J Med* **354**, 166-178 (2006).
27. Campana, D. Role of minimal residual disease monitoring in adult and pediatric acute lymphoblastic leukemia. *Hematol Oncol Clin North Am* **23**, 1083-1098, vii (2009).
28. Borowitz, M.J., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* **111**, 5477-5485 (2008).
29. Bruggemann, M., *et al.* Standardized MRD quantification in European ALL trials: proceedings of the Second International Symposium on MRD assessment in Kiel, Germany, 18-20 September 2008. *Leukemia* **24**, 521-535 (2010).
30. Yoda, A., *et al.* Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **107**, 252-257 (2010).

31. van der Velden, V.H.J., *et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* **21**, 604-611 (2007).
32. Germano, G., Songia, S., Biondi, A. & Basso, G. Rapid detection of clonality in patients with acute lymphoblastic leukemia. *Haematologica* **86**, 382-385 (2001).
33. Flohr, T., *et al.* Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* **22**, 771-782 (2008).
34. Stow, P., *et al.* Clinical significance of low levels of minimal residual disease at the end of remission induction therapy in childhood acute lymphoblastic leukemia. *Blood* **115**, 4657-4663 (2010).
35. Cave, H., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--Childhood Leukemia Cooperative Group. *N Engl J Med* **339**, 591-598 (1998).
36. Coustan-Smith, E., *et al.* Immunological detection of minimal residual disease in children with acute lymphoblastic leukaemia. *Lancet* **351**, 550-554 (1998).
37. van Dongen, J.J., *et al.* Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* **352**, 1731-1738 (1998).
38. Conter, V., *et al.* Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. *Blood* **115**, 3206-3214 (2010).
39. Basso, G., *et al.* Risk of relapse of childhood acute lymphoblastic leukemia is predicted by flow cytometric measurement of residual disease on day 15 bone marrow. *J Clin Oncol* **27**, 5168-5174 (2009).
40. Eckert, C., *et al.* Prognostic value of minimal residual disease in relapsed childhood acute lymphoblastic leukaemia. *Lancet* **358**, 1239-1241 (2001).
41. Coustan-Smith, E., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. *Leukemia* **18**, 499-504 (2004).
42. Paganin, M., *et al.* Minimal residual disease is an important predictive factor of outcome in children with relapsed 'high-risk' acute lymphoblastic leukemia. *Leukemia* **22**, 2193-2200 (2008).
43. Hagedorn, N., *et al.* Submicroscopic bone marrow involvement in isolated extramedullary relapses in childhood acute lymphoblastic leukemia: a more precise definition of "isolated" and its possible clinical implications, a collaborative study of the Resistant Disease Committee of the International BFM study group. *Blood* **110**, 4022-4029 (2007).
44. Knechtli, C.J., *et al.* Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. *Blood* **92**, 4072-4079 (1998).
45. van der Velden, V.H., *et al.* Real-time quantitative PCR for detection of minimal residual disease before allogeneic stem cell transplantation predicts outcome in children with acute lymphoblastic leukemia. *Leukemia* **15**, 1485-1487 (2001).
46. Bader, P., *et al.* Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. *Leukemia* **16**, 1668-1672 (2002).

47. Uzunel, M., Jaksch, M., Mattsson, J. & Ringden, O. Minimal residual disease detection after allogeneic stem cell transplantation is correlated to relapse in patients with acute lymphoblastic leukaemia. *Br J Haematol* **122**, 788-794 (2003).
48. Bader, P., *et al.* Prognostic value of minimal residual disease quantification before allogeneic stem-cell transplantation in relapsed childhood acute lymphoblastic leukemia: the ALL-REZ BFM Study Group. *J Clin Oncol* **27**, 377-384 (2009).

1.1 Minimal residual disease is an important predictive factor of outcome in children with relapsed 'high-risk' acute lymphoblastic leukemia (Leukemia 2008;22:2193-2200)

Maddalena Paganin[1,4], Marco Zecca[2,4], Giulia Fabbri[1], Katia Polato[1], Andrea Biondi[3], Carmelo Rizzari[3], Franco Locatelli[2] and Giuseppe Basso[1]

[1]*Laboratorio di Ematologia e Oncologia Pediatrica, Università di Padova, Padova, Italy.*

[2]*Oncoematologia Pediatrica, Fondazione IRCCS Policlinico San Matteo, Università di Pavia, Pavia, Italy.*

[3]*Centro Ricerca Tettamanti and Clinica Pediatrica, Università di Milano Bicocca, Monza, Italy.*

[4] *These authors contributed equally to this work.*

The aim of the study was to analyze the impact of minimal residual disease (MRD) after reinduction therapy on the outcome of children with relapsed 'high-risk' acute lymphoblastic leukemia (ALL). Sixty patients with isolated or combined marrow relapse were studied. All patients belonged to the S3 or S4 groups, as defined by the Berlin–Frankfurt–Münster stratification for relapsed ALL. MRD was studied by real-time quantitative PCR after the first, second and third chemotherapy course (time points 1 (TP1), 2 (TP2) and 3 (TP3), respectively). MRD results, not used for treatment refinement, were categorized as negative (NEG MRD), positive not-quantifiable (POS-NQ MRD) when MRD level was below quantitative range (a level $<10^{-4}$) or positive within quantitative range (POS MRD) when MRD level was $\geq 10^{-4}$. With a median observation time of 15 months, overall 3-year event-free survival (EFS) was 27%. The 3-year EFS was 73, 45 and 19% for patients with NEG-MRD, POS NQ-MRD and POS-MRD at TP1, respectively ($P<0.05$). The prognostic predictive value of MRD was statistically confirmed in multivariate analysis. MRD quantitation early and efficiently differentiates patients who benefit from conventional treatment, including allogeneic hematopoietic stem cell transplantation, from those needing innovative, experimental therapies.

Introduction

The treatment of acute lymphoblastic leukemia (ALL) in children has improved over the recent decades, and nowadays nearly 80% of patients can be cured with chemotherapy[1]. Nevertheless, disease recurrence remains the leading cause of treatment failure.

While a substantial proportion of children with relapsed ALL achieve a second remission, the overall final outcome remains unsatisfactory[2-7].

A number of studies analyzed factors that influence the outcome of relapsed ALL. The Berlin–Frankfurt–Münster (BFM) ALL Relapse Study Group identified, as poor prognostic factors after first relapse, a short duration of first remission, an isolated bone marrow (BM) relapse and T-cell immunophenotype[2]. Further studies have confirmed these observations[3-4,8-9]. In view of these results, the most recent protocols for relapsed ALL are based on patient stratification according to these risk factors. However, relapsed patients belonging to the higher risk groups (those with an early BM recurrence or with T-ALL marrow relapse) have a low probability of event-free survival (EFS), in particular when treated with chemotherapy alone[6].

Minimal residual disease (MRD) has already proven to be an important prognostic factor for a better stratification of patients in different risk classes in first-line treatment protocols[10-11]. The feasibility of measuring MRD for patients' stratification has been recently reported also in a large cohort of children with ALL[12]. Retrospective studies performed on stored samples have suggested that MRD could be of relevant importance also in intermediate-risk relapsed patients in second remission[13-14].

The aim of this study was to prospectively evaluate whether the early quantization of molecular response to treatment in children with relapsed ALL could also be used to predict the outcome of patients with poor prognostic characteristics.

Materials and methods

Patients

The study included 60 consecutive pediatric patients with a diagnosis of ALL, who experienced a BM or combined (BM plus an extramedullary site) relapse between June 2003 and December 2006, after first-line chemotherapy treatment according to the AIEOP ALL 2000 protocol. All patients belonged to either S3 or S4 risk group, as defined by the BFM risk group stratification for relapsed ALL[2]. Patients' characteristics are summarized in Table 1.

Diagnosis of ALL was made according to standard criteria. An isolated BM relapse was diagnosed with greater than or equal to 25% lymphoblasts among nucleated BM cells, without evidence of extramedullary relapse. In patients with proven extramedullary relapse, a combined relapse was diagnosed when the percentage of lymphoblasts in BM was greater than 5%.

Relapses occurring within 18 months from the start of the treatment were defined as very early, whereas relapses occurring after 18 months from the start of treatment but within 6 months after discontinuation of frontline treatment were defined as early. All patients enrolled presented with a BM or combined, early or very early, relapse.

The S3 group included children with B-cell precursor ALL early BM relapse, whereas the S4 group comprised children with B-cell precursor ALL experiencing very early BM or combined relapse and those with T-cell ALL BM relapse, irrespective of the time between diagnosis and relapse[2].

The inclusion of patients in this study was based exclusively on the following: (i) the attribution to the S3 or S4 risk group; (ii) the achievement of a second hematological and

morphological complete remission (CR); (iii) the availability of BM samples at relapse and at one or more subsequent time points, namely time points 1 (TP1), 2 (TP2) and 3 (TP3), that is, after 3–5 weeks (after the first reinduction cycle), 6–8 weeks (after the second chemotherapy cycle) and 9–16 weeks (after the third chemotherapy cycle) from the beginning of the relapse protocol.

Ten patients who relapsed during the period of study accrual but with no informative molecular marker (5 patients) or with a molecular marker that did not reach the required level of sensitivity of 10^{-4} (5 patients) were excluded from the study.

Written informed consent was obtained from the patients and/or their guardians, and the study was approved by the Institutional Review Boards of the participating centers.

Treatment protocol

All patients received reinduction chemotherapy as planned in the AIEOP ALL REC 2003 protocol for S3 and S4 groups. Briefly, children were randomized to receive either a 5-day first induction course with fludarabine (30 mg/m^2/day for 5 days), cytosine arabinoside (2 g/ m^2/day for 5 days) and lyposomal doxorubicine (50 mg/m^2/day for 3 doses) or cytosine arabinoside (3 g/m^2/day for 5 days) plus idarubicine 40 mg/ m^2 given in a single dose on day +3. Patients achieving CR received a second cycle comprising fludarabine and cytosine arabinoside. The third cycle comprised 2 weeks of oral dexametasone (6 mg/m^2/day), 4 weekly doses of vincristine (1.5 mg/m^2/dose i.v.) and idarubicine (6 mg/m^2/dose i.v.), 4 doses of Escherichia coli L-asparaginase (10 000 IU/m^2 i.m.) followed by one dose of cyclophosphamide (1 g/m^2 i.v.), 8 doses of cytosine arabinoside (75 mg/m^2/dose subcutaneously) and 14 days of oral 6-mercaptopurine (60 mg/m^2/day). Intrathecal methotrexate was given on days 2, 16, 31 and 38.

Owing to the severity of the prognosis for S3 and S4 patients, after induction and consolidation, all children had the indication to receive allogeneic hematopoietic stem cell transplantation (HSCT), this choice being supported by data from Borgmann *et al*[7]. The majority of treating centers performed transplants from matched family or unrelated donors, whereas transplants from HLA-partially matched (that is, haploidentical) family donors were performed only in a limited number of institutions with a specific program on haploidentical T-cell-depleted HSCT. For this reason, some children lacking a matched family or unrelated donor did not receive any form of allogeneic HSCT. In detail, 33 patients received allogeneic HSCT, 4 from HLA-matched family donor, 21 from compatible unrelated volunteer and 8

from HLA-haploidentical family donor. Patients who did not undergo allogeneic HSCT received maintenance chemotherapy according to the AIEOP ALL REC 2003 protocol.

Minimal residual disease results were not available to the treating physician and did not influence treatment decisions.

MRD analysis

DNA of BM mononuclear cells, obtained after Ficoll–Paque centrifugation, was extracted and purified using Gentra kit (Gentra System, Minneapolis, MN, USA). PCR analysis to detect specific TCRγ, TCRδ, and VDJH, DJH, VK and IRSS gene rearrangements was performed at diagnosis and again at the time of relapse using methods published elsewhere[15-17].

Clonal gene rearrangements, identified by homo/heteroduplex analysis, were sequenced by dye-terminator cycle sequencing kit on ABI Prism 310 (Applied Biosystems, Foster City, CA, USA)[18]. MRD levels in follow-up samples were analyzed by real-time quantitative PCR with hydrolysis (TaqMan) probes[19-22].

We used at least one molecular marker identified at the time of relapse. In detail, at relapse, all patients of this cohort conserved at least one of the original gene rearrangements. Only markers with a cutoff level of at least 10^{-4} were considered. In patients with more than one marker available, if MRD levels differed between the two markers, the highest MRD level was chosen for final MRD assessment. The real-time PCR analysis of BM samples was performed in accordance with the guidelines published by the European Study Group on MRD detection in ALL[23].

Minimal residual disease negativity was established when there was absence of specific amplification or amplification within 1 threshold cycle (Ct) of the background or amplification with a distance of more than 20 Ct's (ΔCt) from the undiluted diagnostic sample. A follow-up sample was considered as MRD positive if it showed a specific amplification product of more than 1 Ct lower than the background and separated less than 20 Ct's (ΔCt) from the undiluted diagnostic sample. MRD values were given if the sample gave a Ct within the QR of the PCR and the ΔCt between replicates was less than 1.5.

Definitions for MRD results

Patients were categorized, according to MRD results, in three groups: (i) positive (POS MRD): patients with MRD levels $\geq 10^{-4}$; (ii) positive not-quantifiable (POS-NQ MRD):

patients with positive MRD at a level $<10^{-4}$, very near to the sensitivity threshold of the technique and not precisely quantifiable; (iii) negative (NEG MRD): patients with completely negative MRD results.

Statistical analysis

Statistical analysis used 1 January 2008, as reference date.

Overall survival and EFS were calculated by the Kaplan–Meier method; comparisons between group were performed with the log-rank test[24]. Relapse incidence (RI) and treatment-related mortality (TRM) were calculated as cumulative incidence curves[25-27]. Overall survival, EFS, RI and TRM were estimated from the date of first relapse to the date of analysis or to the date of an adverse event. All results were expressed as 3-year probability or 3-year cumulative incidence (%) and 95% confidence interval (95% CI).

A univariate analysis of EFS, RI and TRM was performed for the whole study population. For multivariate analyses, the Cox proportional hazard regression model was used[28-29]. χ^2 test was used to compare differences in percentages.

All *P*-values are two-sided and values less than 0.05 were considered to be statistically significant. The SAS package (SAS Institute, Cary, NC, USA) and NCSS 2001 (Number Cruncher Statistical Systems, Kaysville, UT, USA) were used for analysis of the data.

Results

No statistical difference, in terms of presenting features or clinical outcome, was observed between the 60 patients analyzed and the 10 patients who relapsed during the study period but who were excluded due to lack of an informative marker (data not shown). For this reason, only the results of the analysis on the 60 patients fulfilling the inclusion criteria are presented.

Overall outcome

Median observation time for surviving patients is 15 months (range, 4–48 months).

All 60 patients achieved a second morphological remission after the first induction course, this being one of the inclusion criteria. Overall, 27 of the 60 patients (45%) are alive: 26 (43%) are in second CR, whereas 1 child is alive after a subsequent BM relapse (2%). The

estimated 3-year overall survival probability is 32% (95% CI, 16–47), whereas the 3-year EFS probability is 27% (95% CI, 8–45) (Figure 1a).

Twenty-six patients (43%) had a second leukemia relapse, at a median of 5 months after the first one (range, 1–31 months). The 3-year cumulative RI is 59% (95% CI, 42–81) (Figure 1b).

Eight patients (13%) died in second CR due to treatment-related causes, at a median of 115 days after the first relapse (range, 55–259). The estimated 3-year cumulative incidence of TRM is 15% (95% CI, 8–28) (Figure 1b).

MRD evaluation

On the whole, 119 MRD analyses were performed on the 60 patients studied, using 70 markers that had already been identified at diagnosis and 21 at relapse. MRD was evaluated at TP1 in 45 cases (75%), at TP2 in 28 cases (47%) and at TP3 in 46 cases (77%). Only 11 patients (18%) had only one time point analyzed, 10 (17%) had all the three estimations and 39 (65%) had two time points evaluated.

As summarized in Table 2, 56% of the tests at TP1 were POS, 24% were POS-NQ and 20% were NEG. At TP3, 48% of the tests were still POS, 22% were POS-NQ and only 30% were NEG (P=NS comparing TP1 vs TP3 results). In particular, considering only the 49 patients with at least two time points studied, we can observe that:

- all the nine children already NEG at TP1 remained NEG at the subsequent time points;
- of the 12 patients with a first POS-NQ result, 5 (42%) became negative at a subsequent control, 5 (42%) remained POS-NQ, whereas only 2 (16%) presented a subsequent worse result, becoming POS;
- considering the 28 patients with a first POS result, only 1 (4%) improved to NEG, 6 cases (21%) partially improved to POS-NQ, whereas in the remaining 21 patients (75%), MRD remained POS also at the following controls.

Impact MRD on clinical outcome

Table 4 shows the results of univariate analysis on EFS. Considering patients' characteristics, a shorter time interval from diagnosis to first relapse was associated with a worse outcome, EFS being 19% (95% CI, 2–36) for very early relapses vs 38% (95 CI, 12–64) for early relapses (P=0.001). The impact of immunophenotype was not statistically

significant, although EFS probability for B-lineage ALL was better than that of T-lineage (30% (95% CI, 9–51) versus 20% (95% CI, 0–44), respectively, P=0.053). Patients belonging to the S3 BFM risk group had a better EFS as compared with S4 patients (35% (95% CI, 3–67) vs 23% (95% CI, 7–39); P=0.0036). Finally, even though this comparison could not be adjusted for patients' clinical features due to the limited number of cases, children who received an allogeneic HSCT (33 cases) had a better EFS as compared with the 27 who did not, EFS being 36% (95% CI, 11–60) and 14% (95% CI, 0–35), respectively (P<0.005).

A NEG or a POS-NQ MRD result was associated with a better outcome. Fourteen of the 17 patients who had at least one NEG MRD result at any time point (82%) are alive and in remission (the remaining three died in remission due to treatment-related complications), and 8 of the 15 children with at least a POS-NQ result at any time point (53%) are alive and disease free. On the contrary, only 4 out of the 28 patients with all POS MRD results (14%), including children with a single evaluation, are alive and disease free (P<0.001).

The influence of MRD result on EFS is shown in Figure 2. The 3-year EFS was 79% (95% CI, 52–100) for patients with a NEG MRD result at first time point analyzed, 43% (95% CI, 4–82) for those with a POS-NQ MRD and 18% (95% CI, 4–32) for children with POS MRD at first time point analyzed (P=0.005). When available, TP1 was also already predictive of the probability of EFS in univariate analysis. In fact, EFS was 73% (95% CI, 41–100), 45% (95% CI, 4–87) and 19% (95% CI, 3–35) for patients with NEG, POS-NQ and POS result at TP1, respectively (P<0.05).

The effect of MRD on patients' EFS was due to a significantly different risk of relapse among the three MRD groups. In fact, considering TP1, RI was 0%, 43% (95% CI, 17–100) and 73% (95% CI, 57–94) for NEG, POS-NQ and POS patients, respectively (P=0.003). Furthermore, considering the first time point available, RI was 0%, 48% (95% CI, 21–100) and 67% (95% CI, 53–86) for NEG, POS-NQ and POS patients, respectively (P=0.0019). On the contrary, none of the MRD time points was associated with differences in TRM (data not shown).

As shown above, only 12 out of the 40 patients with multiple MRD analyses improved their result from POS-NQ to NEG or from POS to POS-NQ or NEG (the remaining 9 cases with multiple analyses were already negative at the first control). EFS of the 12 cases with improving MRD was significantly better than that of patients with stable MRD positivity: 67% (95% CI, 35–98) vs 12% (95% CI, 0–26); P<0.01.

Finally, when a separate analysis for patients who received or did not receive HSCT was performed, the effect of MRD on outcome was similar to that observed on the whole

study population, even though the lower number of patients within each subgroup decreased the power of the statistical tests. For the HSCT group, EFS was 86, 44 and 26%, respectively, for patients NEG, POS-NQ and POS at first time point available (P=0.09), whereas in children who did not receive HSCT, EFS was 67, 50 and 9%, respectively, for patients NEG, POS-NQ and POS at first time point available (P=NS).

The predictive value of MRD on clinical outcome was also confirmed by the multivariate analysis (see Table 4). The relative risk of treatment failure for patients POS at the first time point available vs NEG patients was 4.40 (95% CI, 1.01–19.33; P<0.05). On the contrary, the relative risk of treatment failure for the POS-NQ group vs the NEG group was not statistically significant (RR=2.04; 95% CI, 0.39–10.59; P=NS). The only other factors associated with an increased risk of treatment failure in multivariate analysis were BFM classification (relative risk for S4 vs S3 patients=2.87 (95% CI, 1.31–6.30; P<0.01)) and allogeneic HSCT (relative risk for patients given HSCT vs patient not given HSCT=0.28 (95% CI, 0.13–0.59; P<0.005)).

Discussion

Despite the advances in the treatment of childhood ALL, approximately 20–25% of children suffer leukemia relapse[30-33]. The present salvage protocols offer a probability of second CR between 70 and 95%, but reported cure rates after a first relapse are between 30 and 40%[3-4,34]. In particular, for patients of the subgroups with poorer prognosis, the reported cure rate is about 20% if allogeneic HSCT is employed and below 5% for children not transplanted[2,6]. In all published studies, the most frequent adverse event was a further BM relapse.

In our cohort, the 3-year EFS was 27% and a second relapse was the most frequent adverse event. For this reason, a major task of future clinical trials will be to predict more precisely the occurrence of relapses, thus identifying patients with the indication to alternative therapies aimed at reducing the tumor burden before an allograft.

Minimal residual disease is a powerful parameter for the prediction of relapse in patients with ALL receiving first-line chemotherapy[10,35-37]. Nevertheless, only very few studies on a limited number of patients studied MRD and its prognostic value during second-line chemotherapy for relapsed ALL[14,38]. In the retrospective study by Eckert *et al.* [14], only patients belonging to the intermediate prognosis group (S2, according to the BFM risk group stratification) were evaluated, whereas the study published by Coustan-Smith *et al.*[38] also included high-risk patients, but employed a different methodology, based on flow cytometry.

However, both studies suggested that MRD analysis could predict long-term outcome in relapsed childhood ALL and could be used to guide the choice of post-remission treatment.

In our hands, MRD measurement using PCR in high-risk relapsed ALL permitted the differentiation of patients with a high risk of relapse from those with a good probability of cure. EFS was more than 70% for patients with NEG MRD, about 40% for those with POS-NQ results and about 20–30% for children with POS MRD. Furthermore, MRD analysis discriminated between good and poor prognosis patients already at TP1, after only one cycle of chemotherapy. In fact, considering only TP1 results, the 3-year EFS was 73, 45 and 19% for NEG, POS-NQ and POS patients, respectively ($P<0.05$).

When a separate analysis for patients who received or did not receive HSCT was carried out, the prognostic impact of MRD on outcome was observed in both subgroups of patients. This observation is also in agreement with previously published studies pointing out the predictive function of MDR assessed immediately before allogeneic HSCT[39-41].

Another important finding was that, at least with the chemotherapy adopted in our protocol, only a minority of patients with POS MRD at TP1 improved their MRD status with a second or third chemotherapy course. In fact, we observed only an 8% decrease in the percentage of POS patients from TP1 to TP3 (56 vs 48%, respectively, *P*=NS), and analyzing 49 patients with two or more time points available, only one child (4%) with a POS first result achieved a NEG result at a subsequent TP analysis, and only six children (21%) improved from POS to POS-NQ.

The importance of a POS result at the first MRD evaluation was confirmed in multivariate analysis, the S4 group of BFM classification being the only other additional factor associated with a worse outcome.

Even though the observation time for our population is relatively short (the median follow-up time is 15 months), it must be considered that only patients belonging to the highest risk groups were enrolled, and that all of them experienced a very early (48%) or early (52%) first leukemia recurrence. Moreover, for relapsing patients, the time from first to second leukemia relapse was very short, the median interval being only 5 months. For this reason, it is reasonable to expect only a very low number of further relapses, even with a longer follow-up.

In conclusion, our study shows that MRD analysis by PCR is a feasible and promising approach to stratify patients also in protocols for relapsed, high-risk ALL. MRD evaluation was able to identify patients who could benefit from the treatment and with a probability of cure of about 70%. On the other hand, a POS first MRD result identified, after just one

chemotherapy cycle, a subset of patients who seem not to benefit from further consolidations, including allogeneic HSCT. These patients with extremely severe prognosis could have the indication to experimental treatment strategies on the basis of the use of novel antileukemia agents, such as clofarabine[42], nelarabine[43] or anti-CD22 monoclonal antibodies[44], aimed at obtaining a better leukemia control before allogeneic HSCT.

References

1. Pui, C.-H. & Evans, W.E. Treatment of acute lymphoblastic leukemia. *N Engl J Med* **354**, 166-178 (2006).
2. Henze, G., *et al.* Six-year experience with a comprehensive approach to the treatment of recurrent childhood acute lymphoblastic leukemia (ALL-REZ BFM 85). A relapse study of the BFM group. *Blood* **78**, 1166-1172 (1991).
3. Gaynon, P.S., *et al.* Survival after relapse in childhood acute lymphoblastic leukemia: impact of site and time to first relapse--the Children's Cancer Group Experience. *Cancer* **82**, 1387-1395 (1998).
4. Lawson, S.E., *et al.* The UK experience in treating relapsed childhood acute lymphoblastic leukaemia: a report on the medical research council UKALLR1 study. *Br J Haematol* **108**, 531-543 (2000).
5. Testi, A.M., *et al.* A single high dose of idarubicin combined with high-dose ARA-C for treatment of first relapse in childhood 'high-risk' acute lymphoblastic leukaemia: a study of the AIEOP group. *Br J Haematol* **118**, 741-747 (2002).
6. Einsiedel, H.G., *et al.* Long-term outcome in children with relapsed ALL by risk-stratified salvage therapy: results of trial acute lymphoblastic leukemia-relapse study of the Berlin-Frankfurt-Munster Group 87. *J Clin Oncol* **23**, 7942-7950 (2005).
7. Borgmann, A., *et al.* Unrelated donor stem cell transplantation compared with chemotherapy for children with acute lymphoblastic leukemia in a second remission: a matched-pair analysis. *Blood* **101**, 3835-3839 (2003).
8. Chessells, J.M., Leiper, A.D. & Richards, S.M. A second course of treatment for childhood acute lymphoblastic leukaemia: long-term follow-up is needed to assess results. *Br J Haematol* **86**, 48-54 (1994).
9. Giona, F., *et al.* ALL R-87 protocol in the treatment of children with acute lymphoblastic leukaemia in early bone marrow relapse. *Br J Haematol* **99**, 671-677 (1997).
10. Jacquy, C., *et al.* A prospective study of minimal residual disease in childhood B-lineage acute lymphoblastic leukaemia: MRD level at the end of induction is a strong predictive factor of relapse. *Br J Haematol* **98**, 140-146 (1997).
11. Cazzaniga, G. & Biondi, A. Molecular monitoring of childhood acute lymphoblastic leukemia using antigen receptor gene rearrangements and quantitative polymerase chain reaction technology. *Haematologica* **90**, 382-390 (2005).
12. Flohr, T., *et al.* Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* **22**, 771-782 (2008).

13. Steenbergen, E.J., *et al.* Prolonged persistence of PCR-detectable minimal residual disease after diagnosis or first relapse predicts poor outcome in childhood B-precursor acute lymphoblastic leukemia. *Leukemia* **9**, 1726-1734 (1995).
14. Eckert, C., *et al.* Prognostic value of minimal residual disease in relapsed childhood acute lymphoblastic leukaemia. *Lancet* **358**, 1239-1241 (2001).
15. Szczepanski, T., *et al.* Ig heavy chain gene rearrangements in T-cell acute lymphoblastic leukemia exhibit predominant DH6-19 and DH7-27 gene usage, can result in complete V-D-J rearrangements, and are rare in T-cell receptor alpha beta lineage. *Blood* **93**, 4079-4085 (1999).
16. van Dongen, J.J., *et al.* Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 Concerted Action: investigation of minimal residual disease in acute leukemia. *Leukemia* **13**, 1901-1928 (1999).
17. van Dongen, J.J.M., *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
18. Germano, G., Songia, S., Biondi, A. & Basso, G. Rapid detection of clonality in patients with acute lymphoblastic leukemia. *Haematologica* **86**, 382-385 (2001).
19. Verhagen, O.J., *et al.* Application of germline IGH probes in real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* **14**, 1426-1435 (2000).
20. van der Velden, V.H., *et al.* Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* **16**, 928-936 (2002).
21. van der Velden, V.H.J., *et al.* Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* **17**, 1013-1034 (2003).
22. van der Velden, V.H.J., Wijkhuijs, J.M., Jacobs, D.C.H., van Wering, E.R. & van Dongen, J.J.M. T cell receptor gamma gene rearrangements as targets for detection of minimal residual disease in acute lymphoblastic leukemia by real-time quantitative PCR analysis. *Leukemia* **16**, 1372-1380 (2002).
23. van der Velden, V.H.J., *et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* **21**, 604-611 (2007).
24. Kaplan E.L., *et al.* Nonparametral estimation from incomplete observations. *Am Stat Assoc* **53**(1958).
25. Pepe, M.S., *et al.* Summarizing data on survival, relapse, and chronic graft-versus-host disease after bone marrow transplantation: motivation for and description of new methods. *Br J Haematol* **83**, 602-607 (1993).
26. Gooley, T.A., Leisenring, W., Crowley, J. & Storer, B.E. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* **18**, 695-706 (1999).
27. Klein JP, R.J., Zhang MJ, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: unadjusted analysis. *Bone Marrow Transplant* **2001**, 909-915 (2001).
28. Cox, D. Regression models and life tables. *J Royal Stat Soc B* **34**(1972).

29. Klein, J.P., Rizzo, J.D., Zhang, M.J. & Keiding, N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part 2: Regression modeling. *Bone Marrow Transplant* **28**, 1001-1011 (2001).
30. Harms DO, J.-S.G. Co-operative study group for childhood acute lymphoblastic leukemia (COALL): long-term follow-up of trials 82, 85, 89 and 92. *Leukemia* **14**, 2234-2239 (2000).
31. Schrappe, M., *et al.* Long-term results of four consecutive trials in childhood ALL performed by the ALL-BFM study group from 1981 to 1995. Berlin-Frankfurt-Munster. *Leukemia* **14**, 2205-2222 (2000).
32. Silverman, L.B., *et al.* Improved outcome for children with acute lymphoblastic leukemia: results of Dana-Farber Consortium Protocol 91-01. *Blood* **97**, 1211-1218 (2001).
33. Pui, C.-H., *et al.* Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood* **104**, 2690-2696 (2004).
34. Buchanan, G.R., *et al.* Alternating drug pairs with or without periodic reinduction in children with acute lymphoblastic leukemia in second bone marrow remission: a Pediatric Oncology Group Study. *Cancer* **88**, 1166-1174 (2000).
35. Panzer-Grumayer, E.R., Schneider, M., Panzer, S., Fasching, K. & Gadner, H. Rapid molecular response during early induction chemotherapy predicts a good outcome in childhood acute lymphoblastic leukemia. *Blood* **95**, 790-794 (2000).
36. Cave, H., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--Childhood Leukemia Cooperative Group. *N Engl J Med* **339**, 591-598 (1998).
37. van Dongen, J.J., *et al.* Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* **352**, 1731-1738 (1998).
38. Coustan-Smith, E., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. *Leukemia* **18**, 499-504 (2004).
39. Krejci, O., *et al.* Level of minimal residual disease prior to haematopoietic stem cell transplantation predicts prognosis in paediatric patients with acute lymphoblastic leukaemia: a report of the Pre-BMT MRD Study Group. *Bone Marrow Transplant* **32**, 849-851 (2003).
40. Bader, P., *et al.* Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. *Leukemia* **16**, 1668-1672 (2002).
41. Knechtli, C.J., *et al.* Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. *Blood* **92**, 4072-4079 (1998).
42. Jeha, S., *et al.* Phase II study of clofarabine in pediatric patients with refractory or relapsed acute lymphoblastic leukemia. *J Clin Oncol* **24**, 1917-1923 (2006).
43. DeAngelo, D.J., *et al.* Nelarabine induces complete remissions in adults with relapsed or refractory T-lineage acute lymphoblastic leukemia or lymphoblastic lymphoma: Cancer and Leukemia Group B study 19801. *Blood* **109**, 5136-5142 (2007).
44. Wayne AS, F.H., Lew G. Targeting CD22 inchildhood B-precursor acute lymphoblastic leukemia (pre-BALL): pre-clinical studies and phase I trial of the anti-CD22immunotoxin CAT-3888 (BL22). *Blood, ASH Abstracts* **110** (2007).

Figures

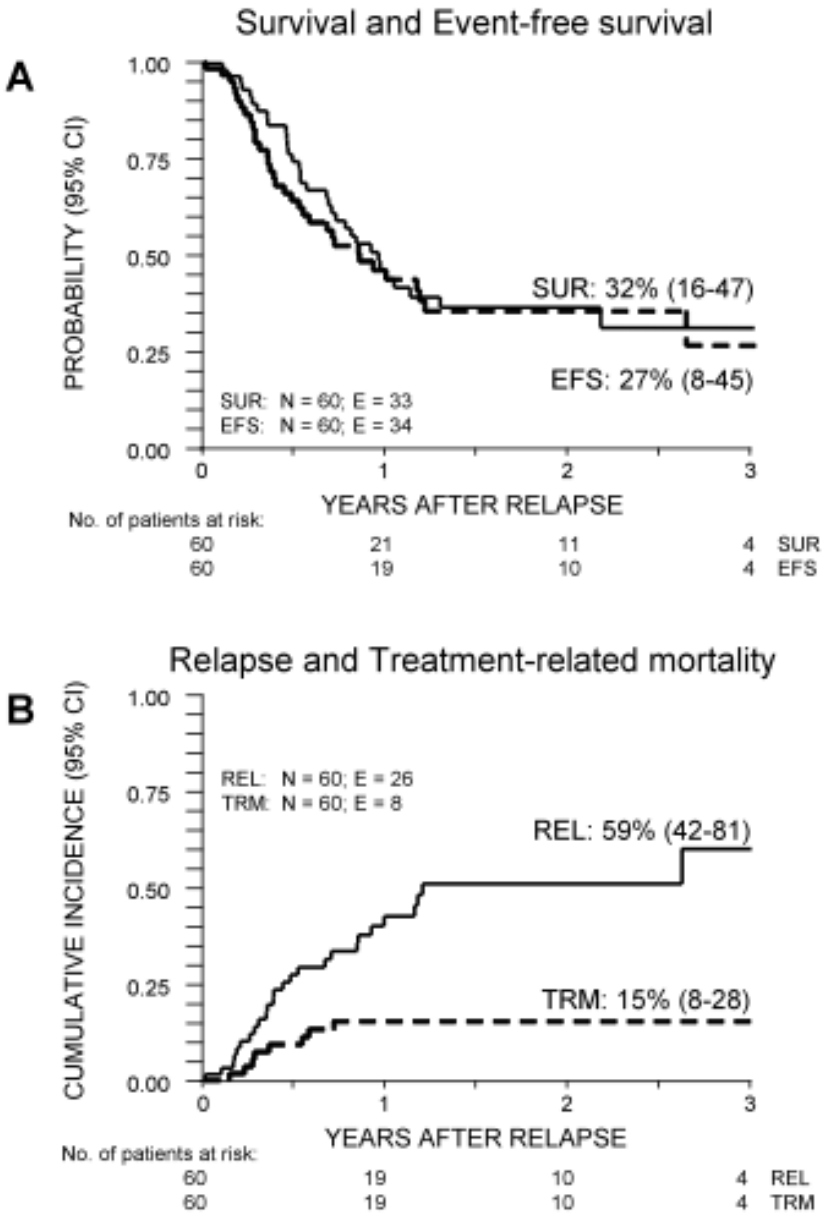


Figure 1. Overall probability of survival (SUR) and event-free survival (EFS) (Figure 1A) and overall cumulative incidence of second relapse (REL) or of treatment-related mortality (TRM) (Figure 1B), from time of first relapse, for the 60 pediatric patients analyzed.

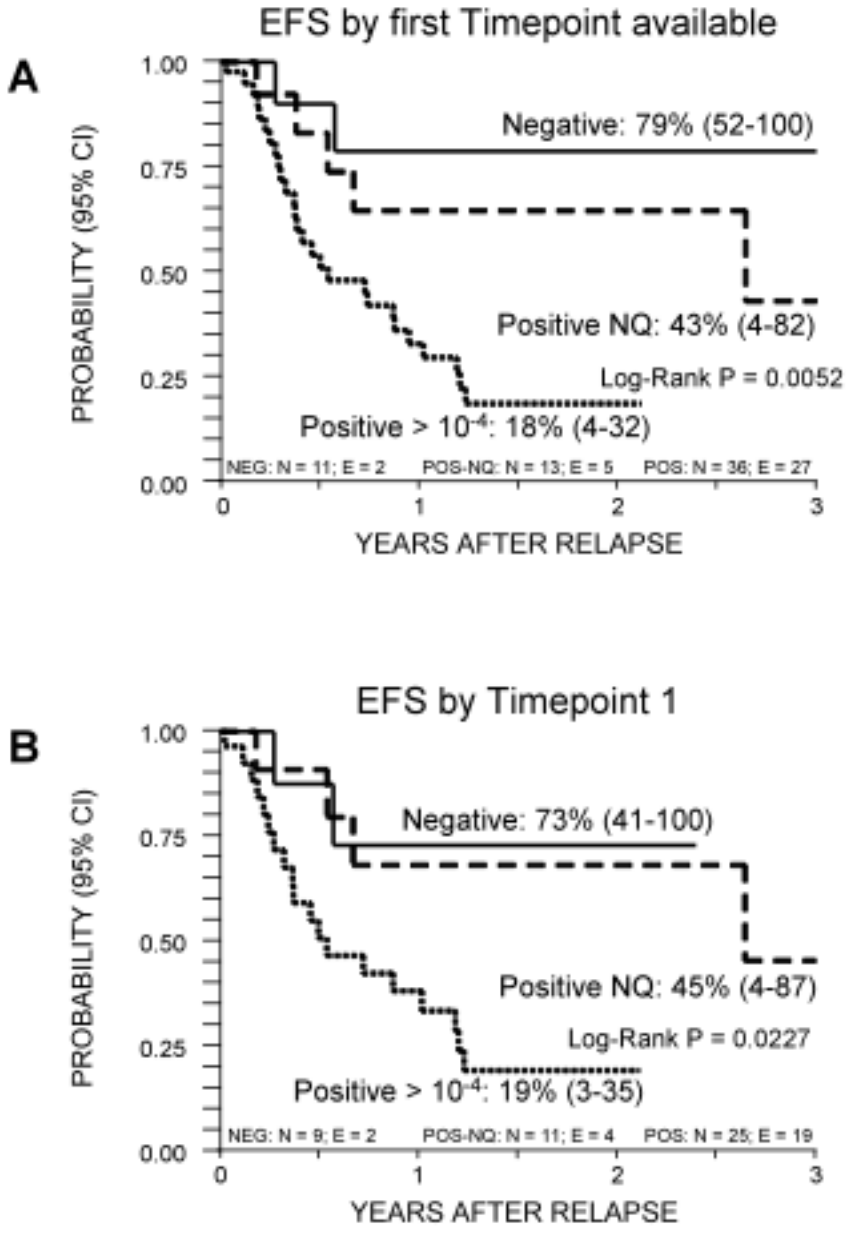


Figure 2. Probability of event-free survival (EFS) according to the results of MRD analysis at time point 1 (Figure 2A) and at first time point available (Figure 1B).

Tables

Table 1. Patients' characteristics

Number of patients enrolled	60 (100%)
Gender: Male / Female	37 / 23
Age at diagnosis (years)	5 (0.6-17)
Period of diagnosis:	January 2001 – June 2006
Period of relapse	June 2003 – December 2006
Phenotype:	
B lineage	45 (75%)
T lineage	15 (25%)
Cytogenetic abnormalities:	
t(9;22)	1 (1.6%)
t(4;11)	2 (3.3%)
First line treatment:	
AIEOP ALL 2000	60 (100%)
Time from diagnosis to relapse (months)	18 (5-29)
Very early relapses	29 (48%)
Early relapses	31 (42%)
Type of relapse:	
Bone marrow only	54 (90%)
Combined	6 (10%)
BFM risk group stratification:	
S3	25 (42%)
S4	35 (58%)
Relapse therapy protocol:	
AIEOP ALL REC 2003	60 (100%)
Consolidation with HSCT:	
No	27 (45%)
Yes, matched family donor	4 (7%)
Yes, unrelated donor	21 (35%)
Yes, HLA-haploidentical family donor	8 (13%)
Time from relapse to HSCT (months)	5 (2-10)

Data are expressed as median and range or as absolute number and percentage, as appropriate
ALL: acute lymphoblastic leukemia; HSCT: Hematopoietic Stem Cell Transplantation

Table 2. Results of MRD analysis

	Evaluated (%)	POS≥ 10^{-4} (%)	POS-NQ (%)	NEG (%)	P value
Time point 1	45 (75)	25 (56)	11 (24)	9 (20)	
Time point 2	28 (47)	17 (61)	4 (14)	7 (25)	
Time point 3	46 (77)	22 (48)	10 (22)	14 (30)	N.S.
First time point available	60 (100)	36 (60)	13 (22)	11 (18)	
Last time point available	60 (100)	31 (48)	12 (20)	17 (28)	N.S.

POS ≥ 10-4: patients with MRD level ≥ 10-4. POS-NQ: patients with positive MRD result, even though at a level < 10-4, very near to the sensitivity threshold of the technique and not precisely quantifiable. NEG: patients with completely negative MRD analysis results.

Time point 1: after the first chemotherapy cycle (after 3 – 5 weeks from relapse). Time point 2: after the second chemotherapy cycle (6 – 8 weeks from relapse. Time point 3: after the third chemotherapy course (9 – 16 weeks from relapse). First time point available: the first time point available for each single patient, irrespective of the number of chemotherapy cycles administered. Last time point available: the last time point available for each single patient, irrespective of the number of chemotherapy cycles administered.

Table 3. Detailed results of MRD analysis and clinical outcome of the patients

Patient No.	BFM Group	Time point 1	Time point 2	Time point 3	Outcome
1	S3	NEG		NEG	Alive & well, no HSCT
2	S4	NEG		NEG	Alive & well, no HSCT
3	S4	NEG	NEG	NEG	Alive & well, no HSCT
4	S3	NEG		NEG	Dead, TRM, no HSCT
5	S3	NEG	NEG		Alive & well after HSCT
6	S3	NEG		NEG	Alive & well after HSCT
7	S4	NEG	NEG	NEG	Alive & well after HSCT
8	S3	NEG	NEG	NEG	Alive & well after HSCT
9	S3	NEG		NEG	Dead, TRM after HSCT
10	S3			NEG	Alive & well after HSCT
11	S4			NEG	Alive & well after HSCT
12	S3	POS-NQ		NEG	Alive & well, no HSCT
13	S4	POS-NQ	NEG	NEG	Alive & well after HSCT
14	S3	POS-NQ		NEG	Alive & well after HSCT
15	S3	POS-NQ	NEG		Alive & well after HSCT
16	S3	POS-NQ	NEG		Dead, TRM after HSCT
17	S3	POS-NQ	POS-NQ		Alive after relapse after HSCT
18	S3	POS-NQ		POS-NQ	Alive & well after HSCT
19	S3	POS-NQ		POS-NQ	Alive & well after HSCT
20	S4	POS-NQ		POS-NQ	Dead, disease progression after HSCT
21	S4			POS-NQ	Alive & well, no HSCT
22	S4	POS-NQ	POS	POS-NQ	Alive & well, no HSCT
23	S4	POS-NQ		POS	Dead, disease progression after HSCT
24	S4		POS	NEG	Alive & well after HSCT
25	S4	POS	POS-NQ		Alive & well, no HSCT
26	S3	POS		POS-NQ	Alive & well, no HSCT
27	S4	POS		POS-NQ	Alive & well after HSCT
28	S3	POS		POS-NQ	Dead, disease progression after HSCT
29	S3		POS	POS-NQ	Alive & well, no HSCT
30	S4		POS	POS-NQ	Dead, TRM, no HSCT
31	S4	POS			Dead, TRM, no HSCT
32	S4	POS			Dead, disease progression, no HSCT
33	S4	POS			Dead, disease progression, no HSCT
34	S4	POS			Dead, disease progression after HSCT
35	S4	POS			Alive & well after HSCT
36	S4	POS	POS		Dead, TRM, no HSCT
37	S3	POS	POS		Alive & well after HSCT
38	S4		POS		Dead, disease progression after HSCT
39	S4		POS		Dead, TRM, no HSCT
40	S3		POS	POS	Dead, disease progression after HSCT
41	S4		POS	POS	Dead, disease progression after HSCT
42	S4		POS	POS	Dead, disease progression, no HSCT
43	S4	POS	POS-NQ	POS	Dead, disease progression, no HSCT
44	S3			POS	Dead, TRM after HSCT
45	S3		POS-NQ	POS	Dead, disease progression after HSCT
46	S4		POS	POS	Dead, disease progression, no HSCT
47	S4		POS	POS	Alive & well, no HSCT
48	S3	POS		POS	Dead, disease progression, no HSCT
49	S4	POS		POS	Dead, disease progression after HSCT
50	S3	POS		POS	Dead, disease progression, no HSCT
51	S4	POS		POS	Dead, disease progression after HSCT
52	S4	POS		POS	Dead, disease progression after HSCT
53	S4	POS		POS	Dead, disease progression, no HSCT
54	S4	POS		POS	Dead, disease progression, no HSCT
55	S4	POS		POS	Dead, disease progression, no HSCT
56	S3	POS		POS	Dead, disease progression, no HSCT
57	S4	POS	POS	POS	Dead, disease progression, no HSCT
58	S4	POS	POS	POS	Dead, disease progression after HSCT
59	S3	POS	POS	POS	Alive & well after HSCT
60	S4	POS	POS	POS	Dead, disease progression after HSCT

Abbreviations: HSCT: hematopoietic stem cell transplantation; TRM: treatment-related mortality.

Time point 1: after the first chemotherapy cycle (after 3–5 weeks from relapse). Time point 2: after the second chemotherapy cycle (6–8 weeks from relapse). Time point 3: after the third chemotherapy course (9–16 weeks from relapse).

POS: patients with MRD level ≥10-4; POS-NQ: patients with positive MRD result, even though at a level <10-4, very near to the sensitivity threshold of the technique and not precisely quantifiable; NEG: patients with completely negative MRD analysis results.

Table 4. Univariate and multivariate analysis for 3-year event-free survival

Univariate Analysis					
Variable	**N**	**E**	**Probability**	**(95% CI)**	**P**
Gender:					
Male	37	23	15%	(0 – 38)	N.S.
Female	23	11	46%	(24 – 69)	
Age at diagnosis:					
< 5 years	30	16	35%	(11 – 44)	N.S.
≥ 5 years	30	18	18%	(0 – 46)	
Phenotype:					
B lineage	45	25	30%	(9 – 51)	0.053
T lineage	15	9	20%	(0 – 44)	
Time to 1st relapse:					
< 18 months	29	20	19%	(2 – 36)	< 0.005
18 – 30 months	31	14	38%	(12 – 64)	
Type of relapse:					
Combined	6	1	80%	(45 – 100)	N.S.
Bone marrow	54	33	25%	(7 – 42)	
BFM Risk group:					
S3	25	11	35%	(3 – 67)	< 0.01
S4	35	23	23%	(7 – 39)	
Allogeneic HSCT:					
No	27	17	14%	(0 – 35)	< 0.001
Yes	33	17	36%	(11 - 60)	
MRD at time-point 1:					
Negative	9	2	73%	(41 – 100)	< 0.05
Positive NQ	11	4	45%	(4 – 87)	
Positive	25	19	19%	(3 – 35)	
MRD at time-point 2:					
Negative	7	1	83%	(54 – 100)	N.S.
Positive NQ	4	3	33%	(0 – 87)	
Positive	17	11	25%	(2 – 43)	
MRD at time-point 3:					
Negative	14	2	83%	(60 – 100)	< 0.005
Positive NQ	10	3	62%	(27 – 97)	
Positive	22	20	5%	(0 – 14)	
MRD at first time-point available:					
Negative	11	2	79%	(52 – 100)	< 0.01
Positive NQ	13	5	43%	(4 – 82)	
Positive	36	27	18%	(4 – 32)	
MRD at last time-point available:					
Negative	17	3	79%	(58 – 100)	< 0.001
Positive NQ	12	4	44%	(3 – 86%)	
Positive	31	27	8%	(8 – 18%)	

Multivariate Analysis			
Variable	**Relative Risk of treatment failure**	**(95% CI)**	**P**
BFM Risk Group			
S4 vs. S3	2.87	(1.31 – 6.30)	< 0.01
MRD at 1st time point available			
Positive NQ vs. Negative	2.04	(0.39 – 10.59)	N.S.
Positive vs. Negative	4.4	(1.01 – 19.33)	< 0.05
Allogeneic HSCT			
Yes vs. no	0.28	(0.13 – 0.59)	< 0.005

Abbreviation: NS: non-significant.

1.2 Molecular relapse is highly predictive of clinical recurrence in children with acute lymphoblastic leukemia enrolled in the AIEOP-BFM ALL 2000 protocol
(paper in preparation)

Giulia Fabbri[1], Maddalena Paganin[1], Katia Polato[1], Emanuela Giarin[1], Daniel Nardo[1], Elena Barisone[2], Gianni Cazzaniga[3], Andrea Biondi [3], Franco Locatelli [4], Giuseppe Basso[1]

[1]*Laboratorio di Ematologia e Oncologia Pediatrica, Università di Padova, Padova, Italy.*

[2] *Dipartimento di Pediatria, Ospedale Infantile Regina Margherita, Torino, Italy.*

[3]*Centro Ricerca Tettamanti and Clinica Pediatrica, Università di Milano Bicocca, Monza, Italy.*

[4]*Oncoematologia Pediatrica, Fondazione IRCCS Policlinico San Matteo, Università di Pavia, Pavia, Italy.*

In adults with acute lymphoblastic leukemia (ALL), detection of minimal residual disease (MRD) positive within the quantitative range (Pos Q) during follow-up is highly predictive of subsequent hematologic relapse. To test whether the same is true in children, we studied MRD by real-time quantitative PCR in bone marrow samples collected from two groups of pediatric ALL patients. We monitored MRD during and after treatment in 113 patients enrolled in the AIEOP-BFM ALL 2000 trial. Six out of the 7 patients who had one or more MRD Pos Q results subsequently developed hematologic relapse (95% confidence interval, 1.85-125.5, *P* < .001). We then performed a case-control study of 37 relapsed ALL children matched with 37 ALL controls in continuous complete remission. MRD Pos Q was found in 11 relapsed children vs. 1 non-relapsed child (*P* < .005, odds ratio, 15.23). Relapse of ALL could have been predicted by MRD monitoring in 82% of patients who were evaluated within 3 months before relapse occurrence. The results of this study suggest that detection of Pos Q MRD during follow-up is highly predictive of subsequent hematologic relapse of pediatric ALL. Accuracy of such a prediction improves by performing monitoring every 3 months.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common cancer in children and adolescents[1]. Nowadays, through the application of reliable prognostic factors permitting to use risk-oriented treatment protocols, around 80% of children with ALL can be cured[1]. The most common cause of treatment failure in pediatric ALL is relapse, which occurs in 20% of patients[2]. Unfortunately, rescue therapies, including allogeneic hematopoietic stem cell transplantation (allo-HSCT), are inadequate in most cases, and the outcome remains poor: Indeed, the overall cure rate of relapsed patients ranges between 30 and 60%, depending on the site of relapse, time from diagnosis to relapse, and ALL immunophenotype[2-3].

Early detection of recurrence could represent an innovative approach capable of improving the chance of being rescued of children with relapsed ALL. Monitoring minimal residual disease (MRD) by a highly sensitive molecular PCR approach or multiparameter flow cytometry has been shown to be clinically relevant to prognostic stratification of newly diagnosed[4-10] and relapsed[11-13] ALL patients, as well as to predict outcome in patients undergoing allo-HSCT[14-20]. MRD monitoring based on the detection of immune gene

rearrangements by real-time quantitative PCR (RQ-PCR) is feasible in almost 85% of patients with ALL[21].

MRD evaluation is being used in several front-line protocols to select the most appropriate treatment by stratifying patients in different classes of risk, and this stratification is expected to contribute to improve the overall outcome. Despite the availability of sophisticated methods for detecting the persistence/re-emergence of the malignant clone, medullary relapse is still currently defined employing morphologic criteria, i.e. presence of a blast percentage exceeding 25% in a bone marrow (BM) sample after having obtained complete remission (CR). To date there are only a few studies suggesting that it is possible to predict impending morphologic relapse through MRD monitoring in hematologic malignancies. In this regard, recently, Raff *et al*. demonstrated that conversion to MRD positivity during the phase of early post-consolidation in adult standard-risk ALL patients is strongly predictive of subsequent morphologic relapse[22]. Earlier diagnosis of relapse through MRD monitoring should in principle improve the efficacy of salvage treatment by starting chemotherapy in the presence of a more limited tumor burden. Indeed, Lo Coco *et al*. demonstrated that, in patients with acute promyelocytic leukemia, the administration of rescue therapy at the time of molecular relapse improves the chance of rescuing patients[23].

The aims of this study were: (i) to evaluate the significance of MRD in pediatric ALL patients as an indicator of impending recurrence of hematologic disease; and (ii) to determine the critical levels of MRD that could predict relapse. Moreover, we wanted to evaluate whether it is clinically advantageous to extend the period of MRD monitoring, which is currently applied only during the early phase of treatment. For this purpose, we analyzed BM samples from pediatric patients who were enrolled in the AIEOP-BFM ALL 2000 clinical trial.

Patients, materials, and methods

Patients

Patients were treated according to the AIEOP-BFM ALL 2000 protocol. The diagnosis of ALL was confirmed through morphologic examination and immunophenotyping in a reference national, centralized laboratory. Initially, we studied a cohort of 113 patients, representative of all patients enrolled in the AIEOP-BFM ALL 2000 protocol in two AIEOP centers (21 patients with relapsed ALL and 92 patients in continuous complete remission

[CCR]). We subsequently studied a further cohort of 47 relapsed patients who were treated in other AIEOP centers.

All patients were less than 18 years of age at time of diagnosis. They were stratified into standard-, intermediate-, and high-risk groups according to (i) MRD data obtained at the end of the induction therapy (time points 1 and 2, corresponding to days +33 and +78, respectively), (ii) prednisone response on day +8, (iii) morphological and clinical remission after induction therapy, and (iv) presence of high-risk chromosomal translocations, namely, t(9;22) and t(4;11). Clinical relapse was defined on the basis of classical morphological criteria, i.e., when a blast percentage ≥25% was detected in a BM smear after CR had been achieved. Patients in relapse were stratified according to (i) duration of first remission, (ii) site of relapse, and (iii) immunophenotype (B-lineage or T-lineage ALL [B- or T-ALL]) in medium- and high-risk groups according to the AIEOP REC 2003 study protocol[24].

After having obtained informed consent from parents or legal guardians, children with newly diagnosed ALL were recruited into the therapeutic protocol AIEOP-BFM ALL 2000, which was approved by the respective institutional ethics committees[25].

Patients were chosen and included in the study on the basis of the availability of BM samples collected at various time points during and after treatment (see "Cell samples and DNA isolation" below). In addition, they had to have at least one marker measurable through RQ-PCR with a sensitivity of at least 10^{-4}.

MRD assessment

Cell samples and DNA isolation. Besides the BM samples collected at diagnosis and at the 2 time points needed to measure MRD for patient stratification, as requested by the protocol, to be eligible for this first study children should have also the following not mandatory BM samples: at the beginning of reinduction therapy, at the beginning of maintenance therapy, during maintenance, at the time of treatment discontinuation, and 1 year after the end of therapy. Mononuclear cells were isolated by Ficoll-Paque gradient centrifugation, and their DNA was extracted and purified through the use of a Puregene DNA Purification Kit (Gentra Systems, Minneapolis, MN) according to the manufacturer's instructions.

Identification of PCR targets and design of allele-specific oligonucleotides. Genomic DNA samples obtained at diagnosis and at relapse were screened for clonal immunoglobulin (IgH/K) and T-cell receptor (TCRγ/δ) gene rearrangements, using published primer sets[26-29]. As an extra target for T-ALL patients, primer sets specific for the SIL-

TAL1/2 deletion were included in the PCR screening protocol using breakpoint-spanning SIL gene forward and TAL reverse oligonucleotides[26,29]. If no T-ALL target was detected, patients were also screened for clonal TCRβ rearrangements[30].

Clonal immune gene rearrangements were identified by heteroduplex analysis (Pharmacia, Uppsala, Sweden) and sequenced by the BigDye Terminator Cycle Sequencing Kit using the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA)[31]. After sequencing, allele-specific oligonucleotides were designed for each PCR target on the basis of the sequence data of the junctional region using Primer Express software v3.0 (Applied Biosystems).

Evaluation and interpretation of MRD RQ-PCR results. The designed forward allele-specific oligonucleotides were then tested in combination with germ-line reverse primers and germ-line TaqMan probes by RQ-PCR using the 7700 or 7900HT Sequence Detection System (Applied Biosystems)[32-36]. For MRD analysis, only primers that reached a sensitivity of at least 10^{-4} were used. PCR analysis was then performed, and results were interpreted according to the guidelines developed by the European Study Group on MRD detection in ALL[37]. If MRD levels differed between the 2 markers, the higher MRD level was assumed to be more accurate.

MRD results were categorized as negative (MRD Neg), positive below quantitative range ($<10^{-4}$) (Pos NQ), or positive within quantitative range ($\geq 10^{-4}$) (Pos Q). An MRD result was considered to be negative when there was an absence of specific amplification or amplification within 1 threshold cycle of the background.[37]

Study design and statistical analysis

As mentioned before, first, we prospectively studied a cohort of 113 patients treated in two AIEOP centers (Padua and Turin, Italy). Cumulative incidence curves for relapse were estimated adjusting for competing risks of other events and were compared with Gray's test. Death in CCR and a diagnosis of second neoplasms were the competing events[38]. All *P* values reported are two-sided, and $P < .05$ was considered to be significant. A patient was considered MRD Neg if MRD continued to be negative at all time points evaluated, or Pos NQ or Pos Q if positivity below or within the quantitative range, respectively, was observed at least once during monitoring[13]. Patients' characteristics are summarized in Table 1a.

In a second phase, we retrospectively analyzed a cohort of 47 relapsed patients who were treated in other AIEOP centers and enrolled in the AIEOP REC 2003 trial[13]. We

designed a blinded, matched case-control study, in which the cases were patients in relapse and the controls were patients in CCR. For each case, we tried to find one control, among patients from the above cohort, who was matched for sex, ethnic origin, age at diagnosis, immunophenotype, and final class of risk (standard, intermediate, or high) of the AIEOP-BFM ALL 2000 protocol. We were able to select 74 patients (37 cases and 37 matched controls) with either T-ALL phenotype (n = 10) or B-ALL phenotype (n = 64) and aged more than 1 year. In two cases, it was not possible to match the controls for sex. Median age at diagnosis was 5.1 and 4.8 years for cases and controls, respectively. Median time from diagnosis to relapse was 23.8 months in the cases (patients in relapse); median follow-up time of controls (patients in CCR) was 65.1 months. Student's t test was used to compare mean ages between cases and controls. We used the Fisher exact test to determine the significance of association between two binary variables—risk factor (presence of a Pos Q result during follow-up) and outcome (relapse)—in a 2x2 contingency table and for estimation of the corresponding common odds ratio. The odds ratio was used as an accurate approximation of the relative risk in this retrospective case-control study. Matching criteria are listed in Table 2a.

Finally, we considered all relapsed patients who were analyzed in this study: 21 from the initial cohort treated in the AIEOP centers in Padua and Turin and 47 treated in other AIEOP centers. Characteristics of these latter relapsed patients' are summarized in Table 3. The GraphPad Prism 4 program (GraphPad Software, La Jolla, CA) was used to analyze the data.

Patients' MRD monitoring and samples

All relapsed patients had been monitored for MRD during the time between diagnosis and relapse (Table 1a and Table 3). Because of the reported risk of clonal evolution[39-43], samples from relapsed patients were evaluated at both diagnosis and relapse, and conserved immune gene rearrangements were considered for the analysis of the data.

Median follow-up time of patients in CCR was 67.2 months (range, 43.9-99.5), and median time between the last time point evaluated for MRD and the end of the study was 38.5 months (range, 12.3-74). We analyzed a total of 943 samples, with a median number of 6 samples per patient. The molecular markers assayed are listed in Supplementary Table 1.

Results

Evaluation and impact of MRD on hematologic relapse

As reported above, we studied a cohort of consecutive 113 patients treated in two representative AIEOP centers participating in the 2000 study[25], 21 of whom relapsed during or after discontinuation of treatment.

Of the 92 patients who remained in CCR, 73 (79%) continued to be MRD Neg, 18 had at least one Pos NQ result during the monitoring, and 1 patient had a Pos Q result that became negative during therapy. Of the 21 relapsed patients, 13 (62%) had one or more Pos NQ (n = 7) or Pos Q (n = 6) results during follow-up, and 8 (38%) continued to be MRD Neg at all time points monitored.

Of the 113 patients included in the study, 7 had one or more Pos Q results during follow-up, and 6 of the 7 subsequently relapsed. Patients with at least one Pos Q result during follow-up had a much higher cumulative incidence of relapse (85.7%) than children with either MRD always Neg (11.1%) or with at least one POS NQ (26.5%) result ($P < .001$) (Figure 1).

Among the 24 patients who had one or more Pos NQ results, 6 (20%) relapsed and 18 (75%) are in CCR (Figure 1). In 9 of the 18 patients in CCR, a Pos NQ result evidenced at one time point was not confirmed at subsequent time points' evaluations. The other 9 patients in CCR had a Pos NQ result at the end of therapy and/or 1 year later. For 4 of these patients, more than 3 years have elapsed since the last time point monitored; for the remaining 5, less than 1 year has elapsed.

Of 82 patients who were MRD Neg during the entire follow-up, only 9 (11%) relapsed (Figure 1). Among these 9 patients, 7 had been last evaluated more than 3 months before relapse, and 2 suffered an extramedullary relapse.

MRD data in this cohort of patients are shown in Table 1b.

Case-control study

The blinded, matched case-control study was aimed at investigating the implication of a Pos Q result in patients who were in complete morphological remission.

In the relapsed and control group the number of patients who had at least one Pos Q result was 11 and 1, respectively. Thus, the occurrence of Pos Q results during MRD monitoring was significantly associated with relapse ($P < .001$; odds ratio, 15.23; 95% confidence interval [CI], 1.85 to 125.5, see also Table 2b).

Of the 37 controls included in this second part of the study, 32 had MRD Neg results at all time points monitored, and 4 had one or more Pos NQ results during follow-up. In 3 of the 4 controls with Pos NQ results, MRD became negative during treatment. The only 1 patient who had a Pos Q result became Pos NQ continuing treatment.

Of the 37 relapsed cases in this study, 11 had Pos Q, 13 had Pos NQ, and 13 had MRD Neg results during the monitoring. Eleven relapsed patients who were MRD Neg at all time points had been last evaluated more than 3 months before relapse, while one suffered an extramedullary relapse.

MRD evaluation in relapsed patients

Considering all 68 relapsed patients (21 from the initial part and 47 from the second part of the study) monitored in this study—independently of the time between the last evaluation and relapse and of the site of relapse—49 had one or more MRD positive results during follow-up: 26 Pos Q and 23 Pos NQ.

Only 3 of the 26 patients who had a Pos Q result during follow-up became Pos NQ, and none became MRD Neg. In 23 of the 68 relapsed ALL patients included in the study, MRD Pos Q was detected at the last evaluation before the conventional diagnosis of morphological relapse. Of 23 patients who had a Pos NQ result at the beginning of the reinduction and/or maintenance therapy, 12 subsequently became MRD Neg.

Thirty-one relapsed patients had an MRD Neg result in the last evaluation before relapse, and 14 had a Pos NQ result. In these patients no Pos Q result was detected before relapse, either because of the extramedullary site of the relapse or because of the time that had elapsed since the last evaluation. Indeed, 3 patients suffered an extramedullary relapse: we found a Pos NQ in 2 cases and a value of 1×10^{-4} in 1 case. One patient suffered a combined relapse with a value of 1×10^{-2}. In 36 cases the last evaluation had been performed more than 3 months before relapse (>6 months in 8 cases and >1 year in 10 cases). In only 5 patients the medullary relapse occurred within 3 months from last evidence of negativity at MRD evaluation.

Excluding those relapsed patients without a detectable Pos Q result before relapse who suffered an extramedullary or combined relapse (4/68) and who were evaluated >3 months before relapse (36/68), the recurrence of leukemia could have been predicted using MRD data in 23 of 28 patients (82%). The median time between the first finding of Pos NQ (that was not followed by an MRD Neg result) and the morphologic relapse was 6.1 months

(range, 0.2-26 months), and the median time between the first finding of Pos Q and relapse was 4.8 months (range, 0.2-22 months).

MRD data for all relapsed patients studied are detailed in Figure 2.

Discussion

In a two-step study on pediatric patients with ALL we investigated the prognostic significance of MRD as a predictor of subsequent recurrence of hematologic disease. While remission in children with ALL is evaluated using both morphological and molecular/immunologic parameters[44], leukemia relapse is still defined exclusively by clinical and morphologic criteria, which have low sensitivity. Morphological examination may be associated to the risk of under- or overestimating the real load of leukemic cells due to the difficulty to discriminate between blasts and normal lympho-hematopoietic cells. Moreover, its sensitivity is hampered by the fact that it does not rely on specific leukemia markers as MRD monitoring does. The persistence of MRD positivity during or after the end of therapy has been shown to be associated with a long-term dismal outcome[4-5,8,45-49]. However, to date, few studies have investigated the role of MRD as a powerful tool to anticipate and define a leukemic relapse of ALL.

To our knowledge, the present study is the first to have evaluated the significance of MRD monitoring to predict leukemic relapse in pediatric ALL patients belonging to different risk classes and treated according to the same therapeutic protocol. MRD monitoring during or after the discontinuation of treatment was able to identify patients whose disease relapsed morphologically after the detection of a Pos Q result during follow-up. In detail, of the 7 patients included in the initial cohort of 113 patients treated at two AIEOP centers who had Pos Q results during follow-up, 6 relapsed. The cumulative incidence of relapse in patients with Pos Q results was significantly higher (85.7%) than that of children with MRD results either negative or Pos NQ. Only one patient who had T-ALL and remains in CCR had a Pos Q MRD evaluation, which later became Pos NQ during follow-up. The clearance of the blast population differs according to the immune-phenotype of ALL, leukemia T-cells having a slower clearance than B-ALL cells. As a consequence, the frequency of MRD positivity during therapy, especially in the initial phases, is higher in patients with T-ALL than in those with B-ALL, and the level of MRD at the first time points obtained during treatment is higher in T-ALL than in B-ALL[50].

To confirm our initial findings, we performed a matched case-control study of 37 patients in relapse (cases) and 37 patients in CCR (controls). The analysis demonstrated that

finding a Pos Q result during monitoring is associated with a 15-fold increase (95% CI, 1.85 to 125.5) in the rate of relapse when compared to finding a negative or Pos NQ MRD (see Table 2b). This observation further confirms the strong predictivity of a Pos Q result as an indicator of impending morphologic relapse of the disease (positive predictive value, 92%).

Based on these findings, we believe that it is reasonable to propose that the detection of a Pos Q result during follow-up be considered a molecular relapse of ALL, highly predictive of a classical morphological recurrence. Such an event justifies the adoption of the choice of offering to the patient salvage therapies, including allo-HSCT, without waiting for the relapse to become morphologically evident.

It must be also underlined that with the current strategies no patients with a Pos Q result improved their MRD status by becoming MRD Neg at later time points: Thus, these patients seemed not to benefit from further administration of conventional front-line chemotherapy. This finding suggests that these patients could be good candidates for more intensive and aggressive therapeutic options, namely allo-HSCT. However, since it has been reported that patients with persistent MRD are more likely to relapse after allo-HSCT than those in molecular remission[12-13], transplantation should be preceded by a course of intensive therapy including novel active agents, such as clofarabine[51], aimed at decreasing the tumor burden. It is reasonable to speculate that salvage therapy in the case of molecular relapse could be more efficient than when used in patients already in morphological relapse. The therapeutic benefit deriving from this choice remains to be demonstrated in properly conducted studies.

It should be noted that not all patients included in this study who had MRD detectable at very low levels later relapsed. In 9 patients in CCR, the Pos NQ detected at one time point was not confirmed at subsequent time points during follow-up: This finding could be explained by the late effect of the chemotherapy or by a clearance of the leukemia cells exerted by the patient's immune system[22,48]. In contrast, in 9 other patients, MRD was detectable only at later time points (i.e., at the end of therapy or 1 year later). In 4 of them, more than 3 years have elapsed since the last monitored time point, and in 5, less than 1 year has elapsed. We cannot exclude that a longer follow-up might show a relapse in some of these patients. In fact, it is known that the cumulative risk of ALL is higher in the first 3 years after the end of therapy[52]. The finding of a Pos NQ result during follow-up suggests the opportunity to repeat the evaluation within 2-3 months in order to detect (i) either an increase in the disease level up to values within the quantitative range (Pos Q), which are highly predictive of a future overt relapse or (ii) a conversion to MRD negativity, which is

associated with a much lower risk of leukemia recurrence. In this regard, it is noteworthy that our study showed that a monitoring interval of 3 months, which was the same interval adopted in a study of the significance of molecular relapse in adult ALL[22], is sufficient to predict a molecular relapse of ALL with an accuracy greater than 80%. The development of an overt leukemia relapse in the 5 patients who had an MRD Neg evaluation in the 3 months preceding recurrence can be interpreted in view of the existence of situation in which the rate of leukemia re-growth is extra-ordinarily high and precluding the possibility of a timely prediction of recurrence. In conclusion, this study suggests the utility of using MRD data to predict and identify relapse in pediatric ALL patients. In particular, the finding of a Pos Q result during follow-up strongly predicts a subsequent hematologic relapse: the accuracy of prediction improves by monitoring every 3 months. In addition, it could be helpful to design and administer personalized salvage therapy before overt hematologic relapse, namely when the tumor burden is lower. Additional clinical studies are needed to prove whether therapeutic intervention based on MRD results can improve the final outcome of pediatric ALL patients. For this purpose, a multicenter clinical trial in which salvage therapy starts in the presence of quantifiable MRD (Pos Q) could be performed, as already done in adult patients in the GMALL trial[22].

References

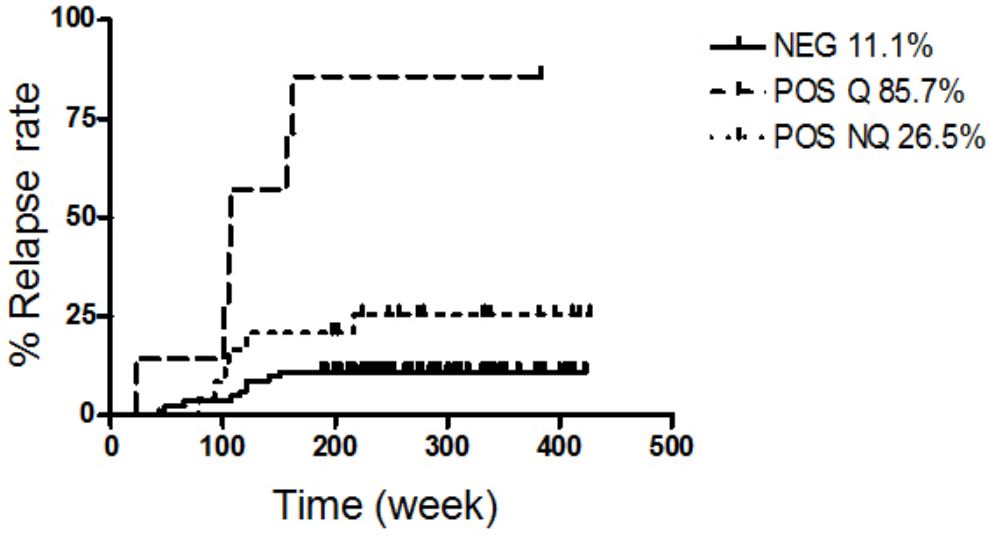
1. Pui, C.H., Robison, L.L. & Look, A.T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030-1043 (2008).
2. Gaynon, P.S. Childhood acute lymphoblastic leukaemia and relapse. *Br J Haematol* **131**, 579-587 (2005).
3. Einsiedel, H.G., *et al.* Long-term outcome in children with relapsed ALL by risk-stratified salvage therapy: results of trial acute lymphoblastic leukemia-relapse study of the Berlin-Frankfurt-Munster Group 87. *J Clin Oncol* **23**, 7942-7950 (2005).
4. Cave, H., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer--Childhood Leukemia Cooperative Group. *N Engl J Med* **339**, 591-598 (1998).
5. van Dongen, J.J., *et al.* Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet* **352**, 1731-1738 (1998).
6. Ciudad, J., *et al.* Prognostic value of immunophenotypic detection of minimal residual disease in acute lymphoblastic leukemia. *J Clin Oncol* **16**, 3774-3781 (1998).
7. Coustan-Smith, E., *et al.* Immunological detection of minimal residual disease in children with acute lymphoblastic leukaemia. *Lancet* **351**, 550-554 (1998).
8. Coustan-Smith, E., *et al.* Clinical importance of minimal residual disease in childhood acute lymphoblastic leukemia. *Blood* **96**, 2691-2696 (2000).

9. Coustan-Smith, E., *et al.* Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia. *Blood* **100**, 52-58 (2002).
10. Borowitz, M.J., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* **111**, 5477-5485 (2008).
11. Eckert, C., *et al.* Prognostic value of minimal residual disease in relapsed childhood acute lymphoblastic leukaemia. *Lancet* **358**, 1239-1241 (2001).
12. Coustan-Smith, E., *et al.* Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia after first relapse. *Leukemia* **18**, 499-504 (2004).
13. Paganin, M., *et al.* Minimal residual disease is an important predictive factor of outcome in children with relapsed 'high-risk' acute lymphoblastic leukemia. *Leukemia* **22**, 2193-2200 (2008).
14. Knechtli, C.J., *et al.* Minimal residual disease status before allogeneic bone marrow transplantation is an important determinant of successful outcome for children and adolescents with acute lymphoblastic leukemia. *Blood* **92**, 4072-4079 (1998).
15. van der Velden, V.H., *et al.* Real-time quantitative PCR for detection of minimal residual disease before allogeneic stem cell transplantation predicts outcome in children with acute lymphoblastic leukemia. *Leukemia* **15**, 1485-1487 (2001).
16. Bader, P., *et al.* Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. *Leukemia* **16**, 1668-1672 (2002).
17. Goulden, N., *et al.* Minimal residual disease prior to stem cell transplant for childhood acute lymphoblastic leukaemia. *Br J Haematol* **122**, 24-29 (2003).
18. Uzunel, M., Jaksch, M., Mattsson, J. & Ringden, O. Minimal residual disease detection after allogeneic stem cell transplantation is correlated to relapse in patients with acute lymphoblastic leukaemia. *Br J Haematol* **122**, 788-794 (2003).
19. Krejci, O., *et al.* Level of minimal residual disease prior to haematopoietic stem cell transplantation predicts prognosis in paediatric patients with acute lymphoblastic leukaemia: a report of the Pre-BMT MRD Study Group. *Bone Marrow Transplant* **32**, 849-851 (2003).
20. Bader, P., *et al.* Prognostic value of minimal residual disease quantification before allogeneic stem-cell transplantation in relapsed childhood acute lymphoblastic leukemia: the ALL-REZ BFM Study Group. *J Clin Oncol* **27**, 377-384 (2009).
21. Flohr, T., *et al.* Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* **22**, 771-782 (2008).
22. Raff, T., *et al.* Molecular relapse in adult standard-risk ALL patients detected by prospective MRD monitoring during and after maintenance treatment: data from the GMALL 06/99 and 07/03 trials. *Blood* **109**, 910-915 (2007).
23. Lo Coco, F., *et al.* Therapy of molecular relapse in acute promyelocytic leukemia. *Blood* **94**, 2225-2229 (1999).
24. Henze, G., *et al.* Six-year experience with a comprehensive approach to the treatment of recurrent childhood acute lymphoblastic leukemia (ALL-REZ BFM 85). A relapse study of the BFM group. *Blood* **78**, 1166-1172 (1991).
25. Conter, V., *et al.* Molecular response to treatment redefines all prognostic factors in children and adolescents with B-cell precursor acute lymphoblastic leukemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. *Blood* **115**, 3206-3214 (2010).

26. Pongers-Willemse, M.J., *et al.* Primers and protocols for standardized detection of minimal residual disease in acute lymphoblastic leukemia using immunoglobulin and T cell receptor gene rearrangements and TAL1 deletions as PCR targets: report of the BIOMED-1 CONCERTED ACTION: investigation of minimal residual disease in acute leukemia. *Leukemia* **13**, 110-118 (1999).
27. Szczepanski, T., *et al.* Ig heavy chain gene rearrangements in T-cell acute lymphoblastic leukemia exhibit predominant DH6-19 and DH7-27 gene usage, can result in complete V-D-J rearrangements, and are rare in T-cell receptor alpha beta lineage. *Blood* **93**, 4079-4085 (1999).
28. Szczepanski, T., *et al.* Precursor-B-ALL with D(H)-J(H) gene rearrangements have an immature immunogenotype with a high frequency of oligoclonality and hyperdiploidy of chromosome 14. *Leukemia* **15**, 1415-1423 (2001).
29. Breit, T.M., *et al.* tal-1 deletions in T-cell acute lymphoblastic leukemia as PCR target for detection of minimal residual disease. *Leukemia* **7**, 2004-2011 (1993).
30. van Dongen, J.J.M., *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* **17**, 2257-2317 (2003).
31. Germano, G., *et al.* Rapid detection of clonality in patients with acute lymphoblastic leukemia. *Haematologica* **86**, 382-385 (2001).
32. Verhagen, O.J., *et al.* Application of germline IGH probes in real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* **14**, 1426-1435 (2000).
33. Bruggemann, M., *et al.* Improved assessment of minimal residual disease in B cell malignancies using fluorogenic consensus probes for real-time quantitative PCR. *Leukemia* **14**, 1419-1425 (2000).
34. van der Velden, V.H., *et al.* Immunoglobulin kappa deleting element rearrangements in precursor-B acute lymphoblastic leukemia are stable targets for detection of minimal residual disease by real-time quantitative PCR. *Leukemia* **16**, 928-936 (2002).
35. van der Velden, V.H.J., Wijkhuijs, J.M., Jacobs, D.C.H., van Wering, E.R. & van Dongen, J.J.M. T cell receptor gamma gene rearrangements as targets for detection of minimal residual disease in acute lymphoblastic leukemia by real-time quantitative PCR analysis. *Leukemia* **16**, 1372-1380 (2002).
36. van der Velden, V.H.J., *et al.* Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* **17**, 1013-1034 (2003).
37. van der Velden, V.H.J., *et al.* Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* **21**, 604-611 (2007).
38. Gray, R. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* **16**, 1141-1154 (1988).
39. Szczepanski, T., *et al.* Comparative analysis of Ig and TCR gene rearrangements at diagnosis and at relapse of childhood precursor-B-ALL provides improved strategies for selection of stable PCR targets for monitoring of minimal residual disease. *Blood* **99**, 2315-2323 (2002).
40. Szczepanski, T., *et al.* Comparative analysis of T-cell receptor gene rearrangements at diagnosis and relapse of T-cell acute lymphoblastic leukemia (T-ALL) shows high stability of clonal markers for monitoring of minimal residual disease and reveals the occurrence of second T-ALL. *Leukemia* **17**, 2149-2156 (2003).

41. Germano, G., *et al.* Clonality profile in relapsed precursor-B-ALL children by GeneScan and sequencing analyses. Consequences on minimal residual disease monitoring. *Leukemia* **17**, 1573-1582 (2003).
42. Li, A., *et al.* Sequence analysis of clonal immunoglobulin and T-cell receptor gene rearrangements in children with acute lymphoblastic leukemia at diagnosis and at relapse: implications for pathogenesis and for the clinical utility of PCR-based methods of minimal residual disease detection. *Blood* **102**, 4520-4526 (2003).
43. Germano, G., *et al.* Comparative sequence analysis of incomplete DJH and TCR gene rearrangements in children with relapses of T-ALL. *Leukemia* **19**, 1687-1689 (2005).
44. Pui, C.H. & Campana, D. New definition of remission in childhood acute lymphoblastic leukemia. *Leukemia* **14**, 783-785 (2000).
45. Mortuza, F.Y., *et al.* Minimal residual disease tests provide an independent predictor of clinical outcome in adult acute lymphoblastic leukemia. *J Clin Oncol* **20**, 1094-1104 (2002).
46. Nizet, Y., *et al.* Long-term follow-up of residual disease in acute lymphoblastic leukemia patients in complete remission using clonogeneic IgH probes and the polymerase chain reaction. *Blood* **82**, 1618-1625 (1993).
47. Cave, H., *et al.* Prospective monitoring and quantitation of residual blasts in childhood acute lymphoblastic leukemia by polymerase chain reaction study of delta and gamma T-cell receptor genes. *Blood* **83**, 1892-1902 (1994).
48. Bjorklund, E., Mazur, J., Soderhall, S. & Porwit-MacDonald, A. Flow cytometric follow-up of minimal residual disease in bone marrow gives prognostic information in children with acute lymphoblastic leukemia. *Leukemia* **17**, 138-148 (2003).
49. Marshall, G.M., *et al.* Importance of minimal residual disease testing during the second year of therapy for children with acute lymphoblastic leukemia. *J Clin Oncol* **21**, 704-709 (2003).
50. Willemse, M.J., *et al.* Detection of minimal residual disease identifies differences in treatment response between T-ALL and precursor B-ALL. *Blood* **99**, 4386-4393 (2002).
51. Bonate, P.L., *et al.* Discovery and development of clofarabine: a nucleoside analogue for treating cancer. *Nat Rev Drug Discov* **5**, 855-863 (2006).
52. Pui, C.-H., *et al.* Risk of adverse events after completion of therapy for childhood acute lymphoblastic leukemia. *J Clin Oncol* **23**, 7936-7941 (2005).

Figures



Group	Patients	Relapses	Cumulative incidence of relapse (95% CI)	*P*
Neg	82	9	11.1 (6-20)	
Pos NQ	24	6	26.5 (13-50)	<.001
Pos Q	7	6	85.71 (54-99)	

Figure1. Cumulative incidence of relapse

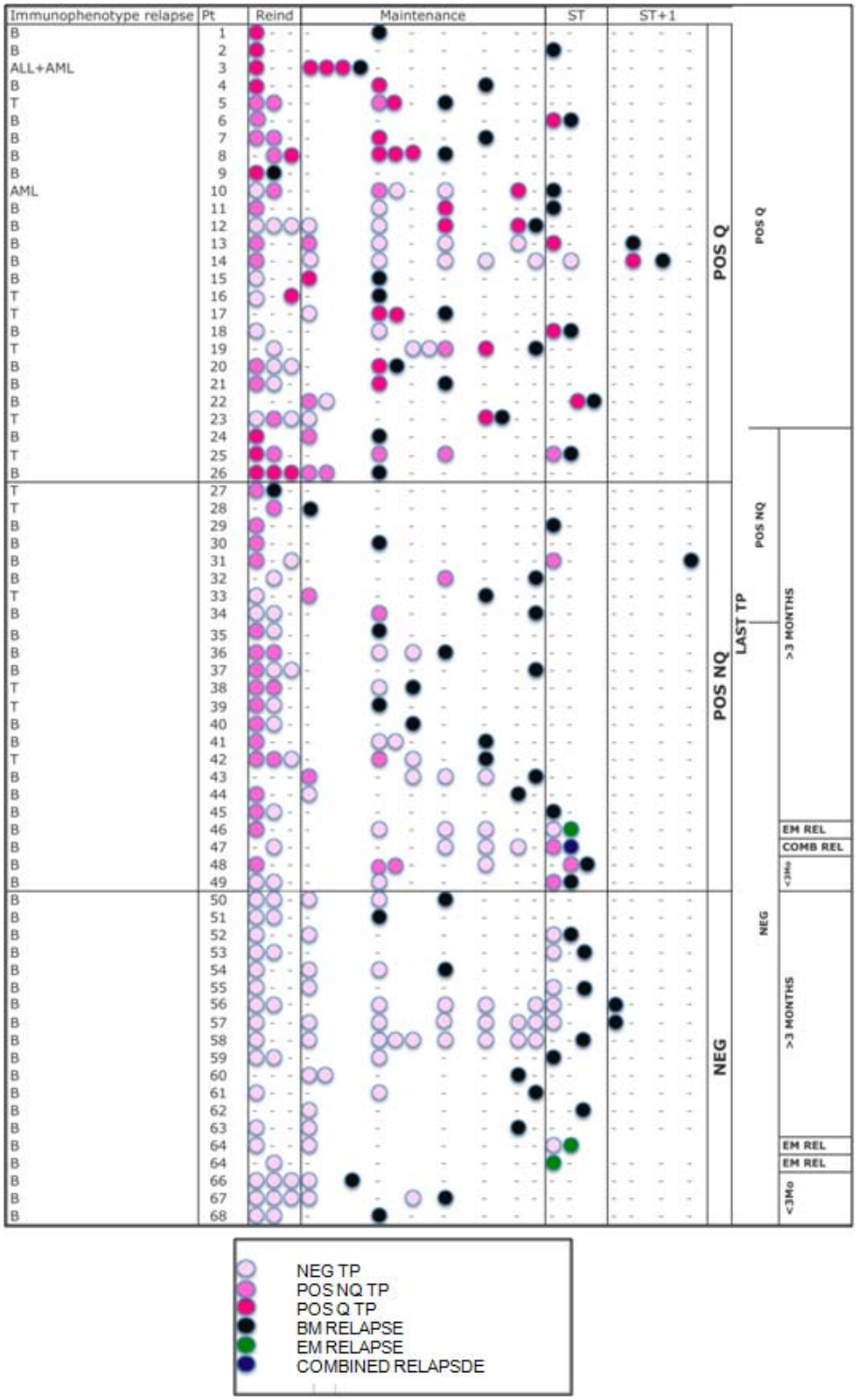


Figure 2. Schematic view of MRD analysis in relapsed patients

Pt indicates patient; Reind reinduction therapy; ST, the end of therapy; ST=1, 1 year after the end of therapy; TP, time point monitored; EM, extramedullary site of relapse; AML, acute myeloid leukemia.

Tables

Table 1. Characteristics of patients

Characteristic	Patients in CCR (n = 92)	Relapsed patients (n = 21)
Median age at diagnosis, y (range)	5 (1.1-20.6*)	4.5 (1.1-16.1)
ALL lineage		
B	85	19
T	7	1
AML	0	1
Sex		
Male	45	12
Female	47	9
Median follow-up: diagnosis to end of study or to relapse, mo (range)	67.2 (43.9-99.5)	21.7 (5.2-50.5)
AIEOP ALL 2000 stratification		
MRD SR	39	3
MRD MR	52	15
MRD HR	1	3
PPR	2 MRD MR pt	1 MRD MR, 1 MRD HR pt
t(4;11)	1 MRD MR pt	1 MRD MR pt
Final class of risk		
SR	39	3
MR	49	13
HR	4	5
REC 2003 stratification		
S2		9 (2 SR, 7MR)
S3		8 (1 SR, 5 MR, 1 HR)
S4		4 (1 MR, 3 HR)

b.

MRD evaluation	MRD Pos Q	MRD Neg or Pos NQ	Total
a) series of relapsing patients	6	15	21
b) series of CCR patients	1	91	92

AML indicates acute myeloid leukemia; SR, standard risk; MR, medium risk; HR, high risk; PPR, poor prednisone responder; S2, class of medium risk of AIEOP REC 2003 protocol; S3 and S4, classes of high risk of AIEOP REC 2003 protocol.
*1 patient >18 y treated according to the AIEOP ALL 2000 protocol.

Table 2. Matching criteria for case-control study and results

a.

Characteristic
Age, y: >1 and <5, ?5 and <10, ?10 and <18
Sex: male or female
Immunophenotype: B- or T-ALL
AIEOP ALL 2000
Prednisone response: good (GPR) or poor (PPR)
High-risk chromosomal translocations: t(9;22), t(4;11)
MRD, day +33 and +78: SR, MR, HR

b.

MRD evaluation	MRD Pos Q	Pos NQ	MRD Neg	TOTAL
a) series of relapsing patients	11	13	13	37
b) series of CCR patients	1	4	32	37

Fisher exact test; odds ratio, 15.23 (95% CI, 1.85-125.5); *P* .003.

Table 3. Characteristics of all relapsed patients

Characteristic	Relapsed patients (n = 47)
Median age at diagnosis, y (range)	6.7 (1.2-15.5)
ALL lineage	
B	34
T	12
ALL+AML	1
Sex	
Male	29
Female	18
Median follow-up: diagnosis to end of study or to relapse, mo (range)	17.6 (6.4-30.5)
AIEOP ALL 2000 stratification	
MRD SR	6
MRD MR	27
MRD HR	14
PPR	6 MRD MR pt, 7 MRD HR pt
t(9;22)	2
Final class of risk	
SR	6
MR	20
HR	21
not known	1
REC 2003 stratification	
S3	19
S4	28

Supplementary Table 1. Characteristics of samples and assays used

MRD evaluation	No. of samples
Molecular markers	
IgH, complete	93
IgH, incomplete	11
KDE	9
TCRδ, complete	27
TCRδ, incomplete	5
TCRγ	11
SIL-TAL	2
TCRβ, complete	2
Sensitivity	
5×10^{-4}	2
1×10^{-4}	61
1×10^{-5}	97
Quantitative range	
1×10^{-3}	54
5×10^{-4}	40
1×10^{-4}	63
1×10^{-5}	3
Total no. of samples	943
Median no. of samples/patient (range)	6 (2-11)

Chapter 2: Genomic characterization of two B cell lymphomas, the Diffuse Large B cell Lymphoma and the Chronic Lymphocytic Leukemia

Diffuse Large B cell Lymphoma (DLBCL) and Chronic Lymphocytic Leukemia (CLL) are two frequent types of B-cell derived lymphomas.

DLBCL is the most common and aggressive subtype of B-Non-Hodgkin Lymphoma (NHL) of the adult, accounting for ~30,000 new diagnoses each year in the U.S. and representing nearly ~40% of all NHLs[1]. This type of lymphoma shows a highly heterogeneous response to therapy, with ~30% of cases not cured by currently available therapeutic regimens.

The WHO classification identifies separate entities of DLBCL: i) primary mediastinal large B-cell lymphoma; ii) intravascular large B-cell lymphoma; iii) primary effusion lymphoma (PEL) and iv) DLBCL, not fitting into one of the previous listed entities[2].

The variability in response of DLBCL to the current available therapeutic strategies reflects the marked heterogeneity in morphologic, phenotypic, clinical and molecular features of this disease that is still largely treated with the same therapeutic strategy (R-CHOP)[3-4].

Morphologically, DLBCL is a neoplasm of large B cells which diffusely disrupt the normal architecture of the lymph node or extranodal site. In 80% of cases DLBCL is composed of cells resembling germinal center centroblasts ("Centroblastic" type), while the "Immunoblastic" type (10% of the cases) has more than 90% immunoblasts[5].

Gene expression profiling can classify DLBCL into at least two phenotypic subtypes differing in molecular features as well as clinical presentation, reflecting the origin from distinct stages of B-cell differentiation, the germinal-center B cell-like (GCB) and the activated B-cell-like (ABC)[6-8].

GCB and ABC DLBCL show a largely distinct pattern of genetic abnormalities, suggesting that the underlying oncogenic pathways involved in the pathogenesis of these two entities are different.

Genetic lesions mainly associated with GCB-DLBCL include translocation t(14;18), amplification of *REL*, *PTEN* deletion, amplification of the miR-17-92 locus (which down-regulates *PTEN*)[8-10]. Somatic mutations associated with the GCB phenotype include gain-of-function lesions affecting *EZH2* and inactivating lesions targeting *CREBBP*[11-12]. Mutations affecting the tumor suppressor *TP53* are detected in 20% of DLBCL, and stratify patients

with GCB-DLBCL, but not ABC-DLBCL, into molecularly distinct subsets with different survivals[13].

Most ABC-DLBCL overexpress BCL2 and many amplify the *BCL2* locus. Frequent ABC-DLBCL associated lesions include deletion of the *INK4A-ARF* locus (encoding for the tumor suppressors *p16* and *p14*), trisomy 3 (encompassing *FOXP1*), gains of 3q (including *NFKBIZ)*, 18q21-q22 gains (associated with overexpression of *BCL2* and *NFATC1*) and 6q21-q22 losses[10,14]. The less curable ABC DLBCLs are also characterized by chromosomal translocations involving the master regulator of germinal center *BCL6* and somatic mutations inactivating *PRDM1*, which encodes for a transcriptional repressor required for plasma-cell differentiation[15-16]. Both these two lesions, which are mutually exclusive and promote lymphomagenesis by blocking terminal differentiation of germinal center B cells, lead to the development of lymphoproliferative disorders recapitulating critical features of the human DLBCL[16-18]. The majority of ABC-DLBCL show NF-κB activated phenotype[19] and ABC-DLBCL derived cells require NF-κB for both proliferation and survival[20-21]. Genetic lesions affecting multiple components of this pathway have been identified. The most frequent are represented by oncogenic activating mutations affecting *CARD11* (10%) and *MYD88* (30%), biallelic genetic lesions inactivating the negative NF-κB regulator *TNFAIP3* (30%), and somatic mutations affecting the BCR-associated signaling molecules *CD79B* (Igβ) (20%) and, less frequently, *CD79*A (Ig)[19,22-25].

A set of genetic lesions specific to DLBCL is represented by mutations showing features of somatic hypermutation (SHM). In DLBCL aberrant somatic hypermutation (ASHM) affects in more than 50% of the cases well know oncogenes, including *MYC*, *RHOH*, *PAX5* and *PIM*1[26]. Because the mutations affect both regulatory and coding sequences, ASHM may represent a major contributor to DLBCL pathogenesis.

CLL, a CD5+ B-cell malignancy, is the most common adult leukemia in the Western World. CLL is generally preceded by a condition, termed Monoclonal B cell Lymphocytosis (MBL), in which a clonal B-cell population, resembling immunophenotypically and cytogenetically a CLL tumor cell, is detected in the peripheral blood of otherwise healthy elderly individuals[27-28]. However, only about 1% of people with MBL develop CLL per year. In 10-15% of the cases CLL can progress into an aggressive form of clonally related lymphoma, most commonly of the diffuse large B cell type[29]. This condition, termed Richter Syndrome, is frequently associated with the acquisition of *TP53* somatic mutations.

Morphologically CLL is characterized by the presence of small mature-appearing B lymphocytes in the peripheral blood.

Despite the major advances of the last decade in understanding the pathogenesis and in improving the treatment[30] of CLL, this disease remains still incurable.

CLL has an extremely heterogeneous clinical course and it can be subdivided into two main subgroups with somatically mutated or unmutated IgV genes, which are associated with good and poor prognosis, respectively[31-32].

Despite the IgV mutational status, gene expression profile studies demonstrated that CLL shows a largely homogeneous phenotype, suggesting a common cell of origin. In addition, these analyses showed that the gene expression profile of CLL is related to that of antigen-experienced B cells[33-34].

More than 80% of CLL patients carry recurrent chromosomal abnormalities, which are independent predicting factors of survival[35].

The most frequent genetic lesion in CLL is so far represented by the deletion of the chromosomal region 13q14, which occurs in more than 50% of the patients. The minimal deleted region (MDR) contains *DLEU2*, a long non-coding RNA gene, the first exon of *DLEU1,* another gene coding for a sterile transcript, and the miR locus *15a/16-1*[36]. Mice carrying a deletion encompassing the entire MDR or only lacking the microRNAs *15a/16-1* develop a variety of lymphoproliferative disorders including MBL, CLL and, in a small fraction of cases, DLBCL[37].

The 11q deletion, which occurs in 20% of the patients, is associated with a poor prognosis. This lesion almost invariably encompasses the gene *ATM*, which is also somatically mutated in a small fraction of cases (12%; 1/3 of the ones carrying the deletion)[35,38].

Deletion of 17p is found in about 10% of CLL patients and always includes *TP53*, which is mutated in 80% of the cases carrying the deletion, while among cases without 17p deletion *TP53* mutations are much rarer. *TP53* mutations are associated with poor prognosis regardless of the presence of 17p deletion[39].

The pathogenetic role and the target genes of trisomy 12, which occurs in up to 20% of the CLL patients, are currently unknown.

Compared to other B-NHL balanced chromosomal translocations are virtually absent in CLL[35,40], suggesting that the genetic remodeling processes which occur in the germinal center do not play a role in this disease.

The known DLBCL- and CLL-associated genetic lesions do not account for all cases, and therefore they most likely represent only a subset of those required for tumor initiation. A further large-scale characterization of mutations and copy number aberrations present in these diseases is required not only to provide insights into their molecular pathogenesis, but also to identify new therapeutic targets[41].

Towards this purpose genome-wide technologies such as next generation whole-exome/genome sequencing and genome-wide high-density single nucleotide polymorphism (SNP) array have been recently applied to fully characterize a wide variety of cancer types in an unbiased fashion[42-55].

Cancer is a genetic disease, and, with the exception of rare cases in which tens of hundreds of genomic rearrangements occur in a single step (phenomenon recently described in Stephens PJ [56]), the mutations found in a cancer genome are progressively acquired over the lifetime of each patient. The somatic mutations in a cancer cell genome may encompass several distinct classes of DNA sequence changes, including single base substitutions, insertions or deletions of small or large segments of DNA, DNA rearrangements, copy number losses and gains. Few mutations, defined as "drivers", provide a selective advantage to the tumor cell and are positively selected during tumor progression. Most of the mutations, instead, are likely to be "passengers" with no functional effect on tumor cell growth. These mutations could have already been present in the cell that gave rise to the cancer or could have been acquired during subsequent tumor growth. The ultimate goal and the greatest challenge of cancer genomics is to distinguish the benign, pre-existing "passenger" mutations from the "drivers" that are relevant for pathogenesis. Mutations that most likely play a role in tumorigenesis are recurrent non-silent events that affect conserved domains or individual nucleotide spots of expressed genes. However, also finding recurring events in non-genic regions of the genome that are minimally annotated may be extremely helpful in identifying functional regions that are currently of unknown significance. The final readout for the role of identified mutations is represented by functional studies in tumor cell lines or in model organisms.

In the studies herein described we took advantage of a combined approach, integrating the whole exome sequencing and the SNP array analysis of a panel of 7 DLBCL and 5 CLL paired tumor/normal DNAs, to identify their respective non-silent mutations (including single base substitutions and small insertions and deletions) and gross copy number aberrations.

The whole exome sequencing was performed using a selective exomic capture approach (Sequence Capture Human Exome 2.1M Array, NimbleGen)[57] followed by next generation pyrosequencing (Roche FLX454 Sequencer)[58]. The schematic view of the microarray direct genomic selection and deep sequencing approaches is shown in Figure 1.

The SNP array analysis was performed using the Affymetrix Genome-Wide Human SNP Array 6.0[59], which interrogates 1.8 million genetic markers, including single nucleotide polymorphisms (SNPs) and copy number variations (CNVs). The experimental pipeline of the genotyping assay is shown in Figure 2.

References

1. Armitage, J.O. & Weisenburger, D.D. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. *J Clin Oncol* **16**, 2780-2795 (1998).
2. Swerdlow, S.H. WHO Classification of Tumours of Haematopoetic and Lymphoid Tissues. *International Agency for research on Cancer (IARC), Lyon* (2008).
3. Coiffier, B., *et al.* CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med* **346**, 235-242 (2002).
4. Coiffier, B. Rituximab in combination with CHOP improves survival in elderly patients with aggressive non-Hodgkin's lymphoma. *Semin Oncol* **29**, 18-22 (2002).
5. Engelhard, M., *et al.* Subclassification of diffuse large B-cell lymphomas according to the Kiel classification: distinction of centroblastic and immunoblastic lymphomas is a significant prognostic risk factor. *Blood* **89**, 2291-2297 (1997).
6. Alizadeh, A.A., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511 (2000).
7. Rosenwald, A., *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* **346**, 1937-1947 (2002).
8. Lenz, G., *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A* **105**, 13520-13525 (2008).
9. Bea, S., *et al.* Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* **106**, 3183-3190 (2005).
10. Tagawa, H., *et al.* Comparison of genome profiles for identification of distinct subgroups of diffuse large B-cell lymphoma. *Blood* **106**, 1770-1777 (2005).
11. Morin, R.D., *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-185 (2010).
12. Pasqualucci, L. Inactivating mutations of acetyltransferase genes in B cell Lymphoma *Nature* **in press** (2011).
13. Young, K.H., *et al.* Structural profiles of TP53 gene mutations predict clinical outcome in diffuse large B-cell lymphoma: an international collaborative study. *Blood* **112**, 3088-3098 (2008).

14. Rosenwald, A., *et al.* Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med* **198**, 851-862 (2003).
15. Pasqualucci, L., *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med* **203**, 311-317 (2006).
16. Mandelbaum, J., *et al.* BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell* **18**, 568-579 (2010).
17. Cattoretti, G., *et al.* Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer Cell* **7**, 445-455 (2005).
18. Calado, D.P., *et al.* Constitutive canonical NF-kappaB activation cooperates with disruption of BLIMP1 in the pathogenesis of activated B cell-like diffuse large cell lymphoma. *Cancer Cell* **18**, 580-589 (2010).
19. Compagno, M., *et al.* Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-721 (2009).
20. Davis, R.E., Brown, K.D., Siebenlist, U. & Staudt, L.M. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* **194**, 1861-1874 (2001).
21. Ngo, V.N., *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106-110 (2006).
22. Kato, M., *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712-716 (2009).
23. Lenz, G., *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676-1679 (2008).
24. Davis, R.E., *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88-92 (2010).
25. Ngo, V.N., *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* (2010).
26. Pasqualucci, L., *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341-346. (2001).
27. Landgren, O., *et al.* B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med* **360**, 659-667 (2009).
28. Rawstron, A.C., *et al.* Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med* **359**, 575-583 (2008).
29. Rossi, D. & Gaidano, G. Richter syndrome: molecular insights and clinical perspectives. *Hematol Oncol* **27**, 1-10 (2009).
30. Hallek, M., *et al.* Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164-1174 (2010).
31. Damle, R.N., *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840-1847 (1999).
32. Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G. & Stevenson, F.K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848-1854 (1999).
33. Klein, U., *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* **194**, 1625-1638 (2001).
34. Rosenwald, A., *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* **194**, 1639-1647 (2001).

35. Dohner, H., *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* **343**, 1910-1916 (2000).
36. Migliazza, A., *et al.* Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood* **97**, 2098-2104. (2001).
37. Klein, U., *et al.* The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell* **17**, 28-40 (2010).
38. Austen, B., *et al.* Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. *J Clin Oncol* **25**, 5448-5457 (2007).
39. Zenz, T., *et al.* TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol* **28**, 4473-4479 (2010).
40. Dohner, H., Stilgenbauer, S., Dohner, K., Bentz, M. & Lichter, P. Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. *J Mol Med* **77**, 266-281 (1999).
41. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719-724 (2009).
42. Ley, T.J., *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72 (2008).
43. Campbell, P.J., *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729 (2008).
44. Stephens, P.J., *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005-1010 (2009).
45. Pleasance, E.D., *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196.
46. Pleasance, E.D., *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190.
47. Mardis, E.R., *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066 (2009).
48. Shah, S.P., *et al.* Mutation of FOXL2 in granulosa-cell tumors of the ovary. *N Engl J Med* **360**, 2719-2729 (2009).
49. Ding, L., *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999-1005 (2010).
50. Beroukhim, R., *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905 (2010).
51. Wiegand, K.C., *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* **363**, 1532-1543 (2010).
52. Jones, S., *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231 (2010).
53. Ley, T.J., *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433 (2010).
54. Jiao, Y., *et al.* DAXX/ATRX, MEN1, and mTOR Pathway Genes Are Frequently Altered in Pancreatic Neuroendocrine Tumors. *Science* (2011).
55. Varela, I., *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* (2011).
56. Stephens, P.J., *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144**, 27-40 (2011).

57. Albert, T.J., *et al.* Direct selection of human genomic loci by microarray hybridization. *Nature methods* **4**, 903-905 (2007).
58. Margulies, M., *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
59. Matsuzaki, H., *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature methods* **1**, 109-111 (2004).

Figures

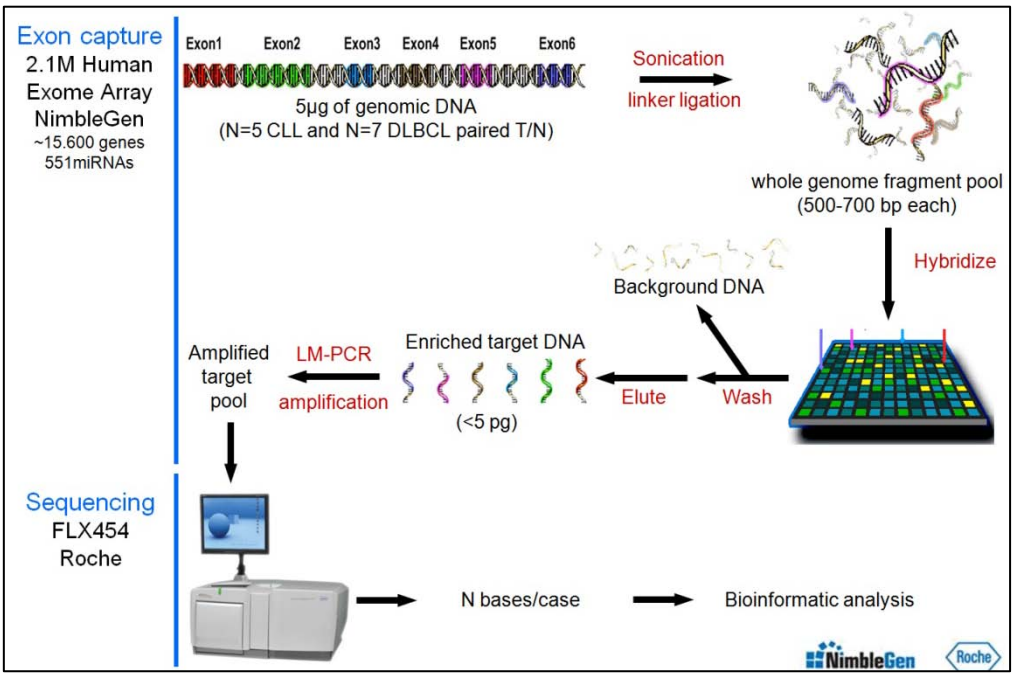


Figure 1. Schematic of the microarray direct genomic selection and deep sequencing approaches

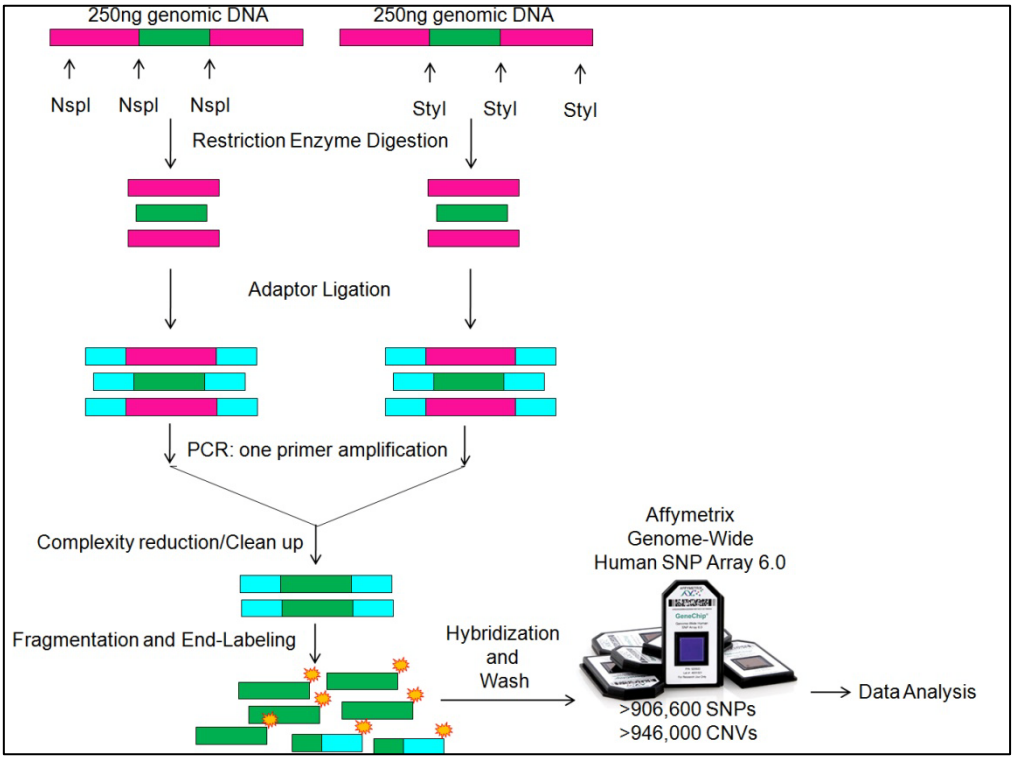


Figure 2. Genotyping SNP 6.0 assay overview

2.1 Inactivating mutations of acetyltransferase genes in B-cell lymphoma (Nature, *in press*)

Laura Pasqualucci[1,2], David Dominguez-Sola[1], Annalisa Chiarenza[1], Giulia Fabbri[1], Adina Grunn[1], Vladimir Trifonov[3], Lawryn H. Kasper[4], Stephanie Lerach[4], Honjyan Tang[1], Jing Ma[5], Davide Rossi[6], Amy Chadburn[7], Vundavalli V. Murty[1,2], Charles G. Mullighan[8], Gianluca Gaidano[6], Raul Rabadan[3], Paul K. Brindle[4] and Riccardo Dalla-Favera[1,2,9].

[1]*Institute for Cancer Genetics and the Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA.*

[2]*Department of Pathology & Cell Biology, Columbia University, New York, NY 10032, USA.*

[3]*Department of Biomedical Informatics and the Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.*

[4]*Department of Biochemistry, St Jude Children's Research Hospital, Memphis, TN 38105, USA.*

[5]*The Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, TN 38105, USA.*

[6]*Division of Hematology, Department of Clinical and Experimental Medicine and IRCAD, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy.*

[7]*Department of Pathology, Northwestern University. Feinberg School of Medicine, Chicago, IL 60611, USA.*

[8]*Department of Pathology, St Jude Children's Research Hospital, Memphis, TN 38105, USA.*

[9]*Department of Genetics & Development, Columbia University, New York, NY 10032, USA.*

B-cell non-Hodgkin lymphoma (B-NHL) comprises biologically and clinically distinct diseases whose pathogenesis is associated with genetic lesions affecting oncogenes and tumor-suppressor genes. We report here that the two most common types, follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL), harbor frequent structural alterations inactivating *CREBBP* and, more rarely, *EP300*, two highly related histone and non-histone acetyltransferases (HATs) that act as transcriptional co-activators in multiple signaling pathways. Overall, ~39% of DLBCL and 41% of FL cases display genomic deletions and/or somatic mutations that remove or inactivate the HAT coding domain of these two genes. These lesions commonly affect one allele, suggesting that reduction in HAT dosage is important for lymphomagenesis. We demonstrate specific defects in acetylation-mediated inactivation of the BCL6 onco-protein and activation of the p53 tumor-suppressor. These results identify *CREBBP/EP300* mutations as a major pathogenetic mechanism shared by common forms of B-NHL, and have direct implications for the use of drugs targeting acetylation/deacetylation mechanisms.

Diffuse Large B-cell Lymphoma (DLBCL) represents the most common form of B-cell non-Hodgkin Lymphoma (B-NHL), accounting for ~30% of the de-novo diagnoses and also arising as a frequent clinical evolution of Follicular Lymphoma (FL)[1]. The molecular pathogenesis of DLBCL is associated with multiple genetic lesions that segregate in part with individual phenotypic subtypes, namely germinal-center B-cell-like (GCB) and activated B-cell-like (ABC) DLBCL, suggesting the involvement of distinct oncogenic pathways[2-8]. However, the full spectrum of lesions that contribute to malignant transformation remains unknown. Genome-wide efforts toward the identification and functional characterization of the entire set of structural alterations present in the DLBCL genome are required for a complete understanding of its pathogenesis[9].

Toward this end, we have integrated next generation whole-exome sequencing analysis of 7 DLBCL cases and genome-wide high-density single nucleotide polymorphism (SNP) array analysis of 72 DLBCL cases. This combined approach led to the identification of >450 loci that are affected by somatic point mutations and/or by recurrent, focal gene copy number (CN) aberrations. Among those that have been so far independently validated, the most commonly involved regions were those harboring the acetyltransferase genes *CREBBP*

(*CBP*) and *EP300 (p300)*. *CREBBP* encodes a highly conserved and ubiquitously expressed nuclear phosphoprotein that, together with the closely related protein EP300, belongs to the KAT3 family of histone/protein lysine acetyltransferases[10,11]. CREBBP and EP300 function as transcriptional coactivators for a large number of DNA-binding transcription factors involved in multiple signaling and developmental pathways, by modifying lysine residues on both histone and non-histone nuclear proteins[12,13]. CREBBP and EP300 enhance transcription by multiple mechanisms, including: i) targeted acetylation of chromatin[12,13]; ii) acetylation of transcriptional activators (e.g., the tumor suppressors p53 and GATA1)[14-17]; and iii) acetylation-mediated inactivation of transcriptional repressors (e.g. the DLBCL-associated oncogene BCL6)[18]. Additionally, both molecules were found to exhibit ubiquitin ligase activity[19,20]. Consistent with the involvement in critical cellular functions, homozygous null mice for either *Crebbp* or *Ep300* are early embryonic lethal[21,22], and the same is true for the compound *Crebbp/Ep300* double heterozygous mice[21,22]. Haploinsufficiency of *CREBBP* (and, rarely, *EP300*) is responsible for the Rubinstein-Taybi syndrome (RTS), an autosomal congenital disorder characterized by mental and growth retardation, skeletal abnormalities and predisposition to tumor development[23-25]. Somatic mutations of these two genes are exceedingly rare in epithelial cancers[26-28], and only three cases were reported to carry *EP300*, but not *CREBBP* mutations in hematologic malignancies[29,30]. Additionally, chromosomal translocations of *CREBBP* are associated with an infrequent type of acute myeloid leukemia (AML) and with therapy-related AML and myelodysplastic syndrome, although the precise consequences of these lesions have not been fully elucidated[31-33].

Monoallelic lesions of *CREBBP* in DLBCL

Following initial observations from whole-exome sequencing analysis of 7 DLBCL cases and paired normal DNAs, we performed targeted re-sequencing of the entire *CREBBP* coding exons in 134 DLBCL samples representative of major phenotypic subtypes. This analysis revealed a total of 34 sequence variants distributed in 30 samples and whose somatic origin was documented by analysis of paired normal DNAs (available in 8 cases)(Fig. 1a,b and Table S1). Of these variants, 17 (50%) were inactivating events, including nonsense mutations (n=9), frameshift insertions/deletions (n=7) and mutations at consensus splice donor/splice acceptor sites (n=1), which generate aberrant transcripts carrying premature stop codons. Based on their distribution along the CREBBP protein, these mutations are all predicted to cause the elimination or truncation of the HAT domain (Fig. 1a and Table S1). The remaining variants included 3 in-frame deletions and 14 missense mutations, primarily

within the HAT domain (Fig. 1a and Table S1), suggesting that they may be functionally important (see below). While *CREBBP* mutations were identified in both DLBCL phenotypic subtypes, their frequency was significantly higher in GCB-DLBCL, where they account for ~32% of the cases (n=21/65) as compared to 13% in ABC/non-GC DLBCL (n=9/69; $p<0.01$)(Fig. 1c).

High density SNP array analysis, available for 72 samples from the same panel, and FISH analysis revealed the presence of monoallelic deletions encompassing or internal to the *CREBBP* locus in 8 additional cases (5 GCB-DLBCL and 3 ABC/non-GC-DLBCL), and a homozygous deletion in one patient (Fig. 2a,b and Fig. S1). Notably, the loss of genetic material was smaller than 240Kb in 4 cases and, in two (2147 and 2043), only involved a limited subset of *CREBBP* exons, thereby identifying this gene as the specific target (Fig. 2a and Table S2). In two additional samples for which CN data were not available, direct sequencing analysis revealed a hemizygous missense mutation, reflecting either the loss of the second allele or copy neutral loss of heterozygosity (Table S1). When combining the sequencing data with the CN data, *CREBBP* mutations and deletions were found to be mutually exclusive in most samples, revealing a predominantly monoallelic distribution (n=33/39 cases) (Fig. 2c). Rare instances of biallelic genetic lesions include one homozygous loss, a missense mutation with loss of heterozygosity (2 primary biopsies), biallelic nucleotide substitutions (2 cases), and a frameshift deletion with missense mutation of the second allele in the OCI-Ly8 cell line. Thus, 29% of all DLBCL patients (n=39/134), corresponding to 41.5% of GCB- and 17% of ABC-DLBCL, harbor genomic alterations affecting the *CREBBP* gene (Fig. 2d).

Frequent mutations of *CREBBP* in FL

We next analyzed various types of mature B-NHL, including FL, Burkitt Lymphoma (BL), Marginal-Zone Lymphoma (MZL) and Chronic Lymphocytic Leukemia (CLL). Mutations analogous to those found in DLBCL were frequent in FL (32.6%, with 16 events distributed in 15/46 cases), but not in other lymphoma types, suggesting a specific role in the pathogenesis of these two diseases (Fig. S2). This analysis also revealed the existence of several mutational hotspots at specific codons within the HAT domain, including R1446, also mutated in B-cell acute lymphoblastic leukemia (B-ALL)[34], Y1503 and D1435; in addition, a 3bp in-frame deletion causing the loss of S1680 was observed in 5 cases, suggesting a functional role for this presently uncharacterized serine (Table S1). None of these recurrent changes were detected in paired normal DNA, thus excluding that they represent germline

polymorphisms. While CN data were not available for the same FL panel, array-CGH analysis performed on a distinct dataset showed deletions spanning the *CREBBP* locus in only 1/68 cases (not shown). Collectively, these findings identify somatic mutations of *CREBBP* as a common event in FL.

EP300 mutations in DLBCL and FL

Given the significant structural and functional similarities between *CREBBP* and *EP300*, we investigated whether this second member of the KAT3 acetyltransferase family is also targeted by structural alterations in B-NHL. Mutational analysis of the same panel identified 19 sequence variants leading to amino acid changes (n=11), in-frame deletions (n=2), and premature stop codons due to frameshift deletions, aberrant splicing or nonsense mutations (n=6) (Fig. S3a and Table S3). These lesions were found in 10% of DLBCL (n=13/134) and 8.7% of FL samples (n=4/46), but were virtually absent in other B-NHLs (Fig. S3b). Seven additional DLBCLs harbored monoallelic deletions spanning, although not limited to, the *EP300* locus (Fig. S3c, Table S4 and data not shown). Notably, structural alterations of *CREBBP* and *EP300* co-exist in only a minority of the affected cases (n=6/53 DLBCL and 0/19 FL) (Fig. S3d), suggesting that inactivation of these loci is at least in part functionally equivalent. Thus, overall ~39% of all DLBCL and at least 41% of FL cases (based on mutations only) display structural alterations of KAT3 family genes.

CREBBP and EP300 protein expression

In order to compare the CREBBP and EP300 protein levels in normal and transformed B-cells, and to investigate the expression of the retained normal allele in cases carrying monoallelic genomic alterations, we examined the expression pattern of these two proteins in reactive human tonsils and in 78 DLBCL primary cases, of which 49 harbored both genes in wild-type configuration, by immunofluorescence (IF) and immunohistochemical (IHC) analysis. Consistent with their reported ubiquitous expression, CREBBP and EP300 were detectable in naïve B-cells within the mantle zone and, at higher levels, in germinal center (GC) B-cells (Fig 3a). Notably, most of the monoallelically-deleted DLBCL cases were positive for the two proteins, although at reduced levels, demonstrating that the residual wild-type allele is expressed (see Fig.3b,c for representative examples). Furthermore, RT-PCR amplification and direct sequencing of the ten CREBBP-mutated cell lines invariably showed the presence of the wild-type allele (Fig. S4), while western blot analysis using antibodies directed against the N-terminal portion of CREBBP revealed the

expression of a full-length protein, corresponding to the wild-type allele, in most of the cell lines carrying truncating gene mutations, as well as in the monoallelically-deleted SUDHL5 cells (Fig. 3d, top panel). Slightly different findings were observed for EP300 where, in contrast with the primary biopsies, the presence of truncating mutations or deletions was associated with the absence of protein expression in four affected cell lines (Fig. 3d, second panel from top). With the exception of SUDHL2, both CREBBP and EP300 mRNAs were readily detected by northern blot analysis in all lines analyzed (Fig 3d, panel 4 and 5 from top).

Interestingly, a few additional lines were found to express very low to undetectable protein levels, despite the presence of mRNA and the absence of structural alterations in the corresponding gene (see CREBBP in BJAB and SUHDL2, or EP300 in SUDHL7 and SUDHL8) (Fig. 3d). Similarly, 6/78 (8%) DLBCL biopsies from patients with intact alleles appeared to lack expression of the two proteins, either simultaneously (3 cases) or individually (2 CREBBP+/EP300- and 1 CREBBP-/EP300- cases)(Fig. 3c,e). Thus, the fraction of DLBCL patients with defective CREBBP and/or EP300 may be higher than that determined based on genetic lesions alone, suggesting that alternative mechanisms of KAT3 gene family inactivation may play a role in this disease.

CREBBP missense mutants fail to acetylate BCL6 and p53

While the presence of gross gene deletions and the distribution of truncating mutations clearly predict a complete gene inactivation or the loss of multiple key functional domains, including the HAT, the functional consequences of the numerous *CREBBP* missense mutations required direct experimental analysis. Notably, of the 30 total events identified, 27 cluster within HAT coding exons, suggesting a selective pressure to alter the CREBBP enzymatic activity (Table S1). In particular, 19 mutations were located within a 68 amino acids stretch that is 96% identical to EP300 and includes the contact surface for coenzyme A (CoA)[35] (Fig. S5). We therefore examined the effect of these mutations on the ability of CREBBP to acetylate known physiologic substrates. We selected BCL6 and p53 because of their biological relevance for GC development and lymphomagenesis[4,36]. In fact, acetylation of the tumor suppressor p53 is indispensable for its transcriptional activity[14,15,37], while EP300-mediated acetylation of the proto-oncoprotein BCL6 leads to inactivation of its transcriptional repressor function[18].

Transient transfection/co-immunoprecipitation assays confirmed that, analogous to EP300[18], CREBBP binds to and acetylates BCL6, leading to a dose dependent impairment in

its ability to repress a BCL6-responsive reporter gene (Fig. S6a,b). We then generated HA-tagged constructs for expression of 9 representative CREBBP alleles harboring missense mutations within (n=6) or immediately outside (n=2) the core HAT domain, as well as a premature stop codon (R1360X) (Fig. 4a). Notably, all of the HAT domain mutant proteins had lost their ability to acetylate BCL6 (Fig. 4b) and to interfere with its transrepression activity, consistent with the reported role of acetylation in inactivating BCL6 (Fig 4c). Conversely, no significant effects were observed from the C1240R and K1320R proteins or from two additional alleles (P1053L, Q1079H) harboring mutations outside the HAT domain (Table S1 and data not shown), although C1240R appeared to have lost its activity in the reporter assay, suggesting that alternative mechanisms may be involved.

When tested on the tumor suppressor p53, the same core HAT mutants were either impaired or severely attenuated (H1487Y and D1435E) in their acetylation activity, as compared to wild-type CREBBP or to C1240R and K1320R (Fig. 4d). It should be noted that C1240R and K1320R represent a second mutational event in cases carrying an additional truncating *CREBBP* mutation (see Table S1), suggesting the possibility that they represent passenger events or they confer more subtle functional consequences. Taken together, these findings demonstrate that CREBBP missense mutant alleles have been selected for their defective HAT activity. In particular, the impairment on BCL6 and p53 acetylation is consistent with specific effects of *CREBBP* genetic lesions in favoring the constitutive activity of the BCL6 oncogene over the function of the p53 tumor suppressor.

CREBBP mutations reduce affinity for Acetyl-CoA binding

To investigate the mechanisms responsible for the observed loss of function on BCL6 and p53, we examined the effect of CREBBP mutations on subcellular localization, enzyme-substrate complex formation, and enzymatic activity. All mutants tested were correctly localized in the nuclear compartment and could be efficiently co-immunoprecipitated with BCL6 or p53 (Fig. S7 and data not shown), indicating that the inability to acetylate these substrates *in vivo* was not due to mislocalization or to impaired physical interaction. Instead, *in vitro* acetylation assays using recombinant GST-p53 and highly purified, wild-type or mutated CREBBP proteins demonstrated a marked, Acetyl-CoA dose-dependent defect in the enzymatic activity of the core-HAT mutants (Fig. 5 and Fig. S8), suggesting that these changes impair the affinity for Acetyl-CoA (see also Fig. S5).

CREBBP missense mutants fail to rescue the phenotype of *Crebbp*/*Ep300* null cells

To further investigate the consequences of CREBBP missense mutations on its function as a transcriptional co-activator, we examined the response of endogenous cAMP response element-binding protein (CREB) target genes in mouse embryonic fibroblasts carrying conditional *Crebbp* and *Ep300* knock-out alleles (dKO MEFs), reconstituted with retroviral vectors expressing either wild-type or DLBCL-associated mutated Crebbp cDNAs (Fig. S9a), and treated with forskolin and IBMX (FI), two inducers of PKA activity. In normal cells, PKA-mediated phosphorylation of CREB is required for the recruitment of CREBBP, which in turn leads to transcriptional activation of cAMP-responsive genes[38]. As previously observed[39], quantitative RT-PCR analysis of CREBBP/EP300-dependent CREB target genes confirmed their upregulation in dKO cells reconstituted with wild-type CREBBP expression constructs (Fig. S9b, column 2 from left). In contrast, all four DLBCL-derived CREBBP mutants tested were generally deficient for cAMP-responsive transcription, as was the HAT-dead control W1502A/Y1503S[40] (Fig. S9b and Fig. S10). These mutants were also associated with reduced endogenous histone H3K18 acetylation (Fig. S11). Finally, we tested whether the mutant CREBBP polypeptides were proficient in rescuing the proliferative defect of the dKO cells, as this system provides a specific readout for the biological activity of the CREBBP point mutants in a physiologic setting. Notably, cells transduced with the four DLBCL-associated alleles displayed significantly impaired cell growth compared to wild-type reconstituted cells, as measured in the YFP+ (dKO) population (Fig. S9c). Collectively, the data presented above provide direct experimental evidence for a role of HAT missense mutations in impairing CREBBP acetyltransferase activity and its function as a transcriptional coactivator.

Discussion

The results herein indicate that inactivation of *CREBBP/EP300* represents a common event in the two most frequent forms of B-NHL, namely FL and DLBCL. Previous extensive surveys in malignancies of epithelial origin have reported inactivating mutations of *EP300* and *CREBBP* in exceedingly rare cases (<2% of primary biopsies)[26-28]. Considering their virtual absence in solid tumors, and the finding of recurrent mutations in B-ALL[34], our results point to a specific role for *CREBBP/EP300* inactivation in the pathogenesis of malignancies derived from B-lymphocytes. Overall, *CREBBP/EP300* lesions are among the most frequent structural alterations yet detected in FL and DLBCL, thus representing an important feature of the pathogenesis of these common diseases. Moreover, the observation of significantly reduced CREBBP and EP300 expression levels in a sizeable fraction of DLBCL samples,

independent of genetic lesions (Fig. 3e and Fig. S12), suggests that additional epigenetic mechanisms may cause reduction of HAT dosage in a larger fraction of tumors.

One key observation of this study is that *CREBBP/EP300* lesions are mostly detected in heterozygosity, suggesting a haploinsufficent role in tumor suppression. This notion is supported by several observations. First, congenital heterozygous mutations of *CREBBP/EP300* are sufficient to cause significant pathologic and developmental phenotypes, including tumorigenesis, in patients with RTS[23-25], thereby confirming the deleterious effect of reduced HAT activity. Second, a fraction of mice with conditional deletion of *Crebbp* in mature B-cells showed reduced survival past 12-months of age[41]. Hematologic malignancies were also reported in 22% of constitutive 10-21 month-old *Crebbp* heterozygous mice and ~30% of chimeric animals after bone marrow or spleen cell transplantation from *Crebbp* heterozygotes[42]. Third, in contrast with the abundantly expressed histone deacetylases (HDACs), HATs are limiting in the cell, suggesting that small dosage variations can have severe biological consequences[43]. Overall, these data provide direct evidence for CREBBP (and, more rarely, EP300) as haploinsufficient tumor suppressors, whose specific role in lymphomagenesis will have to be tested by conditional deletion of these alleles in GC B-cells.

Given the global involvement of CREBBP/EP300 on gene transcriptional regulation, it is difficult to predict which cellular targets/pathways may be critically affected by HAT reduction in lymphomagenesis. At this stage, our results demonstrate that mutant CREBBP and EP300 proteins are deficient in acetylating BCL6 and p53, leading to constitutive activation of the oncoprotein and to decreased p53 tumor suppressor activity. The balance between the activities of these two genes is critical for the regulation of DNA damage responses in mature GC cells during immunoglobulin gene remodeling[36,44]. Thus, the consequences of BCL6 activity overriding p53 would be an increased tolerance for DNA damage in the context of diminished apoptotic and cell cycle arrest responses.

These results have important therapeutic implications in view of current attempts to utilize a variety of HDAC inhibitors as anti-cancer drugs. While the benefits of these compounds has been proven in certain malignancies, such as mature T-cell lymphoproliferative disorders, their efficacy in other common cancers, including B-NHL, is uncertain at this stage[45]. The findings of this study suggest that the use of HDAC inhibitors has a rational basis in B-NHL, as it may contribute to re-establishing physiologic acetylation levels. On the other hand, their efficacy should be re-evaluated by stratifying patients based

on the presence of HAT defects as well as by testing the numerous HDAC and sirtuin inhibitors with target specificity.

Methods Summary

Mutation analysis. The complete coding sequences and exon/intron junctions of *CREBBP* and *EP300* were analyzed by PCR amplification and direct sequencing of whole genome amplified DNA using the oligonucleotides reported in Tables S5 and S6. Mutations were confirmed from both strands on independent PCR products amplified from genomic DNA, and their somatic origin was documented by analysis of matched normal DNA in available cases.

SNP array analysis was performed using Affymetrix Genome-Wide 6.0 Arrays and a computational workflow described in detail in the Supplementary Information.

***In vivo* and *in vitro* characterization of HAT activity.** The ability of CREBBP mutants to acetylate BCL6 and p53 was assessed in HEK293T cells after co-transfection of pCMV-Flag-BCL6 (or pCIN4-Flag-p53) with plasmids encoding wild-type vs mutant HA-tagged mouse Crebbp. BCL6 acetylation was evaluated on Flag/M2 immunoprecipitates using antibodies directed against acetyl lysines; for p53, western blot analysis was performed on whole cell extracts using a specific anti-acetylated p53 antibody. The amounts of exogenous CREBBP were monitored using anti-HA and anti-CREBBP (A22, Santa Cruz Biotechnology) antibodies. *In vitro* acetylation assays were performed using recombinant GST-p53 and purified CREBBP-HA proteins, in the presence of the indicated amounts of acetyl-CoA.

Transient Transfection/Reporter gene assays. The effect of CREBBP on BCL6-dependent transcription was assessed in HEK293T cells co-transfected with a luciferase reporter vector containing 5 BCL6 consensus binding sites upstream of the SV40 promoter (5XBCL6)[46] and the pCMV-Flag-BCL6 construct, in the absence or presence of wild-type vs mutant CREBBP-HA expression vectors (see Full Methods).

Reconstitution of *Crebbp/Ep300* null MEFs. *Crebbp*$^{flox/flox}$*;Ep300*$^{flox/flox}$*;*YFP^{+} conditional (dKO) MEFs have been described[39]. CREBBP expression was restored by retroviral infection with constructs encoding for HA-tagged wild-type or mutant Crebbp, and cells were analyzed for H3K18 acetylation, cAMP-dependent transcriptional responses and cell proliferation (see Full Methods).

.

References

1. Swerdlow, S.H. et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, (International Agency for Research on Cancer (IARC), Lyon, 2008).
2. Compagno, M. et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-21 (2009).
3. Davis, R.E. et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88-92 (2010).
4. Klein, U. & Dalla-Favera, R. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol* **8**, 22-33 (2008).
5. Lenz, G. et al. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676-9 (2008).
6. Lenz, G. & Staudt, L.M. Aggressive lymphomas. *N Engl J Med* **362**, 1417-29 (2010).
7. Mandelbaum, J. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B-cell like diffuse large B-cell lymphoma. *Cancer Cell* (2010).
8. Morin, R.D. et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-5 (2010).
9. Downing, J.R. Cancer genomes--continuing progress. *N Engl J Med* **361**, 1111-2 (2009).
10. Goodman, R.H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**, 1553-77 (2000).
11. Kalkhoven, E. CBP and p300: HATs for different occasions. *Biochem Pharmacol* **68**, 1145-55 (2004).
12. Bannister, A.J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641-3 (1996).
13. Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953-9 (1996).
14. Gu, W., Shi, X.L. & Roeder, R.G. Synergistic activation of transcription by CBP and p53. *Nature* **387**, 819-23 (1997).
15. Lill, N.L., Grossman, S.R., Ginsberg, D., DeCaprio, J. & Livingston, D.M. Binding and modulation of p53 by p300/CBP coactivators. *Nature* **387**, 823-7 (1997).
16. Avantaggiati, M.L. et al. Recruitment of p300/CBP in p53-dependent signal pathways. *Cell* **89**, 1175-84 (1997).
17. Blobel, G.A., Nakajima, T., Eckner, R., Montminy, M. & Orkin, S.H. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci U S A* **95**, 2061-6 (1998).
18. Bereshchenko, O.R., Gu, W. & Dalla-Favera, R. Acetylation inactivates the transcriptional repressor BCL6. *Nat Genet* **32**, 606-13 (2002).
19. Grossman, S.R. et al. Polyubiquitination of p53 by a ubiquitin ligase activity of p300. *Science* **300**, 342-4 (2003).
20. Shi, D. et al. CBP and p300 are cytoplasmic E4 polyubiquitin ligases for p53. *Proc Natl Acad Sci U S A* **106**, 16275-80 (2009).
21. Oike, Y. et al. Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. *Blood* **93**, 2771-9 (1999).
22. Yao, T.P. et al. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361-72 (1998).

23. Roelfsema, J.H. & Peters, D.J. Rubinstein-Taybi syndrome: clinical and molecular overview. *Expert Rev Mol Med* **9**, 1-16 (2007).
24. Petrij, F. et al. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348-51 (1995).
25. Miller, R.W. & Rubinstein, J.H. Tumors in Rubinstein-Taybi syndrome. *Am J Med Genet* **56**, 112-5 (1995).
26. Iyer, N.G., Ozdag, H. & Caldas, C. p300/CBP and cancer. *Oncogene* **23**, 4225-31 (2004).
27. Gayther, S.A. et al. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet* **24**, 300-3 (2000).
28. Ward, R., Johnson, M., Shridhar, V., van Deursen, J. & Couch, F.J. CBP truncating mutations in ovarian cancer. *J Med Genet* **42**, 514-8 (2005).
29. Garbati, M.R., Alco, G. & Gilmore, T.D. Histone acetyltransferase p300 is a coactivator for transcription factor REL and is C-terminally truncated in the human diffuse large B-cell lymphoma cell line RC-K8. *Cancer Lett* **291**, 237-45 (2010).
30. Shigeno, K. et al. Disease-related potential of mutations in transcriptional cofactors CREB-binding protein and p300 in leukemias. *Cancer Lett* **213**, 11-20 (2004).
31. Borrow, J. et al. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet* **14**, 33-41 (1996).
32. Rowley, J.D. et al. All patients with the T(11;16)(q23;p13.3) that involves MLL and CBP have treatment-related hematologic disorders. *Blood* **90**, 535-41 (1997).
33. Sobulo, O.M. et al. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc Natl Acad Sci U S A* **94**, 8732-7 (1997).
34. Mullighan, C.G. et al. CREBBP mutations in relapsed acute lymphoblastic leukemia. *Nature* (2010).
35. Liu, X. et al. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846-50 (2008).
36. Phan, R.T. & Dalla-Favera, R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* **432**, 635-9 (2004).
37. Tang, Y., Zhao, W., Chen, Y., Zhao, Y. & Gu, W. Acetylation is indispensable for p53 activation. *Cell* **133**, 612-26 (2008).
38. Kwok, R.P. et al. Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature* **370**, 223-6 (1994).
39. Kasper, L.H. et al. CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation. *EMBO J* (2010).
40. Bordoli, L. et al. Functional analysis of the p300 acetyltransferase domain: the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. *Nucleic Acids Res* **29**, 4462-71 (2001).
41. Xu, W. et al. Global transcriptional coactivators CREB-binding protein and p300 are highly essential collectively but not individually in peripheral B cells. *Blood* **107**, 4407-16 (2006).
42. Kung, A.L. et al. Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev* **14**, 272-7 (2000).
43. Legube, G. & Trouche, D. Regulating histone acetyltransferases and deacetylases. *EMBO Rep* **4**, 944-7 (2003).
44. Phan, R.T., Saito, M., Basso, K., Niu, H. & Dalla-Favera, R. BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nat Immunol* **6**, 1054-60 (2005).

45. Stimson, L., Wood, V., Khan, O., Fotheringham, S. & La Thangue, N.B. HDAC inhibitor-based therapies and haematological malignancy. *Ann Oncol* **20**, 1293-302 (2009).
46. Huynh, K.D., Fischle, W., Verdin, E. & Bardwell, V.J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev* **14**, 1810-23 (2000).

Accession numbers. The Affymetrix expression data reported in this paper have been deposited in the NCBI Gene Expression Omnibus (GEO) database (Series Accession Number GSE12195). The SNP Array 6.0 data and the whole exome sequencing data from the 7 DLBCL cases have been deposited in dbGaP under accession no. phs000328.v1.p1.

Full Methods

DNA extraction, amplification and sequencing. Genomic DNA was extracted by standard methods, and whole genome amplification was performed using the RepliG kit (QIAGEN) according to the manufacturer's instructions. Sequences for all annotated exons and flanking introns of *CREBBP* and *EP300* were obtained from the UCSC Human Genome database, using the corresponding mRNA accession number as a reference (NM_004380.2 and NM_001429.3, respectively). PCR primers, located≥50 bp upstream or downstream to target exon boundaries, were designed in the Primer 3 program (http://frodo.wi.mit.edu/primer3/) and filtered using UCSC In Silico PCR to exclude pairs yielding more than a single product (Table S5 and Table S6). Purified amplicons were sequenced directly from both strands (Genewiz, Inc, South Plainfield, NJ) as described, and compared to the corresponding germline sequences, using the Mutation Surveyor Version 2.41 software package (Softgenetics, State College, PA; http://www.softgenetics.com). Somatic mutations were confirmed on independent PCR products obtained from high molecular weight genomic DNA. Synonymous mutations, previously reported polymorphisms (Human dbSNP Database at NCBI, Build 130, and Ensembl Database) and changes present in the matched normal DNA, when available, were excluded. In cases carrying multiple events within a single gene, the allelic distribution of the mutations was determined by cloning and sequencing PCR products obtained from cDNA and spanning both events (N=10 clones each).

Northern Blot analysis of CREBBP and EP300 expression. Total RNA (12μg) was extracted from exponentially growing cell lines by TRIzol (Invitrogen) and Northern blot analysis was performed according to standard procedures, with radiolabeled probes corresponding to a 0.9Kb fragment of the human *CREBBP* cDNA (region 204-1143, GenBank accession No. NM_004380.2) or a 1.2Kb fragment of the human *EP300* cDNA (region 6277-7524, GenBank accession No. NM_001429.3), and GAPDH as a control for loading. *CREBBP* and *EP300* expression levels were then quantitated by Phosphorimager analysis and normalized with the GAPDH levels.

Tissue microarrays, immunohistochemistry (IHC) and immunofluorescence (IF) analysis. DLBCL tissue microarrays were constructed according to standard procedures and analyzed by IHC, using rabbit polyclonal antibodies directed against the N-terminus of CREBBP (A22) or EP300 (N15)(Santa Cruz Biotechnology). Cases were scored as positive if

≥20% tumor cells were stained by the antibody. IF analysis of CREBBP expression in conditional *Crebbp/Ep300* dKO MEFs was performed 3 days after deletion of the endogenous loci and 5 days after infection with HA-tagged CREBBP retrovirus, using the CREBBP (A22) antibody. The specificity of both the A22 and the N15 antibodies had been previously validated on paraffin-embedded cell pellets from HEK293T cells transfected with control and CREBBP-specific shRNAs as well as by IF staining of *Crebbp/Ep300* dKO MEFs (data not shown).

Protein extraction and western blot analysis. Whole cell extracts were prepared in RIPA buffer containing protease inhibitors as described[18] and were analyzed by western blotting according to standard methods, using the following primary antibodies: anti-FLAG/M2 (Sigma), anti-HA (3F10)(Roche), anti-BCL6 (Cell Marque, GI191E/A8), anti-acetyl lysines (Cell Signaling, rabbit), anti-acetylated p53 (kind gift of W. Gu), anti p53 (DO1), anti-CREBBP (A22), anti-EP300 (N15)(all from Santa Cruz Biotechnology), anti-GFP (JL-8)(Clontech), anti-β-actin (clone AC-15) and anti-β-tubulin (clone B-5-1-2)(both from Sigma). Proteins were resolved by SDS-PAGE in 3-8% NuPAGE Tris-Acetate gels (Invitrogen), and visualized using a chemiluminescence detection kit (Pierce) as recommended by the manufacturer.

Transient transfections/reporter gene assays. Transient transfections were performed in HEK293T cells using polyethylenimine (PEI), as described[47]. For reporter assays, cells were seeded on a 24-well plate and transfected using 100 ng of a Luciferase reporter construct driven by a PGL3-SV40-based backbone with 5 BCL6 binding sites[46], 2.5 ng of pRL-SV40 (Promega), 1 ng of BCL6-encoding plasmid and the indicated doses of wild-type and mutant CREBBP expression vectors. The total amount of transfected DNA was kept constant in each experiment by adding pCMV-HA or pcDNA3 vector sequences to a final amount of 600ng/well, and 3.5 μg of PEI. All experiments were performed in duplicate and luciferase activities were measured thirty-six hours after transfection using the Dual-Luciferase Reporter Assay System (Promega), according to the manufacturer's instructions. Given the known effects of CREBBP on the SV40 promoter and on other regulatory sequences, the response of the BCL6 reporter was normalized first to that of a SV40-Renilla reporter construct, and then to the basal activity of CREBBP on the 5X-BCL6 reporter, in the absence

of BCL6. Data are expressed as differences relative to the basal activity of the reporter construct (set as 1) after the above-mentioned normalization.

Co-immunoprecipitation assays. To assess the interaction between CREBBP and BCL6 or p53, HEK293T cells were co-transfected with plasmids encoding for the two proteins, together with wild-type or mutant CREBBP-HA. Thirty-six hours after transfection (or twenty-four hours for p53), cells were lysed in IP buffer (50 mM Tris, pH 7.0, 250 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.05% NP40, 10 mM sodium fluoride, 0.1mM sodium orthovanadate and protease inhibitor cocktail [Sigma]), and the cleared lysates were incubated overnight at 4°C with anti-HA or anti-Flag/M2 beads (Sigma). Immunocomplexes were eluted by incubating the beads in IP buffer containing HA or 3X-Flag peptide, respectively (Sigma). A fraction of the final eluates was resolved by SDS-PAGE and analyzed by western blot.

***In vitro* acetyl-transferase assays.** Recombinant GST-p53 and CREBBP-HA proteins were obtained as described in detail in the Supplementary Information and used for *in vitro* acteylation assays according to published protocols, with minor modifications[14,48,49]. Briefly, 10-24 ng of recombinant CREBBP-HA and 100 ng of GST-p53 were combined in 40 uL reactions containing 50 mM Tris-Cl (pH 8.0), 10% Glycerol, 1mM DTT, 1mM PMSF, 0.1 mM EDTA, 10 mM Butyric Acid (#B103500, Sigma), and variable amounts of Acetyl-CoA (Sigma)(2mM to 25 nM). As shown in Fig. 6, Acetyl-CoA concentrations as low as 25 nM were sufficient to obtain efficient acetylation of the substrate by CREBBP. Reactions were performed for 60 minutes at 30°C, and stopped by addition of an equal volume of 2X Laemmli buffer followed by heating for 10 minutes at 70°C. A fraction of the final product was resolved by SDS-PAGE in 3-8% Tris-Acetate gradient gels (Invitrogen) and analyzed by western blotting.

Reconstitution of conditional *Crebbp/Ep300* dKO MEFs for analysis of CREB target genes transcription, H3K18 acetylation and cell growth. The protocols for MEF isolation, cell culture and retroviral transduction, as well as the murine Crebbp (CBP)-HA retroviral construct used as template for the generation of various CREBBP mutants have been described previously[50]. In this system, Cre mediated recombination induces expression of YFP, allowing for specific identification of the deleted cells. *Crebbp*$^{flox/flox}$; *Ep300*$^{flox/flox}$;

YFP$^+$ MEFs were first infected with retroviruses encoding either wild-type or selected CREBBP point mutants, and the endogenous *Crebbp*^{flox} and *Ep300*^{flox} loci were deleted after 48 hours by infection with Cre-expressing adenovirus (Ad-Cre). The W1502A/Y1503S HAT-dead mutation was used as negative control[40]. In all experiments, transduction efficiencies were 70% or higher, as assessed by IF analysis of HA positive cells 3 days after deletion of endogenous *Crebbp/Ep300* and 5 days after retroviral infection. To examine cAMP-dependent gene expression, MEFs were cultured for 16 hrs in DMEM containing 0.1% FBS and treated for 90 min with 10μM forskolin + 100μM IBMX (or ethanol vehicle) before harvesting in TRIzol (Invitrogen); qRT-PCR assays were performed as reported[48]. Expression of Crebbp and Ep300 was verified by IF analysis using rabbit polyclonal anti-CREBBP (A-22) and anti-EP300 (N-20) antibodies (Santa Cruz Biotechnology). Analysis of H3K18 acetylation was performed as described, using the anti-H3K18Ac antibody (ab1191) (Abcam) and the HA-11 monoclonal antibody against the HA epitope (Boehringer Mannheim). Confocal images were taken using the same settings for all mutants and mean intensity ratios for individual nuclei were collected using SlideBook 5 software. Nuclei where the Crebbp (HA) signal was at least 2.5 fold above background were used to calculate the ratio of the H3K18Ac mean signal intensity to the Crebbp (HA) mean signal intensity. For the growth assays shown in Fig. S11, *Crebbp*$^{flox/flox}$*;Ep300*$^{flox/flox}$*;YFP*$^+$ MEFs were infected with CREBBP retroviruses 48 hours before Ad-Cre mediated deletion of endogenous *Crebbp* and *Ep300.* From each reconstituted population, equivalent numbers of YFP$^+$ (dKO) MEFs were seeded at day 1 (i.e., the day following overnight Ad-Cre treatment), and the total number of YFP$^+$ cells was calculated on Day 11 from the total cell number, based on the percentage of YFP$^+$ cells as assessed by flow cytometric analysis.

References not cited in the main text

47. Bieber, T. & Elsasser, H.P. Preparation of a low molecular weight polyethylenimine for efficient cell transfection. *Biotechniques* **30**, 74-7, 80-1 (2001).
48. Kuninger, D., Lundblad, J., Semirale, A. & Rotwein, P. A non-isotopic in vitro assay for histone acetylation. *J Biotechnol* **131**, 253-60 (2007).
49. Tang, Y., Luo, J., Zhang, W. & Gu, W. Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol Cell* **24**, 827-39 (2006).
50. Bedford, D.C., Kasper, L.H., Fukuyama, T. & Brindle, P.K. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* **5**, 9-15 (2010).

Main Figures

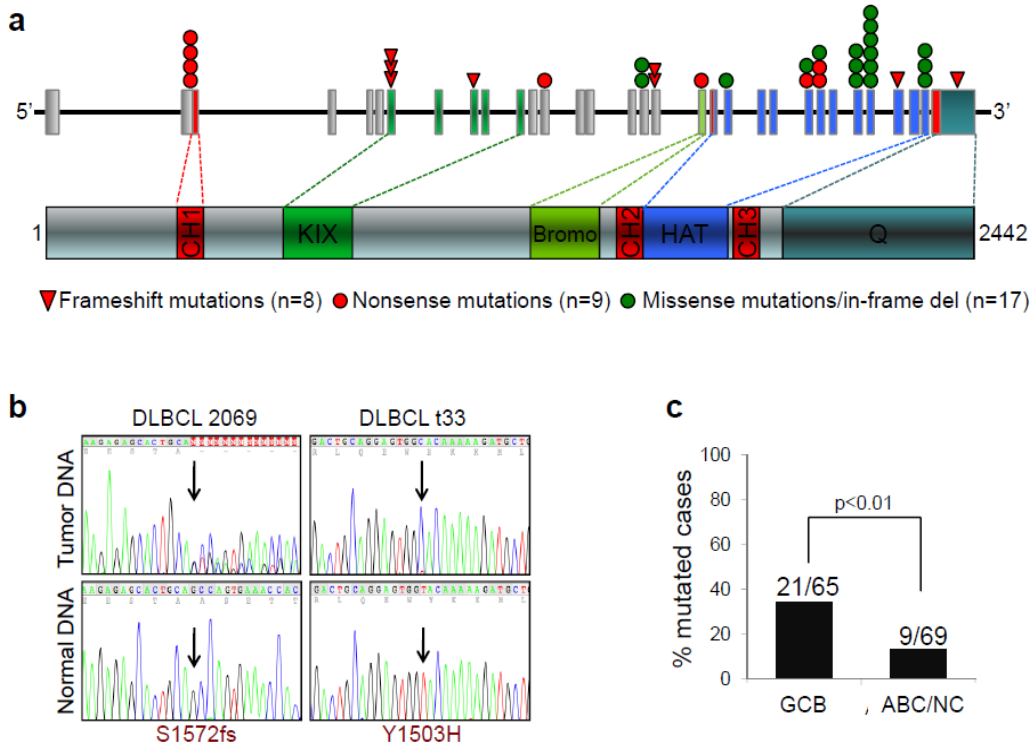


Figure 1. The *CREBBP* gene is mutated in DLBCL. a, Schematic diagram of the *CREBBP* gene (top) and protein (bottom). Exons are color coded according to the corresponding protein functional domains (CH, cysteine-histidine rich; KIX, CREB-binding; Bromo, bromodomain; HAT, histone acetyltransferase; Q, poly glutamine stretch). Color-coded symbols depict distinct types of mutations. **b,** Sequencing traces of representative mutated DLBCL tumor samples and paired normal DNA; arrows point to the position of the nucleotide change (amino acid change shown at the bottom). **c,** Distribution of CREBBP mutations in major DLBCL subtypes; on top, actual number of mutated samples over total analyzed.

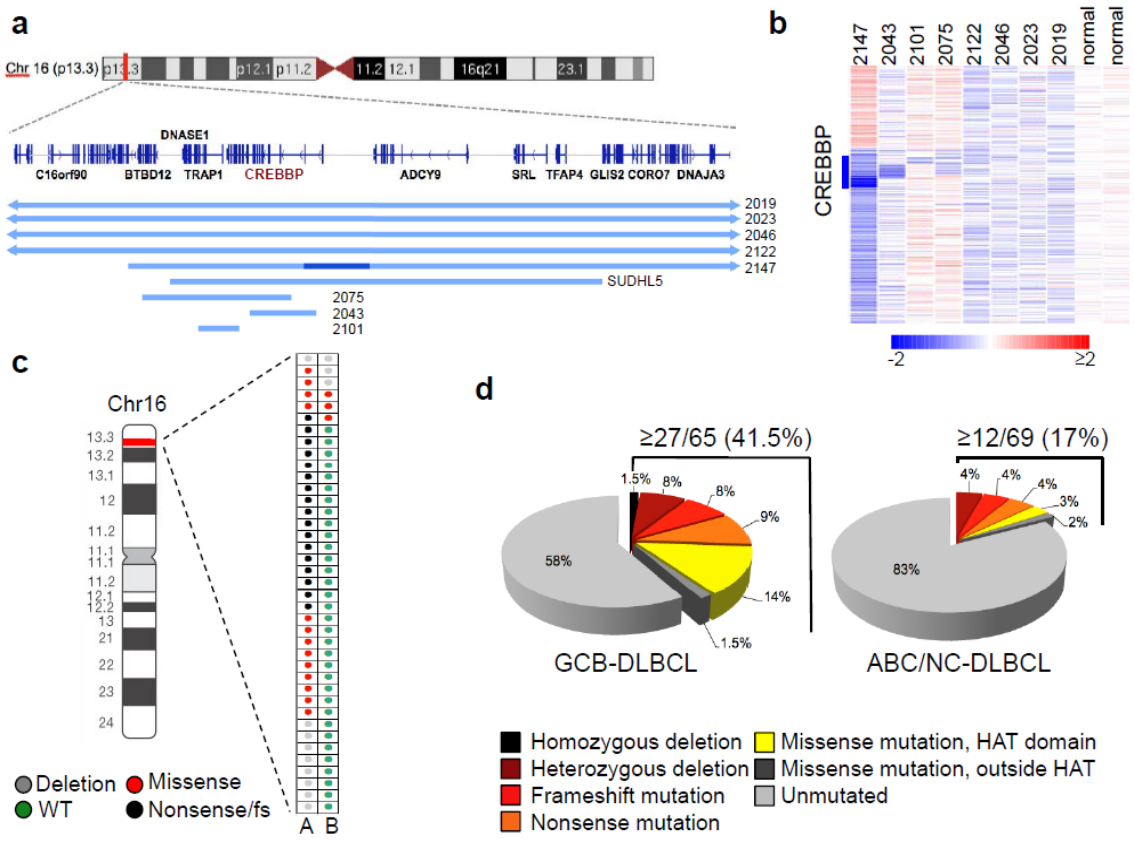


Figure 2. Mutations and deletions of *CREBBP* are predominantly monoallelic. a, Map of the genomic region encompassing *CREBBP*. Blue lines below the map indicate the extent of the deletions identified in 9 DLBCL samples, with the dark blue segment corresponding to a homozygous loss. **b,** dChipSNP heatmap showing median-smoothed log2 CN ratio for 8 DLBCL biopsies harboring *CREBBP* deletions, and two normal DNAs (N). A vertical blue bar indicates the location of the *CREBBP* locus; in the red-blue scale, white corresponds to a normal (diploid) CN log-ratio, blue is deletion and red is gain. **c,** Allelic distribution of *CREBBP* genetic lesions in individual DLBCL samples. **d,** Overall frequency of *CREBBP* structural alterations in DLBCL subtypes (mutations and deletions, combined).

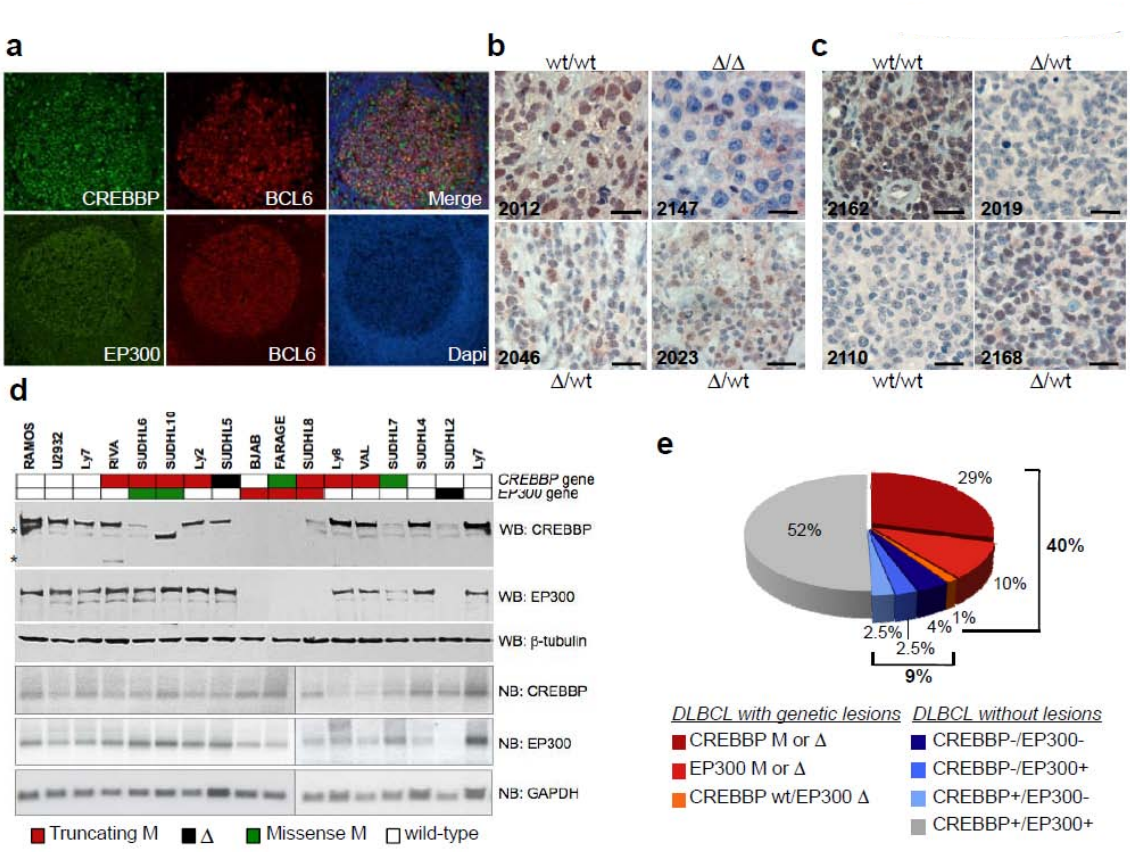


Figure 3. CREBBP and EP300 expression in normal and transformed B-cells. a, Immunofluorescence analysis of reactive tonsils. BCL6 identifies GC B-cells, and Dapi is used to detect nuclei. **b, c,** Immunohistochemistry analysis of CREBBP (b) and EP300 (c) protein expression in representative DLBCL biopsies (genomic status as indicated; scale bar, 100μm). Sample 2147, which harbors a homozygous *CREBBP* deletion, serves as negative control. **d,** Western blot and northern blot analysis of DLBCL cell lines carrying wild-type or aberrant *CREBBP* and *EP300* alleles (color coded as indicated). The aberrant band in SUDHL10 corresponds in size to the predicted ~220kD CREBBP truncated protein. * non-specific bands. Tubulin and GAPDH control for total protein and RNA loading, respectively. **e,** Overall proportion of DLBCL biopsies showing defective CREBBP/EP300 function due to genetic lesions (red scale) and/or lack of protein expression (blue scale).

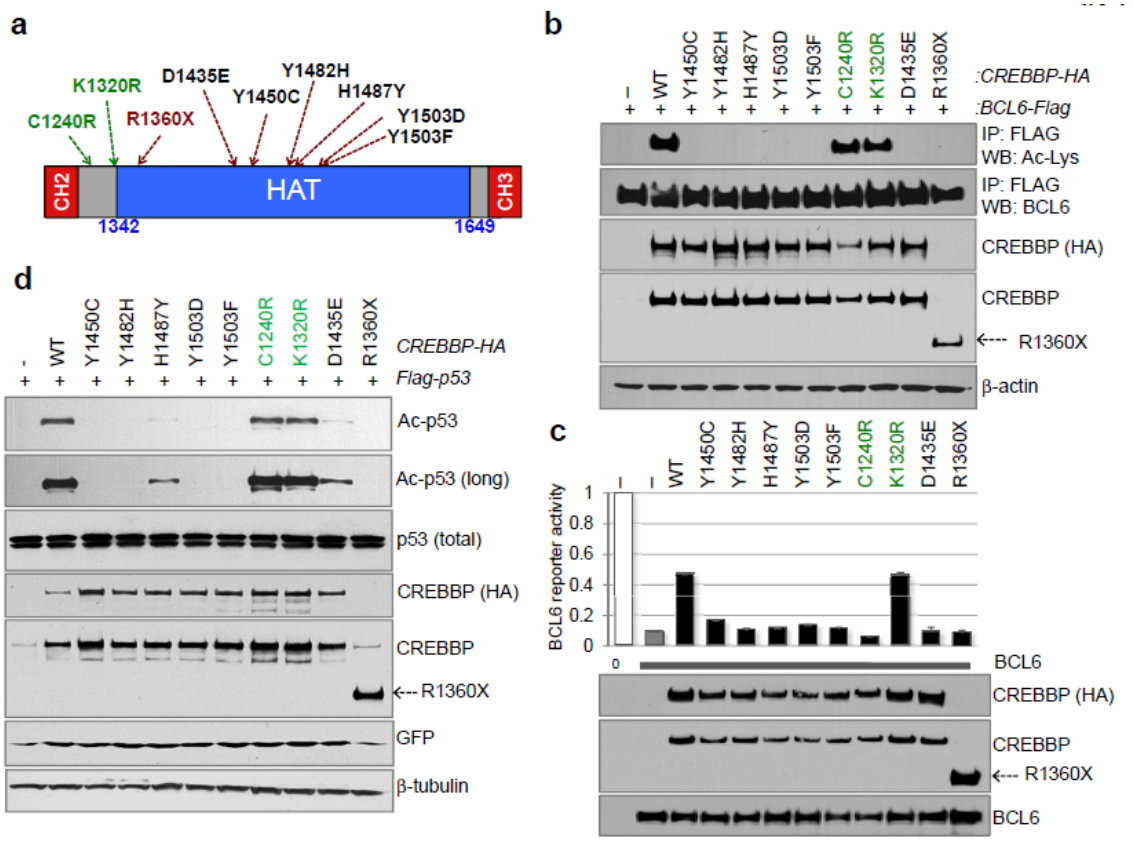


Figure 4. CREBBP missense mutations impair its ability to acetylate BCL6 and p53. a, Schematic diagram of the CREBBP HAT and CH domains, with the CREBBP point mutations tested in *b-d* (in green, residues located immediately outside the HAT domain). **b,** Acetylation levels of exogenous BCL6 in Flag immunoprecipitates obtained from HEK293T cells co-transfected with wild-type or mutant CREBBP expression vectors. Actin, input loading control. **c,** Luciferase reporter assays using a synthetic 5X-BCL6 reporter. Results are shown as relative activity compared to the basal activity of the reporter, set as 1 (mean +/- SD, n=2). Bottom, BCL6 and CREBBP-HA protein levels in the same lysates. Note that the amount of transfected BCL6 and CREBBP-encoding plasmids was adjusted to achieve equal protein amounts. **d,** p53 acetylation in HEK293T cells co-transfected with the indicated CREBBP expression vectors. The anti-p53 antibody documents comparable amounts of p53 (exogenous+endogenous). GFP monitors for transfection efficiency, and actin is used as loading control.

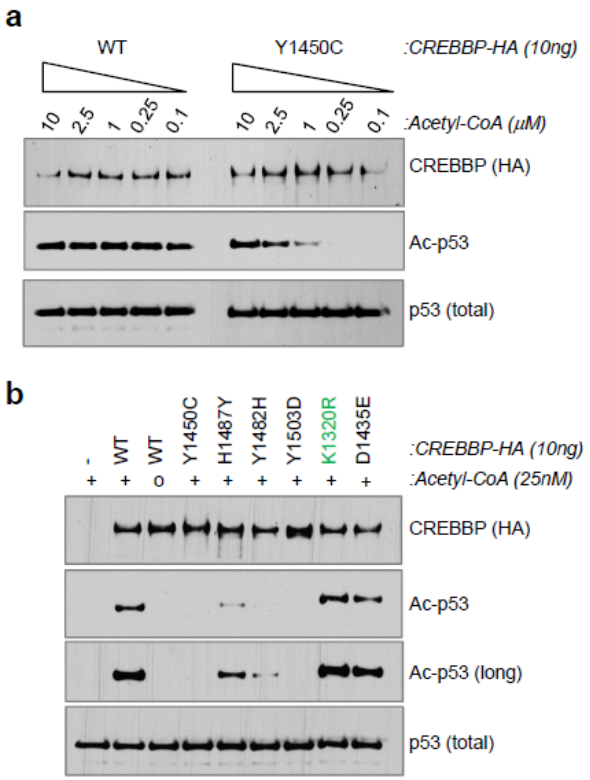


Figure 5. DLBCL-associated mutations in the CREBBP HAT domain decrease its affinity for Acetyl-CoA. a, Western blot analysis of *in vitro* acetyltransferase reactions performed using the mutant Y1450C CREBBP-HA recombinant protein and a GST-p53 substrate in the presence of decreasing amounts of Acetyl-CoA. Anti-HA and anti-p53 antibodies document the presence of equivalent amounts of effector and substrate proteins in the reaction. **b,** *In vitro* acetyltransferase activity of the indicated CREBBP-HA mutant proteins in the same assay using 25nM Acetyl-CoA.

Supplementary Material

Cell lines. The following DLBCL cell lines were used in the study: OCI-Ly1, OCI-Ly3, OCI-Ly4, OCI-Ly7, OCI-Ly8, OCI-Ly10, OCI-Ly18, HBL-1, SUDHL-2, U2932, RIVA, RC-K8, SUDHL-4, SUDHL-5, SUDHL-6, SUDHL-7, SUDHL-8, SUDHL-10, DB, BJAB, FARAGE, VAL, and WSU. All cell lines were maintained in Iscove's Modified Dulbecco Medium (IMDM) supplemented with 10% FBS, 100 μg/mL penicillin, 100 μg/mL streptomycin and 2 mM L-glutamine, except for OCI-Ly10 and OCI-Ly4, which were cultured in IMDM with 20% heparinized human plasma and 55 μM β-mercaptoethanol. HEK293T, H1299 and HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 100 μg/mL penicillin, and 100 μg/mL streptomycin. The following BL cell lines were included in the mutation analysis of *CREBBP* and *EP300*: Ramos, P3HRI, BL113, DAUDI, RAJI, ODHI1, MUTU1.

Primary tumor samples. Primary biopsies from 111 newly diagnosed, previously untreated DLBCL patients were obtained as paraffin-embedded and/or frozen material from the archives of the Departments of Pathology at Columbia University, Weill Cornell Medical College, and the Department of Clinical and Experimental Medicine at the University of Novara, after approval by the respective Institutional Review Boards. The fraction of tumor cells, assessed by southern blot analysis of the rearranged immunoglobulin heavy chain locus and/or by histologic analysis of frozen sections isolated before and after obtaining tissue for molecular studies, corresponded to >80% in most of the cases and to >50% in all cases. The DLBCL cohort (cell lines and primary biopsies) comprised 65 GCB-DLBCL, 54 ABC/non-GC-DLBCL, and 5 unclassified (NC) cases, as determined previously[1] based on gene expression profile analysis[2] (available for 113 samples) or immunohistochemical stains[3] (n=11). None of the patients included in the study had a clinical history of RTS. High molecular weight genomic DNA from 46 FL, 53 CLL, 23 BL and 11 MZL patients were provided by the Department of Clinical and Experimental Medicine, University of Novara.

Copy number determination of *CREBBP* and *EP300* in DLBCL by high-density SNP array analysis. Genome-wide DNA profiles were obtained from high molecular weight genomic DNA of DLBCL patients using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA), following the manufacturer's instructions. Image data analysis and quality control for the hybridized samples were performed using the

Affymetrix Genotyping Console 3.0.1 software, and only samples passing the Affymetrix recommended contrast QC and SNP call rates threshold (in the Birdseed v2.0 algorithm) were considered for analysis. Affymetrix CEL files and corresponding SNP genotype call files generated by the Affymetrix Genotyping Console tool were then analyzed using the dCHIP software[4] according to a previously published workflow[5,6]. Model-based expression was performed using the perfect-match (PM) model to summarize signal intensities for each probe set. Probe intensity data for each array were normalized using a diploid reference set of 3 normal (non tumor) DNA samples that had been processed and hybridized in the same experiment as the tumor samples. The standard invariant-set normalization approach in dCHIP was implemented by a karyotype guided normalization method as described in detail in Ref. 5 and 6. Candidate genomic regions of amplification and deletion were identified by applying the circular binary segmentation (CBS) algorithm to the SNP array data as described[5,7], with the following criteria: i) mean log2 ratios of ≥ 0.2 or ≤ -0.2 (≥ 0.15 or ≤ -0.15 in four cases where the % of tumor cells in the biopsy was estimated to be ~60%); ii) ≥8 SNP markers (SNP and/or CNV) within a segment. The results of the CBS algorithm were then compared to those of dChip. To exclude calls of genomic gains or loss arising from inherited genomic CNV, the dChipSNP algorithm was also applied to 130 normal DNAs from an independent study[5] as well as to 230 normal DNAs from the HapMap project; alterations identified in the pool of reference samples were assumed to be inherited and therefore excluded. In addition, CNV were excluded if present in the Database of Genomic Variants (http://projects.tcag.ca/cgi-bin/variation/gbrowse/hg18/).

cDNA synthesis and RT-PCR analysis. Total RNA extracted by TRIzol was treated with DNAse prior to cDNA synthesis, according to the manufacturer's instructions. The cDNA was then used as a template in PCR amplification reactions. For detection of the normal and mutant *CREBBP* and *EP300* alleles in affected samples, primers were designed based on the transcript sequences surrounding the validated mutations, and the amplified PCR products were then analyzed by direct sequencing or by cloning and sequencing.

Gene expression data. Gene expression profile analysis of normal B cells and primary DLBCL cases was performed using Affymetrix HG-U133Plus 2.0 arrays as part of an independent study (GEO database GSE12195)[1]. The probe sets used in Fig. S12 are 228177_at (CREBBP) and 202221_s_at (EP300).

Construction of mammalian expression vectors encoding for mutated HA-tagged CREBBP proteins. All CREBBP missense mutants were generated by a site-directed mutagenesis protocol using a previously described mammalian expression vector encoding the mouse Crebbp protein with a C-terminal HA tag as a template (pCIN4-CREBBP-HA, kind gift of Dr. W. Gu). The presence of the desired mutations and the integrity of the open reading frame where confirmed in all constructs by enzymatic digestion and full-length sequencing. In western blot analyses, the R1360X construct, which contains a stop codon N-terminal to the HA tag, can only be detected by the anti-CREBBP specific N-22 antibody.

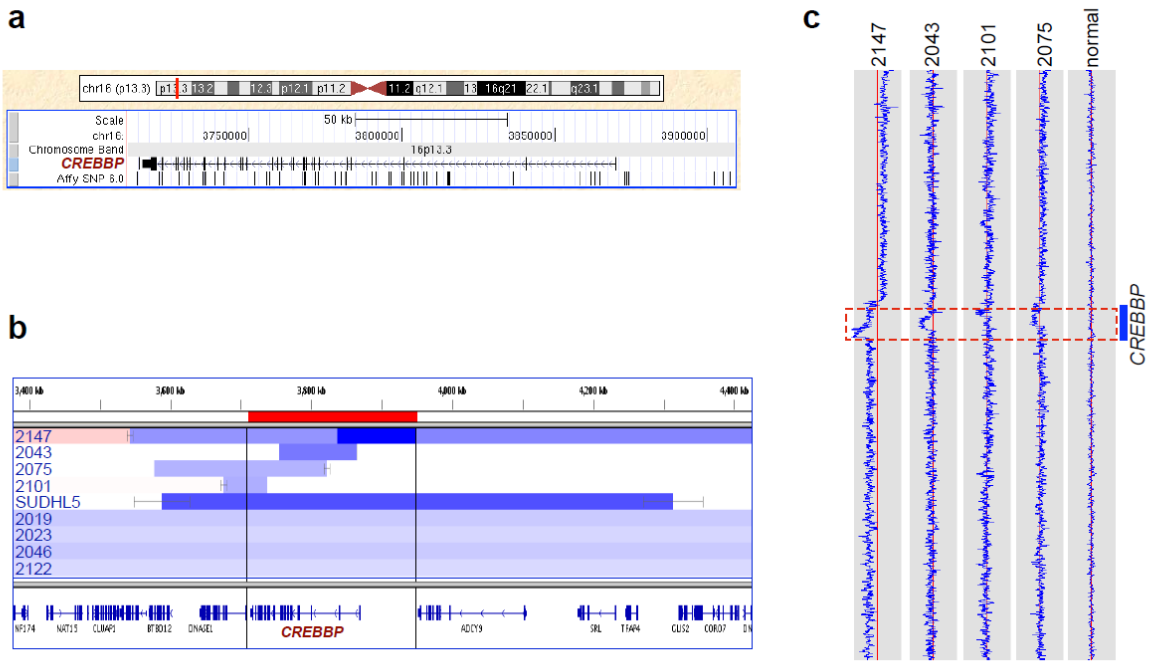
Modeling of CREBBP HAT domain missense mutations. The structural view of the CREBBP HAT domain was generated in the PyMOL v0.99 software (available at http://www.pymol.org) using the coordinates of the crystal structure of the EP300 HAT domain (85% identity with CREBBP) in complex with the Lys-CoA (MMDB ID 62423).

Immunofluorescence analysis of exogenous CREBBP proteins. For analysis of mutant CREBBP subcellular localization, HeLa cells were transiently transfected with plasmids encoding wild-type or mutant CREBBP-HA, harvested 30 hr after transfection, and used to prepare cytospins according to standard methods. After fixation in 10% buffered formalin for 15 minutes at room temperature, cells were permeabilized for 10 minutes in 0.2% Triton X-100 in PBS, blocked for 30 min in PBS-Tween with 3% bovine serum albumin (BSA), and incubated 90 min at room temperature in blocking buffer containing a 1:200 dilution of a FITC-conjugated anti-HA antibody (clone 11.1, Covance). Images were captured using the Nikon NES software coupled to a Nikon ES400 fluorescence microscope with a B/W camera, and artificially colored using Adobe Photoshop.

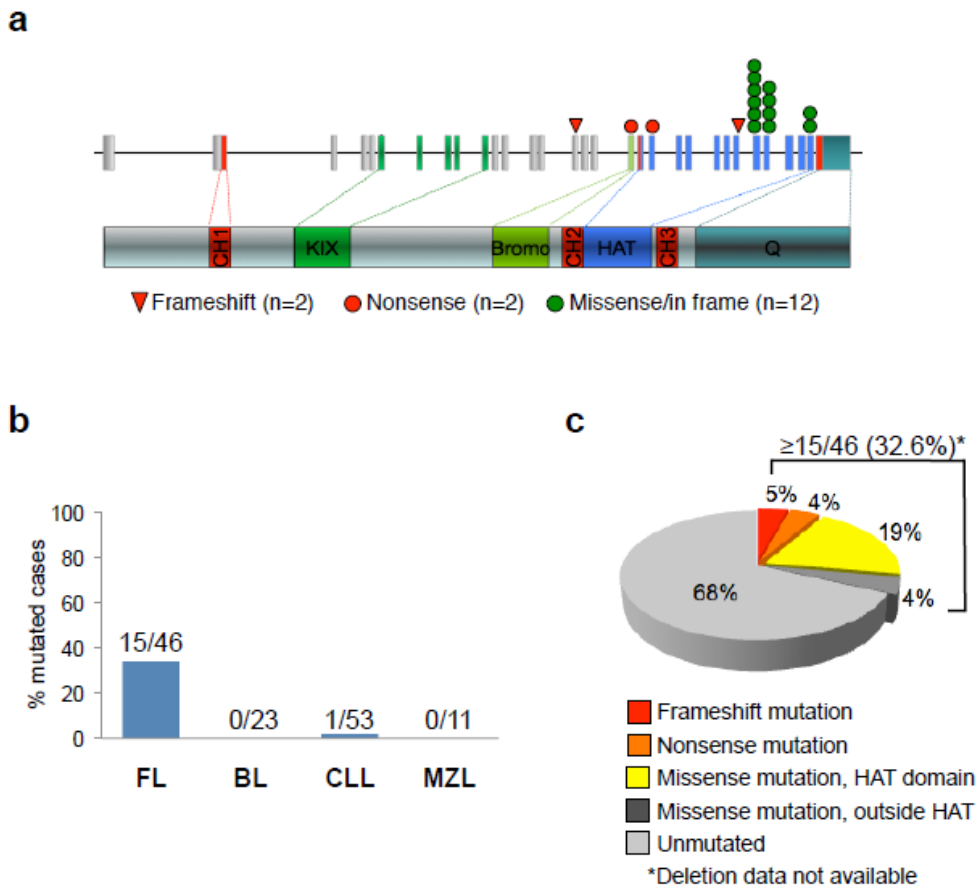
Isolation of recombinant GST-p53 and CREBBP-HA proteins. Rosetta *E.coli* bacteria containing a GST-p53 expression plasmid were kindly provided by Dr. W. Gu and the procedure used for purification of the recombinant protein is described in detail elsewhere[8]. Briefly, GST-p53 expression was induced with 1mM IPTG overnight at 23-25°C and the recombinant protein was obtained from bacteria cell lysates in BC500 using gluthatione-coated agarose beads, upon 90 minutes incubation at 4°C. After several washing steps, GST-p53 recombinant protein was eluted using BC100 buffer containing 20 mM gluthatione. The concentration of GST-p53 in the eluates was assessed by commassie staining upon SDS-

PAGE (Figure S8b). Recombinant CREBBP-HA proteins (WT and selected mutants) were isolated from transiently transfected HEK293T cells, 48 hr after transfection. Briefly, cells were harvested in cold PBS and lysed in BC500 (20mM Tris pH 7.5, 500 mM NaCl, 10% Glycerol, 0.2 mM EDTA, 1% Triton X-100, 1mM PMSF, 0.1 mM sodium orthovanadate and protease inhibitors). Cleared lysates were supplemented with 5M NaCl to a final concentration to 650 mM, and CREBBP-HA was immunoprecipitated overnight at 4°C using agarose-conjugated HA beads (Sigma). After several washes in BC500 and BC100, proteins were eluted twice in BC100 containing 500 µg/mL HA peptide (Sigma) at 4°C. The two fractions were pooled and resolved by SDS-PAGE, followed by commassie blue staining to assess protein purity and concentration (see Fig. S8a).

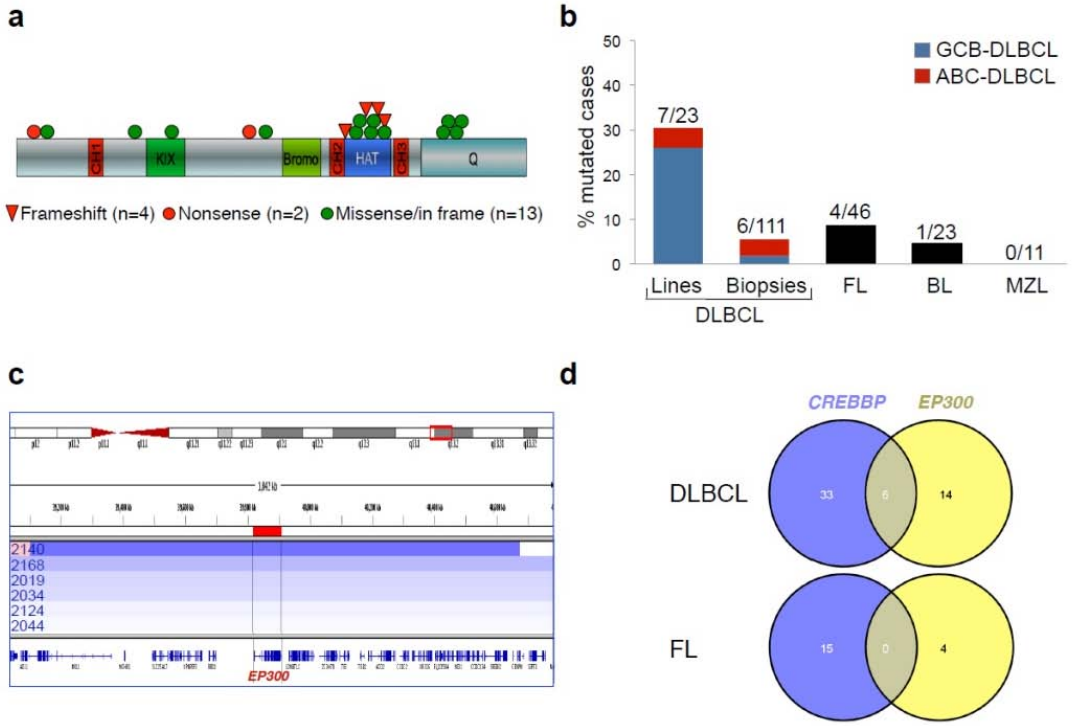
Supplementary Figures



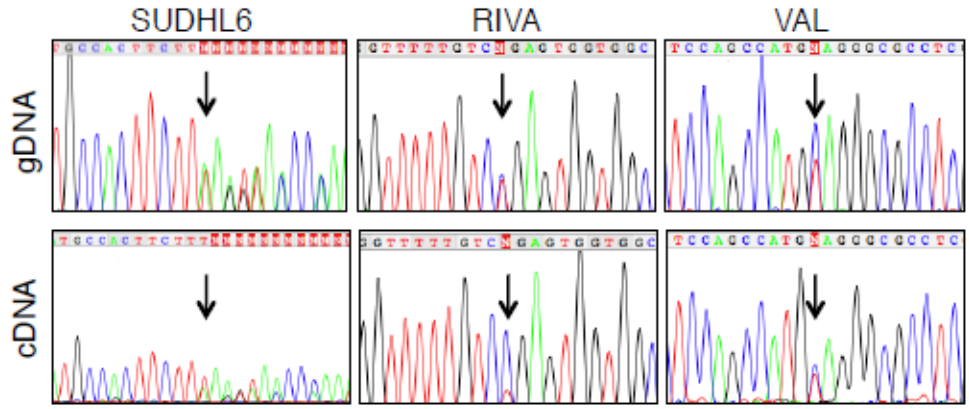
Supplementary Figure 1. The *CREBBP* locus is targeted by focal deletions in DLBCL. a, Distribution of SNP markers covering the *CREBBP* locus in the Affymetrix Genome-Wide Human SNP array 6.0, as obtained from the University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) using the NCBI36.1/hg18 annotation. **b,** Graphic representation of segmentation data from the 9 DLBCL samples carrying CREBBP deletions, visualized using the Integrative Genomics Viewer software (http://www.broadinstitute.org/igv). Each track represents one sample, where white denotes a normal (diploid) copy number, blue indicates a region of copy number loss, and red corresponds to a region of copy number gain (data range for minimum, baseline and maximum value: -1.5, 0, 1.5). Note that, due to the presence of non-tumor cells infiltrating the biopsy, the inferred copy number, and the corresponding color intensity, may vary across samples (see actual values in Table S2). Individual genes in the region are aligned in the bottom panel, and the boxed area (defined by the red bar at the top) corresponds to the *CREBBP* locus. Numbers indicate the genomic coordinates of the displayed region on chromosome 16p. **c,** Copy number plots of the four DLBCL samples showing focal deletions encompassing the *CREBBP* locus (boxed area), as compared to a normal diploid DNA. In each plot, a red line denotes a baseline level of 2.



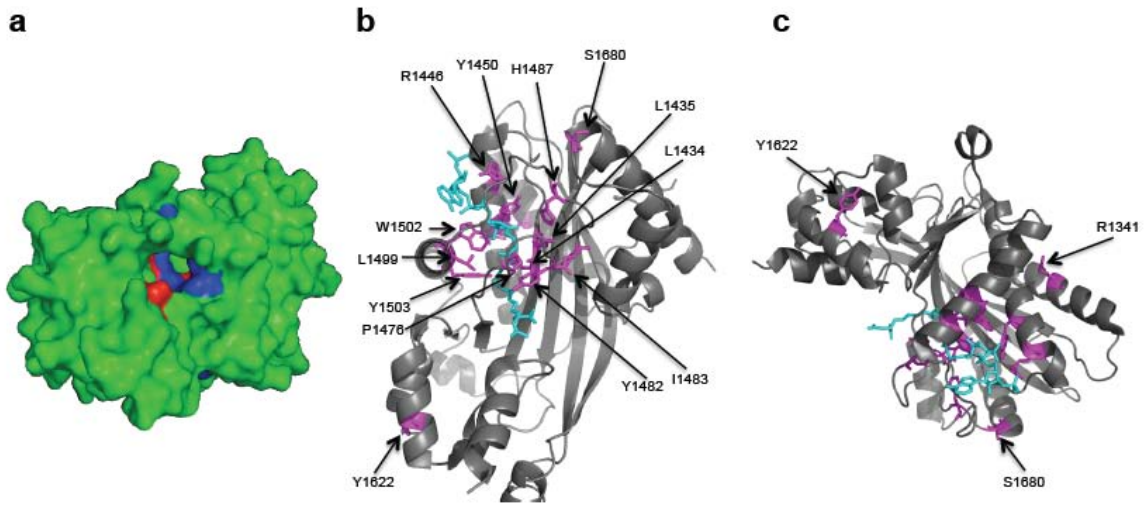
Supplementary Figure 2. *CREBBP* is mutated in a large fraction of FL cases. a, Distribution of FL-associated point mutations along the *CREBBP* gene and protein, graphically represented as described in Fig. 1a. **b,** Prevalence of *CREBBP* mutated cases in various B-NHL types. **c,** Overall frequency of *CREBBP* mutations, according to mutation type.



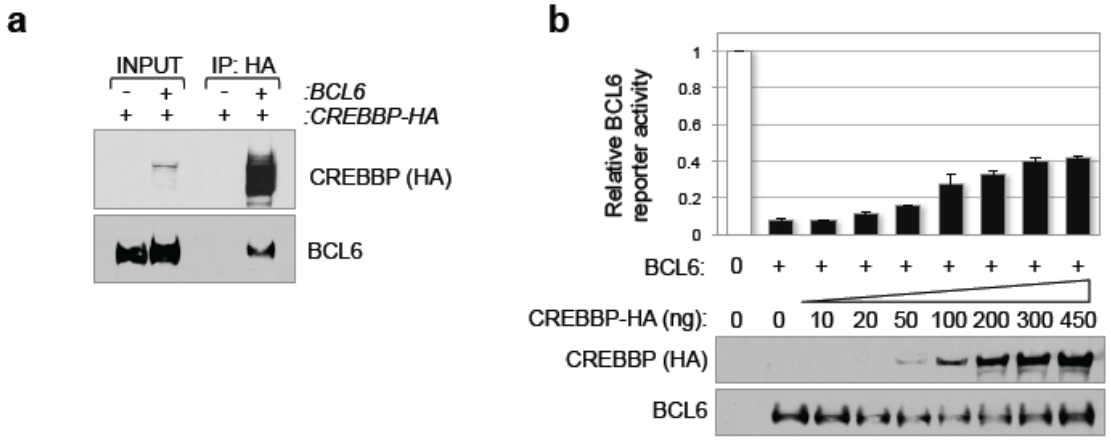
Supplementary Figure 3. The *EP300* gene is mutated in a small subset of DLBCL and FL. a, Schematic representation of the EP300 protein, with its key functional domains. Color-coded symbols indicate distinct types of mutational events, and the corresponding number is given in brackets. **b,** Overall frequency of *EP300* mutated samples in various B-NHL, including DLBCL cell lines, DLBCL primary biopsies, FL, BL, and MZL. In the DLBCL bar graph, red and blue colors are used to indicate the phenotypic subtype of the affected cases. **c,** Graphic representation of segmentation data from 6 DLBCL primary biopsies carrying *EP300* deletions, visualized as described in Fig. S1b. Individual genes in the region are aligned in the bottom panel, and numbers indicate the genomic coordinates of the displayed region on chromosome 22q13.2. Two additional cell lines carrying aberrant *EP300* alleles do not appear in the plot since the loss of *EP300* was documented by different methods. These include SUDHL2, where deletion of one allele was documented by FISH analysis, and RC-K8, which has been reported in the literature[9] (see also Table S4). **d,** Venn diagram showing the overlap between cases carrying genomic lesions of *CREBBP* and *EP300* in DLBCL (top) and FL (bottom). Note that four of the six DLBCL samples harboring alterations in both genes are represented by cell lines. Diagrams were generated by using the interactive tool VENNY at http://bioinfogp.cnb.csic.es/tools/venny/index.html.



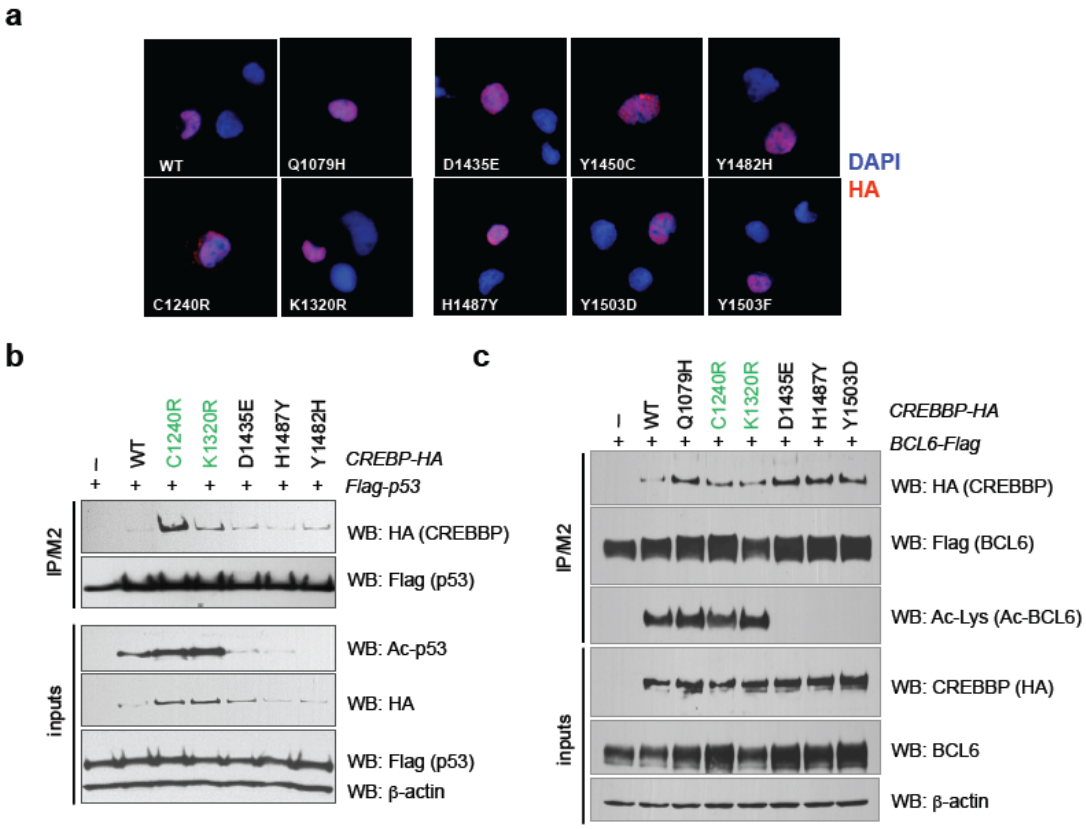
Supplementary Figure 4. The residual wild-type *CREBBP* allele is expressed in mutated cases, suggesting haploinsufficiency. Chromatograms of PCR products amplified from genomic DNA (gDNA, top) and cDNA (bottom) of representative DLBCL cell lines harboring *CREBBP* mutations. The presence of a double peak in the cDNA sequence (arrows) documents expression of both wild-type and mutated allele.



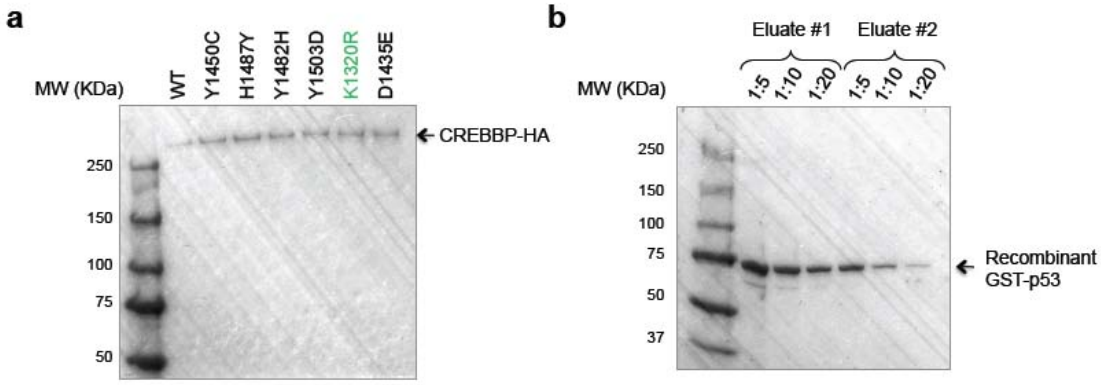
Supplementary Figure 5. Most CREBBP missense mutations cluster in the HAT domain, proximal to the lysine acetylation reaction center. a, Crystal structure of the EP300 HAT domain (85% identity with CREBBP) in complex with the Lys-CoA (MMDB ID 62423). Residues targeted by somatic point mutations in DLBCL and FL are highlighted in blue (or red if recurrently mutated). Note their clustering around the pocket that accommodates the Lys-CoA bisubstrate, mimicking the lysine side chain from native protein substrates, in close proximity to Ac-CoA[10]. **b,** Ribbon diagram of the HAT domain (front view). Lys-CoA is shown in aquamarine stick figure representation, while the mutated residues are color-coded in pink. **c,** Ribbon diagram of the HAT domain (back view), with the Lys-CoA represented in aquamarine, and the mutated residues color-coded in pink.



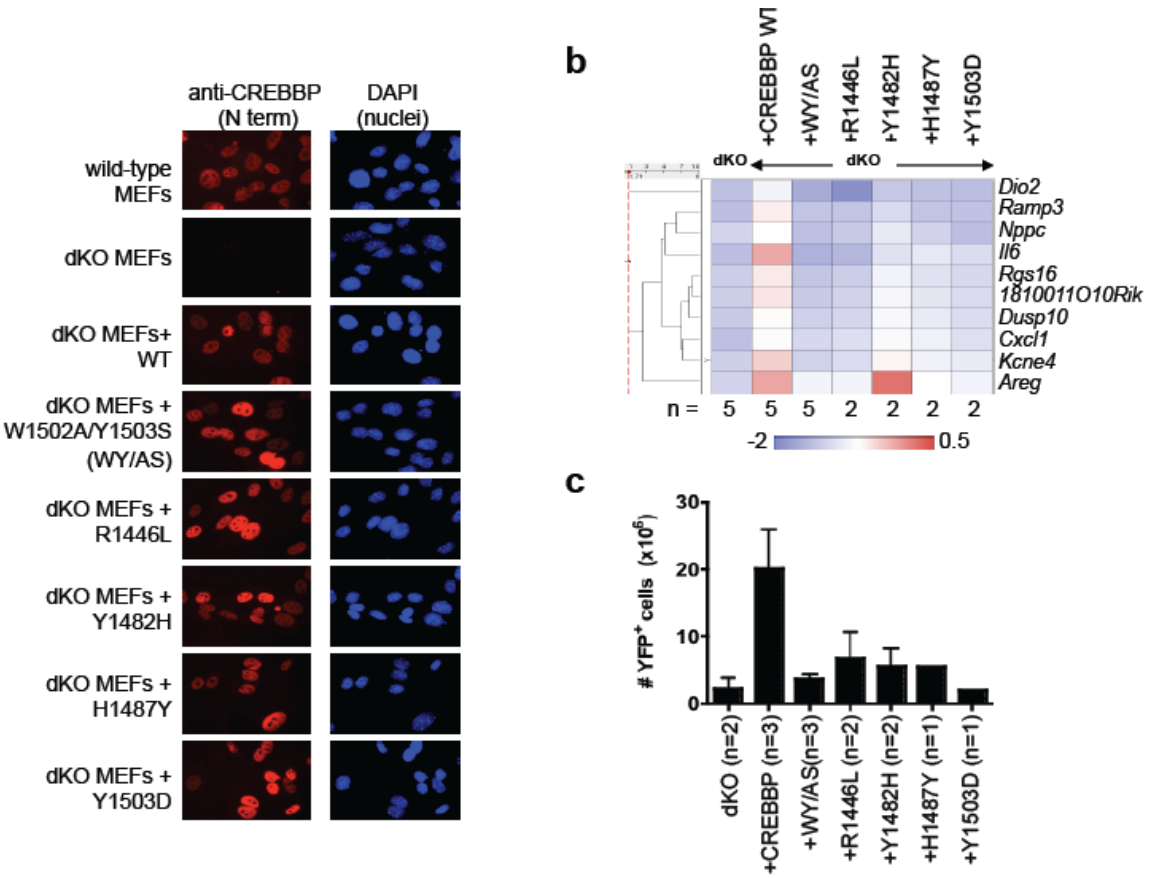
Supplementary Figure 6. CREBBP acetylates and inactivates the BCL6 transcriptional repressor. a, Co-immunoprecipitation of CREBBP-HA and BCL6 in transfected HEK293T cells confirms their physical interaction, as previously shown for EP300[11]. **b,** Dose dependent response of a 5X-BCL6 luciferase reporter construct to BCL6, in the absence or presence of increasing CREBBP-HA amounts. Results are shown as relative activity compared to the basal reporter activity, set as 1 (white bar) (mean +/- standard deviation, as obtained from one of two independent experiments, performed in duplicate). Western blot analysis using anti-HA and anti-BCL6 antibodies monitors for the relative expression levels of exogenous CREBBP and BCL6 in the same lysates (bottom panels).



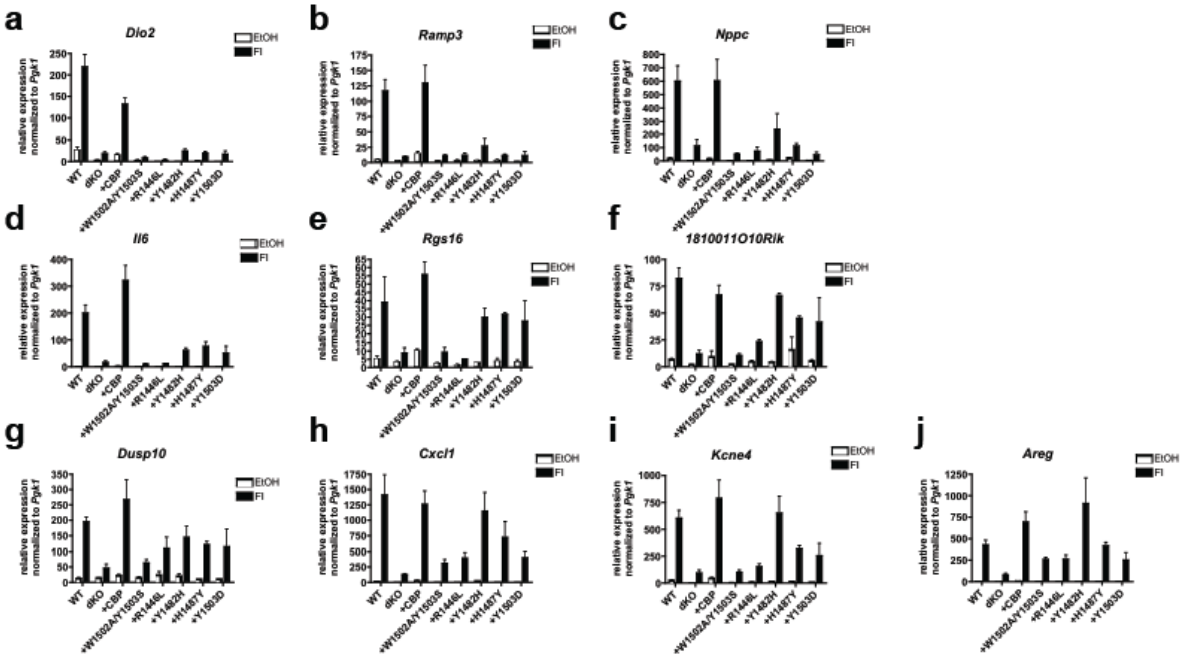
Supplementary Figure 7. Failure of CREBBP missense mutants to acetylate BCL6 and p53 is not due to loss of physical interaction. a, Subcellular localization of wild-type and mutant CREBBP proteins in HeLa cells, transiently transfected with expression plasmids encoding for the indicated HA-tagged mutants and analyzed by immuofluorescence using specific antibodies directed against the HA tag. Nuclei are counterstained with DAPI, and the signal corresponding to HA is artificially colored in red. **b,** Western blot analysis of whole cell extracts from HEK293T cells co-transfected with Flag-p53 and the indicated CREBBP-HA derivatives, before (inputs) and after immunoprecipitation with Flag/M2 antibodies. **c,** HEK293T cells were transiently transfected with plasmids expressing Flag-tagged human BCL6 and HA-tagged mouse CREBBP proteins (wild-type vs the indicated DLBCL-derived point mutants), and the presence of exogenous CREBBP in the complex was measured by western blot analysis using anti-HA antibodies after immunoprecipitation with Flag/M2 beads (top panel). Western blot analysis with anti-BCL6 antibodies in the total protein extracts before immunoprecipitation (inputs) controls for the expression levels of exogenous BCL6, and β-actin is used as protein loading control. The acetylation status of BCL6 was assessed in the same Flag immunoprecipitates by using anti-acetyl-lysine specific antibodies. Note that the amounts of pCMV-Flag-BCL6 and pCIN4-CREBBP-HA DNAs were adjusted in the transfection in order to obtain comparable expression levels for each of the proteins across the samples. Numbering of the mutated residues is based on the human CREBBP amino acid sequence.



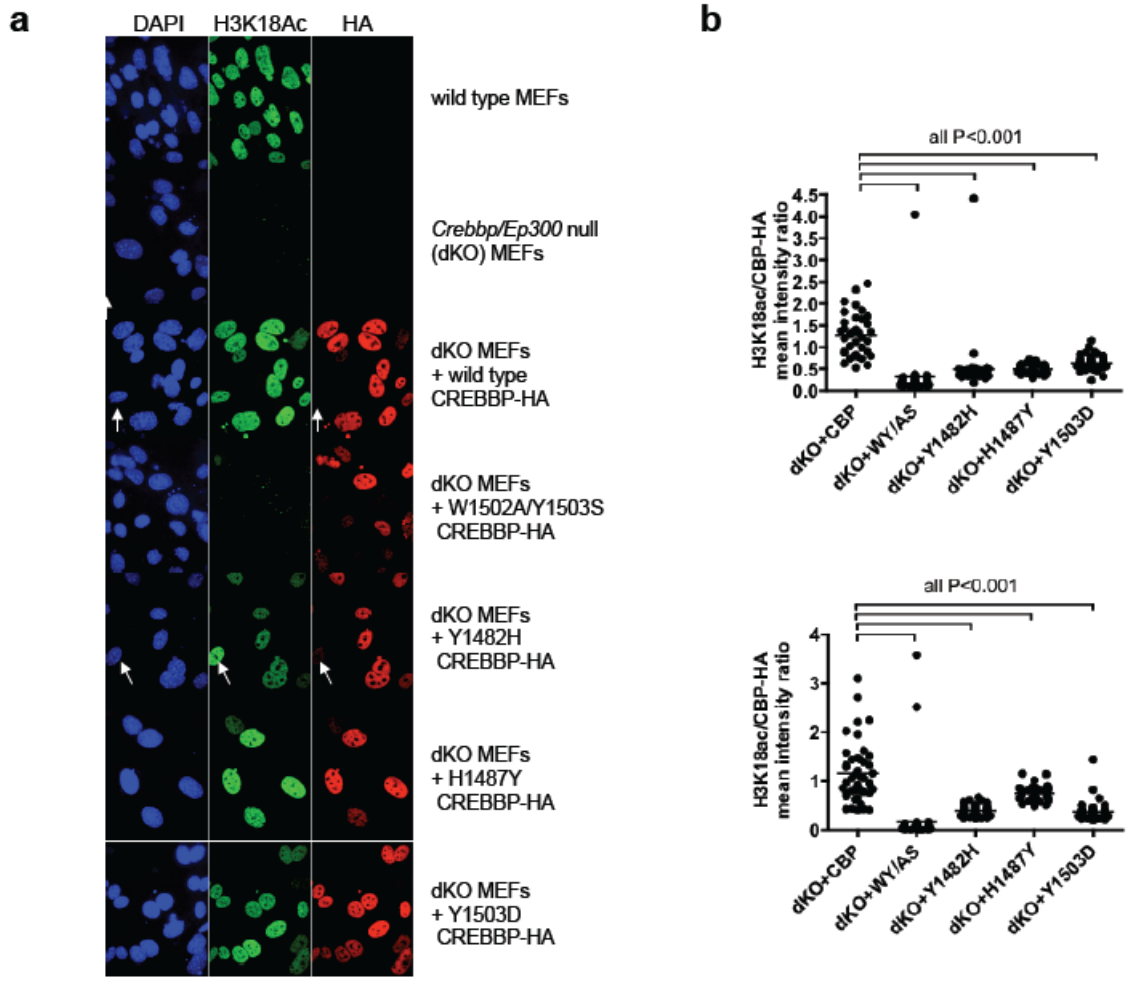
Supplementary Figure 8. Coomassie blue staining of purified recombinant CREBBP-HA (a) and GST-p53 (b) proteins used in the *in vitro* acetylation assays. In (b), each lane corresponds to increasing dilutions of the original stock eluate.



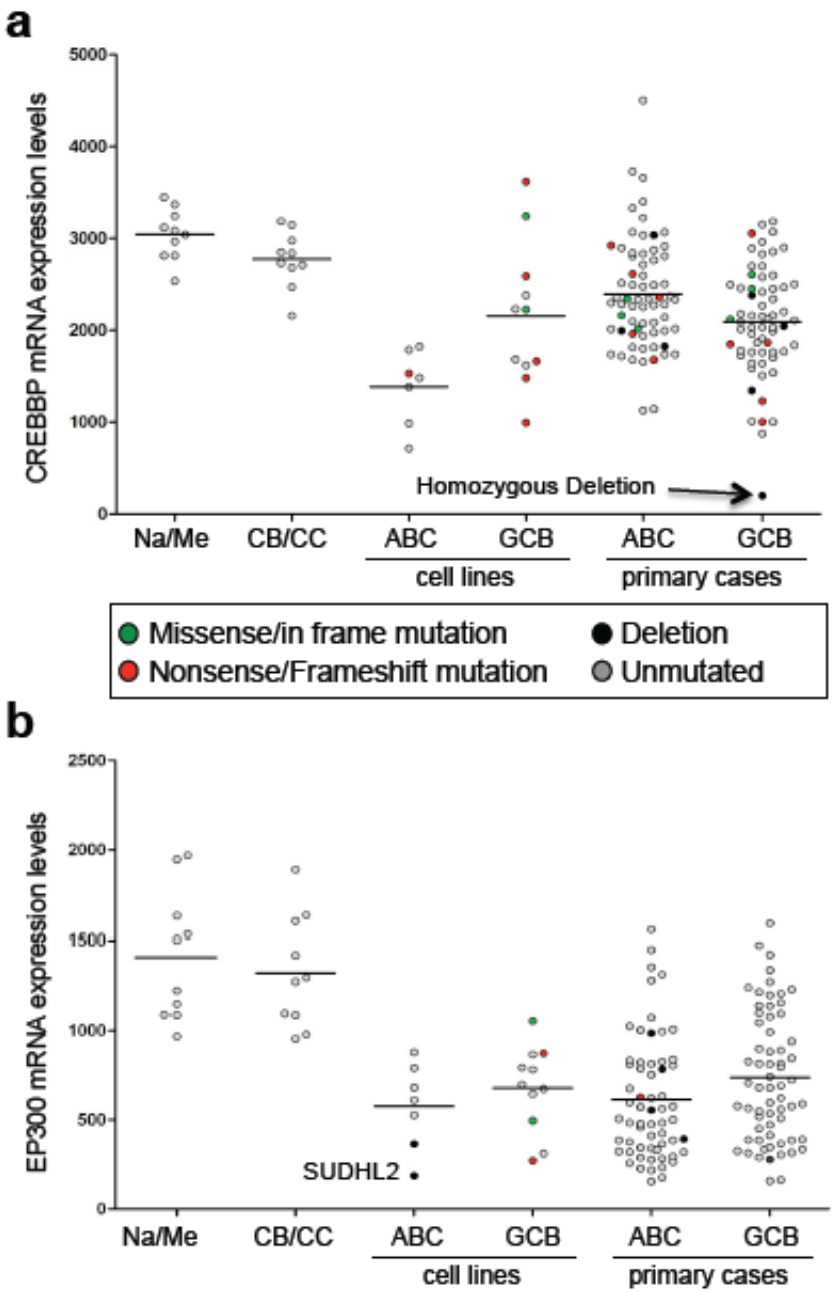
Supplementary Figure 9. DLBCL-associated CREBBP mutations have variable effects on cAMP-dependent gene expression and are incapable of rescuing the growth phenotype of conditional *Crebbp*/*Ep300* double null MEFs. a, Immunofluorescence analysis of CREBBP protein expression in wild-type and *Crebbp/Ep300* null (dKO) MEFs transduced with retroviral vectors expressing various HA-tagged CREBBP proteins. DAPI identifies the nuclei. **b,** Heat map of hierarchically clustered qRT-PCR gene expression data in *Crebbp/Ep300* dKO MEFs reconstituted with retroviruses expressing wild-type CREBBP (+CREBBP) or various CREBBP mutants (as indicated), and treated with forskolin + IBMX (FI) for 90 min (n=2-5). Data are expressed as log ratio of the wild-type MEF signal. All four DLBCL-derived CREBBP mutants tested were generally deficient for cAMP-responsive transcription, as was the W1502A/Y1503S HAT-dead mutation[12] (used as negative control), although individual CREB target genes displayed different levels of dependence on HAT activity. **c,** Proliferative capacity of conditional *Crebbp/Ep300* dKO MEFs reconstituted with retroviruses expressing wild-type (WT) CREBBP or the indicated CREBBP mutant alleles. From each reconstituted population, equivalent numbers of YFP^+ (dKO) MEFs were seeded at day 1, and the total number of YFP^+ cells was measured on Day 11 by flow cytometry, as described in the methods section. Data are expressed as a percentage of the total cell number (n=1-3; mean+/- S.E.M.).



Supplementary Figure 10. Quantitative RT-PCR analysis of CREB target gene expression. a-j, qRT-PCR gene expression data from wild-type MEFs, untransduced *Crebbp*/*Ep300* dKO MEFs, and dKO MEFs reconstituted with retroviruses expressing HA-tagged wild-type (+CREBBP) or mutant CREBBP, and treated for 90 min with forskolin+IBMX (FI) or Ethanol vehicle (EtOH). The W1502A/Y1503S HAT-dead mutation is included as control[12]. Gene expression of individual genes was normalized to expression of Pgk1, and is shown as mean values ± SEM, as obtained from 2-5 independent experiments.



Supplementary Figure 11. DLBCL-associated mutations of *CREBBP* affect H3K18 acetylation. a, Immunofluorescence analysis of histone H3 lysine 18 acetylation (H3K18Ac) in wild type and dKO MEFs, non-transduced or transduced with retrovirus expressing HA-tagged CREBBP proteins (wild type and the indicated mutants). White arrows point to cells that did not delete endogenous *Crebbp* and *Ep300*, and do not express the HA-tagged CREBBP construct, thus serving as a reference for endogenous levels of H3K18Ac. **b,** Signal intensity ratio of H3K18Ac versus exogenous CREBBP (HA) for individual nuclei from the same populations shown in (a) (mean +/- SEM; n=34-65). Mutated CREBBP proteins showed significantly reduced ability to induce H3K18Ac, independent of retrovirus transduction efficiency (Tukey's post-hoc test of one-way ANOVA for the pairs indicated).



Supplementary Figure 12. Expression of CREBBP and EP300 mRNA in normal and transformed human B cells. a, b, CREBBP and EP300 mRNA expression levels in purified normal B cell subpopulations and DLBCL samples (cell lines and primary biopsies), as measured by Affymetrix U133p2 gene expression profile analysis. Data are expressed as absolute probe intensity values after MAS5 normalization. Each dot corresponds to an individual sample, color-coded based on the presence or absence of distinct genetic lesions. CB/CC, GC-derived centroblasts and centrocytes; Na/Me, non-GC naïve and memory B cells (each purified from 5 different individuals); ABC, activated B cell type; GCB, germinal center B cell type. Arrow points to the homozygously deleted DLBCL case 2147. In the SUDHL2 cell line, northern blot analysis documented lack of EP300 mRNA expression (see Fig. 3d).

Tables

Table S1. Mutations of the *CREBBP* gene in B-NHL

Sample ID	Diagnosis	Exon	Nucleotide change*	AA change*	Affected Domain	HAT activity^^
Missense/In-frame Mutations						
2108	ABC-DLBCL	E16	C3362T	P1053L	interdomain region	retained
FARAGE	GCB-DLBCL	E16	G3441T	Q1079H	interdomain region	retained
Ly8	GCB-DLBCL	E20	T3922C	C1240R	PHD domain	retained
2040^	GCB-DLBCL	E23	A4163G	K1320R	HAT domain	retained
2026	ABC-DLBCL	E24	G4226A	R1341Q	HAT domain	retained
2062	GCB-DLBCL	E26	T4509A	D1435E	HAT domain	lost
2126^	GCB-DLBCL	E26	G4541T (+/+)	R1446L	HAT domain	lost
SUDHL7	GCB-DLBCL	E26	A4553G	Y1450C	HAT domain	lost
34t	GC-DLBCL	E27	C4631G	P1476R	HAT domain	nd
2134	GCB-DLBCL	E27	T4648C	Y1482H	HAT domain	lost
WSU	GCB-DLBCL	E27	C4663T	H1487Y	HAT domain	lost
2015^	ABC-DLBCL	E27	T4711G	Y1503D	HAT domain	lost
33t^	GC-DLBCL	E27	T4711C (+/+)	Y1503H	HAT domain	nd
2109^	GCB-DLBCL	E27	A4712T	Y1503F	HAT domain	lost
SUDHL7	GCB-DLBCL	E30	ΔTCC (5239-5241)	ΔS1680#	HAT domain	lost
2109^	GCB-DLBCL	E30	ΔTCC (5239-5241)	ΔS1680#	HAT domain	lost
2191	GC-DLBCL	E30	ΔTCC (5239-5241)	ΔS1680#	HAT domain	lost
FL22^	FL	E26	A4508G	D1435G	HAT domain	nd
BOR	FL	E26	G4541A	R1446H	HAT domain	lost
FL15^	FL	E26	G4541T	R1446H	HAT domain	lost
FL28^	FL	E26	C4540T	R1446C	HAT domain	nd
FL6^	FL	E26	C4540T	R1446C	HAT domain	nd
FL8	FL	E26	T4505C	L1434P	HAT domain	nd
FL25	FL	E27	T4652G	I1483S	HAT domain	nd
1518^	FL	E27	T4700C	L1499P	HAT domain	nd
BOR	FL	E27	G4710T	W1502C	HAT domain	nd
1647	FL	E27	A4712T	Y1503F	HAT domain	lost
1514	FL	E30	ΔTCC (5239-5241)	ΔS1680#	HAT domain	lost
FL26^	FL	E30	ΔTCC (5239-5241)	ΔS1680#	HAT domain	lost
CLL44^	B-CLL	E29	A5069G	Y1622C	HAT domain	nd
Nonsense Mutations						
2209^	non-GC-DLBCL	E2	C427T	R75X	KIX, BD, HAT, IBiD domain	lost
2152^	ABC-DLBCL	E2	C490T	Q96X	KIX, BD, HAT, IBiD domain	lost
2161	GCB-DLBCL	E2	C670T	Q156X	KIX, BD, HAT, IBiD domain	lost
VAL	GCB-DLBCL	E2	C898T	Q232X	KIX, BD, HAT, IBiD domain	lost
2179	GCB-DLBCL	E12	C2458T	Q752X#	KIX, BD, HAT, IBiD domain	lost
2106	GCB-DLBCL	E18	C3721T	R1173X#	BD, HAT, IBiD domain	lost
2040^	GCB-DLBCL	E23	A4171T	K1323X	HAT, IBiD domain	lost
2020	GCB-DLBCL	E24	C4225T	R1341X	HAT, IBiD domain	lost
RIVA	ABC-DLBCL	E24	C4282T	R1360X	HAT, IBiD domain	lost
FL7	FL	E18	T3579A	Y1125X#	BD, HAT, IBiD domain	lost
FL17	FL	E20	T3906G	Y1234X	BD, HAT, IBiD domain	lost
Frameshift/Splice Site Mutations						
2183	ABC-DLBCL	E6	dupl CCAT (1535-1538)	L446fs	KIX, BD, HAT, IBiD domain	lost
SUDHL8	GCB-DLBCL	I6	T(+2)C	L525fs	KIX, BD, HAT, IBiD domain	lost
SUDHL6	GCB-DLBCL	E6	ΔT (1611)	L470X	KIX, BD, HAT, IBiD domain	lost
Ly1	GCB-DLBCL	E8	ΔC (1922)	T573fs	KIX, BD, HAT, IBiD domain	lost
Ly8	GCB-DLBCL	E17	ΔCT (3505-3506)	L1101fs	BD, HAT, IBiD domain	lost
2170	ABC-DLBCL	E17	dupl 40bp (3465-3505)	L1101fs	BD, HAT, IBiD domain	lost
2069^	ABC-DLBCL	E28	dupl GCACT (4907-4911)	S1572fs	HAT domain	lost
SUDHL10	GCB-DLBCL	E31	ΔC (6035)	A1944fs	IBiD domain	nd
1546	FL	E15	A3263T (splice site)	E1020fs	BD, HAT, IBiD domain	lost
1536^	FL	E25	ΔC (4428)	C1408fs	HAT domain	lost

*Numbering according to GenBank accession No. NM_004380.2 (mRNA) and NP_004371.2 (protein) respectively.

These mutations were also reported in RTS patients (see Roelfsema and Peters, 2007; Schorry et al, 2008; and Leiden Open Variation Database at http://chromium.liacs.nl/LOVD2/home.php).

^ For these samples, paired normal DNA was available and confirmed the somatic origin of the mutation.

^^ as assessed by in vivo acetylation assays (missense mutations) and/or by prediction (nonsense mutations)

Abbreviations: ABC, activated B cell type; GCB, germinal center B cell type; NC, unclassified; non-GC and GC, non germinal center and germinal center type, respectively (based on IHC); FL, follicular lymphoma; B-CLL, B-cell chronic lymphocytic leukemia; Δ, deletion; i, intron; fs, frameshift; dupl, duplication; +/+, hemizygous mutation; nd, not determined.

KIX, CREB-binding; BD, bromodomain; PHD, plant homeodomain; HAT, histone acetyltransferase;IBiD, IRF3-binding

Table S2. SNP array copy number analysis of the *CREBBP* gene in DLBCL

Sample ID	DLBCL Subtype	Chr	Cytoband	Start position*	End position*	Mean CN	Size (Kb)	Aberration	No.of Genes	Annotated genes in the region**
2101	non-GC	16	16p13.3	3681563	3738462	1.48	56.9	loss	2	CREBBP, TRAP1
2043	ABC	16	16p13.3	3756551	3864655	1.21	108.1	loss	1	CREBBP
2147	GCB	16	16p13.3	3838219	3949571	0.65	111.4	loss	1	CREBBP
2147	GCB	16	16p13.3	3547655	3836389	1.35	288.7	loss	5	BTBD12, CREBBP, DNASE1, NLRC3, TRAP1
2075	GCB	16	16p13.3	3580275	3818710	1.51	238.4	loss	4	BTBD12, CREBBP, DNASE1, TRAP1
SUDHL5^	GCB	16	16p13.3	3636366	4282743	1.05	646.4	loss	7	BTBD12, CREBBP, DNASE1, TRAP1, ADCY9, SRL, TFAP4
2023	GCB	16	16p13.3	765	5547879	1.80	5547.1	loss	206	CREBBP and >200 other genes
2046	GCB/NC	16	16p13.3	765	5547879	1.78	5547.1	loss	206	CREBBP and >200 other genes
2019	ABC	16	16p13.3 - 16q21	26113	62521051	1.67	62494.9	loss	525	CREBBP and >200 other genes
2122	GC	16	16p13.3 - 16q24.3	765	88815024	1.64	88814.3	loss (entire chromosome)	776	CREBBP and >200 other genes

* Numbering according to NCBI Build 36.1. Position 765 corresponds to the first marker on chromosome 16 in the Affymetrix SNP6.0 array.
**NCBI RefSeq database.
Abbreviations: ABC, activated B cell type; GCB, germinal center B cell type; NC, unclassified; non-GC and GC, non germinal center and germinal center type, respectively (based on immunohistochemistry).
^ Data for the SUDHL5 cell line were obtained from the Tumorscape database at http://www.broadinstitute.org/tumorscape, and were confirmed by FISH analysis using BAC clone RP11-292B10 and a control centromere 16 probe.

Table S3. Mutations of the *EP300* gene in B-NHL

Sample ID	Diagnosis	Exon	Nucleotide change*	AA change*	Affected Domain
Missense/In-frame Mutations					
2180	ABC-DLBCL	E2	A771G	M126V	none
2161	GCB-DLBCL	E6	A1914G	S507G	none
2141	ABC-DLBCL	E9	G2163C	A590P	KIX domain
2025^	ABC-DLBCL	E14	C3168A	P925T	none
FL1	FL	E26	T4584C	H1397Y	HAT domain
DB	GCB-DLBCL	E28	G4912T	G1506V	HAT domain
SUDHL8	GCB-DLBCL	E29	A5105T	K1570N	HAT domain
SUDHL10	GCB-DLBCL	E30	G5191A	R1599H	HAT domain
SUDHL6	GCB-DLBCL	E30	C5274T	R1627W	HAT domain
2041	NC-DLBCL	E31	A6106C	Q1904P	none
FL14	FL	E31	A6918C	M2175L	none
2135	ABC-DLBCL	E31	ΔCAGCAGCAACAG (6969-6980)	Δ4Q (2192-2195)	none
FL19	FL	E31	Δ12bp (7022-7033)	ΔQFQQ (2210-2213)	none
Nonsense Mutations					
RAMOS	BL	E2	C582T	Q63X	KIX, BD, HAT, IBiD domain
SUDHL2	ABC-DLBCL	E14	C2856T (+/+)	Q821X	KIX, BD, HAT, IBiD domain
Frameshift/Splice Site Mutations					
BJAB	GCB-DLBCL	E24/I24	G(i24+1)A	ΔE24 (R1292fs)**	HAT, IBiD domain
FARAGE	GCB-DLBCL	E27	ΔA4803	M1470fs	HAT, IBiD domain
SUDHL8	GCB-DLBCL	E29	ΔA5031	K1547fs	HAT, IBiD domain
1658	FL	E28/I28	G(i28+1)A	V1540fs	HAT, IBiD domain

*Numbering according to GenBank accession No NM_001429.3 (mRNA) and NP_ 001420.2 (protein) respectively.
Note that none of these mutations has been previously observed in RTS patients (see Roelfsema and Peters, 2007; Bartsch et al, 2009; and Leiden Open Variants Database at http://chromium.liacs.nl/LOVD2/home.php).
** as verified by cDNA amplification followed by both direct sequencing and sequencing of individual clones.
^ For this sample, paired normal DNA was available and confirmed the somatic origin of the mutation. One additional sequence variant identified in 4 patients (G1026A, corresponding to AA change G211S) most likely represents a germline polymorphism since, although no matched normal tissue was available for these patients, the same change was present in normal DNA in a separate study (Ref. 34).
Abbreviations: ABC, activated B cell type; GCB, germinal center B cell type; NC, unclassified; FL, follicular lymphoma; BL, Burkitt lymphoma; Δ, deletion; I, intron; fs, frameshift; +/+, hemyzigous mutation

Table S4. SNP Array copy number analysis of the *EP300* gene in DLBCL

Sample ID^	DLBCL Subtype	Chr	Cytoband	Start position*	End position*	Mean CN	Size (Kb)	Aberration	No.of Genes	Annotated genes in the region**
2140	GC	22	22q13.1 - 22q13.2	39102833	40674243	1.10	1571.41	loss	28	EP300 and 27 other genes
2168	ABC	22	22q11.21 - 22q13.33	20175343	49535360	1.60	29360.02	loss	352	EP300 and >200 other genes
2019	ABC	22	22q11.1 - q13.33	14432516	49581309	1.64	35148.79	loss (entire chromosome)	416	EP300 and >200 other genes
2034	ABC	22	22q11.1 - 22q13.33	14805205	49518684	1.44	34713.48	loss (entire chromosome)	412	EP300 and >200 other genes
2124	GCB	22	22q11.1 - 22q13.33	14432516	49581309	1.81	35148.79	loss (entire chromosome)	416	EP300 and >200 other genes
2044	ABC	22	22q11.1 - 22q13.33	14857504	49581309	1.82	34723.81	loss (entire chromosome)	414	EP300 and >200 other genes

^ Two additional cell lines carrying aberrant *EP300* alleles include SUHDL2, which showed a monoallelic gene deletion by FISH analysis using a labeled probe for the *EP300* locus (BAC RP11-1078O11) and a control probe for chromosome 22, and RC-K8 (Garbati et al., Ref. 30 of the manuscript)

* Numbering according to NCBI Build 36.1.

**NCBI RefSeq database.

Abbreviations: ABC, activated B cell type; GCB, germinal center B cell type; GC, germinal center type (based on immunohistochemistry).

Table S5. Oligonucleotides used for genomic amplification of the *CREBBP* gene

Exon	Oligonucleotide name	Position^	Oligonucleotide Sequence (5'-3')
Exon 1	CREBBP-E1F1	13*	GTTGCTGTGGCTGAGATTTG
	CREBBP-E1R1	+463	GACAGTCCTGCGACGAACTT
Exon 2-I	CREBBP-E2F1	+781	ACTTTCCAACTGCCTGACCTG
	CREBBP-E2R1	743*	ACTTTAACCAGACCCACCCAG
Exon 2-II	CREBBP-E2F2	943*	AAACCTGCGTTAGGGTCTCAG
	CREBBP-E2R2	368*	ACAGTGGGAACCTTGTTCCAG
Exon 2-III	CREBBP-E2F3	779*	TTACTATTGAGGAGGCCTGGG
	CREBBP-E2R3	-73	TCATAGAAACGTGGCAGTTGG
Exon 3	CREBBP-E3F	+200	CCTCACCTATCTCTCCTGTTGC
	CREBBP-E3R	-80	AACTGTGTGAGCATTTCCCAG
Exon 4	CREBBP-E4F	+323	TGAACCAGAGAGCTGCTGTAAG
	CREBBP-E4R	-107	TGGTCGGTATTATCCATCAGC
Exon 5	CREBBP-E5F	+202	CTGCCCACTCCCTACCTACTC
	CREBBP-E5R	-240	GCCCTTTATCACCCATCCTTC
Exon 6	CREBBP-E6F	+316	TTGGGTTCCATCACTCCATTC
	CREBBP-E6R	-149	TTCTCCCTAGGCTTGTTTACTGC
Exon 7	CREBBP-E7F	+327	AAGTCAGGAGAGCCAAGTGATG
	CREBBP-E7R	-101	CACCAATTGTTTATATGGTGGC
Exon 8	CREBBP-E8F	+249	GCACGTGACTTGTATAGGCTCC
	CREBBP-E8R	-124	ACATCACTTGGCTCTCCTGAC
Exon 9	CREBBP-E9F	+263	GGAAGTCTCCTTGGTCAGTGG
	CREBBP-E9R	-161	GCCAGGCAGATCTCAAACTG
Exon 10	CREBBP-E10F	+298	ACATCAACAGCTTCTGCAGGG
	CREBBP-E10R	-145	TGGGCTTCTGGTGTATCAGG
Exon 11	CREBBP-E11F	+148	GAAAGGGTTAGAAAGAAATATACAGGG
	CREBBP-E11R	-191	GGACACTGAGTTTCTCTCTTGGG
Exon 12	CREBBP-E12F	+208	CAAGTGACATGAATTCTGCTGC
	CREBBP-E12R	-112	TTCTGTTGCCTGTGCGTTC
Exon 13	CREBBP-E13F	+300	AGACATGAAATGTGCATTCTGG
	CREBBP-E13R	-78	GGATCATTCTGGCTCACCTTG
Exon 14	CREBBP-E14F1	+558	CAAGTGATTCTCCCACCTTGG
	CREBBP-E14R1	-11	ATTTCTGGTAGGGACAGGTGCC
Exon 15	CREBBP-E15F	+311	CCGAACCCACATTCAAACAG
	CREBBP-E15R	-83	ATTGTAGGTTGCATGAGCAGC
Exon 16	CREBBP-E16F	+350	CCTGAGTGTTTCTGCAGGG
	CREBBP-E16R	-83	AGGCTTGTAAGAGTCTTCCCG
Exon 17	CREBBP-E17F	+231	AGAACAATCTTCAAGGCAGGG
	CREBBP-E17R	-93	ATGTCACAGCAGCCAGGATTC
Exon 18	CREBBP-E18F	+328	CCAGTATACAGGCGTGGTCTC
	CREBBP-E18R	-131	TTCTTACTGTTGGGAATGGAAGT
Exon 19	CREBBP-E19F1	-163	TCACATGCTATCCCAAAATGTC
	CREBBP-E19R1	+176	TTCAGGAAAGAAATAATGTACA
Exon 20	CREBBP-E20F	+176	GGCACCGGTACCTTCCTTATAG
	CREBBP-E20R	-149	TTTCATTGTCCTCACTGACCC
Exon 21	CREBBP-E21F	+139	ACCCACAACCCACTCCATAAG
	CREBBP-E21R	-259	GGGCAACAGAGCAAGACC
Exon 22	CREBBP-E22F	+238	AAGCGGACAAACGCTTAGAAC
	CREBBP-E22R	-83	GTGGACGCACACACAGACTTC
Exon 23	CREBBP-E23F	+165	GAACCATGTGTTGAGAGGAACC

	CREBBP-E23R	-178	AAGTTCTAAGCGTTTGTCCGC
Exon 24	CREBBP-E24F1	-132	TTCCGTGTTTGGAGAACATT
	CREBBP-E24R1	+317	TAAAACCCCAGACAGGACAA
Exon 25	CREBBP-E25F	+221	GGACACTTAAGAGCCCTGGTC
	CREBBP-E25R	-225	CATTCACAGAGGTGCAGTTCC
Exon 26	CREBBP-E26F	+268	CACCTGGAAAGAGGAGCTTTG
	CREBBP-E26R	-78	CAGGGTGTTGTTTGTTGCTTG
Exon 27	CREBBP-E27F	+376	CTCCAACTGTGCTGCTCTCAG
	CREBBP-E27R	-96	TCCTGGCTTTAGTCCTTGCTC
Exon 28	CREBBP-E28F	+259	AGGACCTAACAGTCGACACGC
	CREBBP-E28R	-122	CACACATGCATGGGACTCTG
Exon 29	CREBBP-E29F	+268	ACTTCCCTCCCACCACAGAC
	CREBBP-E29R	-111	GTGACCTACTTTGGCCTGAGC
Exon 30	CREBBP-E30F	+389	CAGCCACCATCAGGTACAGAC
	CREBBP-E30R	-107	CTCAGCCACCTGCCTATTCTG
Exon 31-I	CREBBP-E31F12	5756*	CGGAGCTTGTGTTTGATGTTG
	CREBBP-E31R12	-198	GTCCATGATCCCATCTTGTCC
Exon 31-II	CREBBP-E31F11	5965*	AGCCAGCTGGTGACATGC
	CREBBP-E31R11	5581*	CCATCCTGCCAGAAGATGAAG
Exon 31-III	CREBBP-E31F10	6030*	CACCTGGCTGGTAGGCTTC
	CREBBP-E31R10	5731*	TGCCTCAACATCAAACACAAG
Exon 31-IV	CREBBP-E31F9	6230*	CTCACCTGGTTGGGTCGG
	CREBBP-E31R9	5842-*	CAGAGTCTGCCTTCTCCTACCTC
Exon 31-V	CREBBP-E31F8	6395*	GGTGAGATGCTCCTGGGTG
	CREBBP-E31R8	6006*	CACAGGGAAGCCTACCAGC
Exon 31-VI	CREBBP-E31F7	6509*	TTGATGAAAGCTGCCATTAGC
	CREBBP-E31R7	6113*	AGCACCTGTACCGGGTGAAC
Exon 31-VII	CREBBP-E31F6	6639*	CATGGCATTCAGGTTCTGC
	CREBBP-E31R6	6313*	CCTGTGATATCCATGCAGGC
Exon 31-VIII	CREBBP-E31F5	6792*	CTGCCTCCGTAACATTTCTCG
	CREBBP-E31R5	6400*	GCTCTGCAAGACCTGCTGC
Exon 31-IX	CREBBP-E31F4	6913*	GTCCTTGAGGCTGCTGGAAC
	CREBBP-E31R4	6521*	CCAAGTACGTGGCCAATCAG
Exon 31-X	CREBBP-E31F3	7234*	GCACCTGGTTACTAAGGGACG
	CREBBP-E31R3	6714*	CTTGAACATCATGAACCCAGG
Exon 31-XI	CREBBP-E31F2	7406*	CCCAAGTGTCCCTGATCTATG
	CREBBP-E31R2	7060*	AACATCCAGCAAGCCCTG
Exon 31-XII	CREBBP-E31F1	7655*	ATCCACCCTTCCATGGCTC
	CREBBP-E31R1	7269*	GTCCCAGCCTCCACATTCC

^ Numbers indicate the distance from the corresponding exon, except where indicated by an asterisk (primers annealing to cDNA sequences)

* Numbering according to GenBank accession No NM_004380.2

Table S6. Oligonucleotides used for genomic amplification of the *EP300* gene

Exon	Oligonucleotide name	Position^	Oligonucleotide Sequence (5'-3')
Exon 1	EP300-E1F	206*	TTTCTATCGAGTCCGCATCC
	EP300-E1R	+73	AATGGAAGAATAAAGGCGCAC
Exon 2	EP300-E2F	-111	CATGGAGTGAGGTTGGGAAAT
	EP300-E2R	+82	ACAATGTAAGGCAAACCCTCC
Exon 3	EP300-E3F	-151	TCACGTTGCCCAAGCTGTAG
	EP300-E3R	+111	GCTCCTAGTGGGTACAAATCCC
Exon 4	EP300-E4F	-148	TGGTTTATGCATTCCCTGTGT
	EP300-E4R	+130	CATACAGACACCATCACCACAA
Exon 5	EP300-E5F	-201	CTTTGTGCAAATTGCTTACCC
	EP300-E5R	+100	CAACACCACAGGTCCCTCAC
Exon 6	EP300-E6F	-135	TGGAAGACAAGATCCACATACTC
	EP300-E6R	+189	GGAAGATACAAACCAGTGGCA
Exon 7	EP300-E7F	-171	GGTGCTTCAAATCAGCCTTTAC
	EP300-E7R	+106	AAAGACATCCTCAAACCGAGG
Exon 8	EP300-E8F	-136	CCCTGCCTAGCTCCTTAATGC
	EP300-E8R	+192	ATGATGGTGGGAAAGGTTGAG
Exon 9	EP300-E9F	-202	AGAGTCATTTCTTATATTGTGAACGG
	EP300-E9R	+100	CCGTCTGAAGCATATGTACTACTTG
Exon 10	EP300-E10F	-125	TGGCACCAGTTCTTAATGCAG
	EP300-E10R	+203	GTTAAGTACCATGCCCAGACATC
Exon 11	EP300-E11F	-207	CAACAAATCCACTTGGAGGC
	EP300-E11R	+84	TAAAGCGGGTGTTCAGGTAGC
Exon 12	EP300-E12F	-190	AGAATGGTGTGAACCCAGGTG
	EP300-E12R	+200	CACTGACACTCCAGGGACAAG
Exon 13	EP300-E13F	-90	GCCTGACGTTAGGAGCATTTG
	EP300-E13R	+95	ACCCACTATTTGCTGCCACTC
Exon 14	EP300-E14F	-169	TGTCCAAAGATACATGCCCAG
	EP300-E14R	+81	AGAGAATGGAAATGGCCCAG
Exons 15-16	EP300-E15F	-100	GGTGGCTAATTCTGCTATCCTG
	EP300-E16R	+102	CAATAATGGCAACTTCTGAGGC
Exon 17	EP300-E17F	-181	AACAAAGCGGGGCTTAGAAT
	EP300-E17R	+88	TGGCTATACTGTTTGGAATGTGA
Exon 18	EP300-E18F	-124	GGGAATATAGACAGGCCAGAAAC
	EP300-E18R	+91	CAGAAGCAGGATATTCTCTTATCCC
Exon 19	EP300-E19F	-140	TCAGGAACTGAATTAGCCCATC
	EP300-E19R	+86	CATGACTGTATGTGTGCTGGC
Exon 20	EP300-E20F	-72	GCTTCGTTGCTTGGCTTG
	EP300-E20R	+300	CCCATTGCTGACATATTCCC
Exon 21	EP300-E21F	-232	TTCCTGGGTTCTCCATTTCTG
	EP300-E21R	+88	TCATCTGATTGGTCATGCAAAC
Exon 22	EP300-E22F1	-177	CCACTCCAGCCTGTACAACA
	EP300-E22R1	+132	GCCAAACCCAAAGAAAACAA
Exon 23	EP300-E23F	-204	TTTCTTCATGTCTGTTGCTTGA
	EP300-E23R	+299	GCCAATCATTTCCACACCTTA
Exons 24-25	EP300-E24F	-191	GATTAGCATGTTCCCTGCACTC
	EP300-E25R	+114	TTGCATTTCCAAACCAAACAC
Exon 26	EP300-E26F	-106	CAAAGAGCCTGGGAGAGTGAG
	EP300-E26R	+190	GGGCCAACATATCCATTTCTC
Exon 27	EP300-E27F	-94	GGTATCTATATCAACTCCAACTTGTGG
	EP300-E27R	+198	GAATGGCATGAACCCAGG
Exon 28	EP300-E28F	-134	GCTTAGGTATAAAGTCTCTGCCAGC
	EP300-E28R	+178	AAAGTTTAATACCACTGAAACAGAACC
Exon 29	EP300-E29F	-356	ACCAAGGTCTTCTGGCTCC
	EP300-E29R	+373	GCCCAGCGAACAGTCAGT

Exon 30	EP300-E30F	-128	CCATGGTGGGATAATTGCTTG
	EP300-E30R	+74	AAATACGTGGCTGCATGGC
Exon 31-I	EP300-E31F1	-126	GGGCAGAGCTGAAGAGGC
	EP300-E31R1	6069*	CTTGAGTCCTGGGCAAGTAGG
Exon 31-II	EP300-E31F2	5811*	TGCCTAAACATCAAGCAGAAGC
	EP300-E31R2	6546*	AGGCTTGTTGAGACACAGTGC
Exon 31-III	EP300-E31F3	6463*	AACCTTTGAACATGGCTCCAC
	EP300-E31R3	7202*	CATCTGTTGCTGAAGGAGTCG
Exon 31-IV	EP300-E31F4	6918*	ATGCCTTCACAATTCCGAGAC
	EP300-E31R4	7515*	AAAGCATTGAATTCTGGTCCG
Exon 31-V	EP300-E31F5	7272*	CACCTACAAGGCCAGCAGAT
	EP300-E31R5	7778*	CGGCTACTGCACAGTTCTTATG

^ Numbers indicate the distance from the corresponding exon, except where indicated by an asterisk (primers annealing to cDNA sequences)
* Numbering according to GenBank accession No. NM_001429.3

References of Supplementary Material

1. Compagno, M. *et al*. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**, 717-21 (2009).
2. Wright, G. *et al*. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A* **100**, 9991-6 (2003).
3. Hans, C.P. *et al*. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103**, 275-82 (2004).
4. Lin, M. *et al*. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-40 (2004).
5. Mullighan, C.G. *et al*. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-64 (2007).
6. Pounds, S. *et al*. Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics* **25**, 315-21 (2009).
7. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-63 (2007).
8. Gu, W., Shi, X.L. & Roeder, R.G. Synergistic activation of transcription by CBP and p53. *Nature* **387**, 819-23 (1997).
9. Garbati, M.R., Alco, G. & Gilmore, T.D. Histone acetyltransferase p300 is a coactivator for transcription factor REL and is C-terminally truncated in the human diffuse large B-cell lymphoma cell line RC-K8. *Cancer Lett* **291**, 237-45 (2010).
10. Liu, X. *et al*. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846-50 (2008).
11. Bereshchenko, O.R., Gu, W. & Dalla-Favera, R. Acetylation inactivates the transcriptional repressor BCL6. *Nat Genet* **32**, 606-13 (2002).
12. Bordoli, L. *et al.* Functional analysis of the p300 acetyltransferase domain: the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. *Nucleic Acids Res* **29**, 4462-71 (2001).

2.2 The Genetic Landscape of Chronic Lymphocytic Leukemia (paper in preparation)

Giulia Fabbri[1], Vladimir Trifonov[2], Davide Rossi[3], Oliver Elliott[2], Andrea Rinaldi[4], Ivo Kwee[4], Francesco Bertoni[4], Gianluca Gaidano[3], Laura Pasqualucci[1,5], Raul Rabadan[2], and Riccardo Dalla-Favera[1,5,6].

[1]*Institute for Cancer Genetics and the Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA.*

[2]*Department of Biomedical Informatics and the Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA.*

[3]*Division of Hematology, Department of Clinical and Experimental Medicine and IRCAD, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy.*

[4]*Oncology Institute of Southern Switzerland (IOSI), Bellinzona, Switzerland.*

[5]*Department of Pathology & Cell Biology, Columbia University, New York, NY 10032, USA.*

[6]*Department of Genetics & Development, Columbia University, New York, NY 10032, USA.*

Chronic lymphocytic leukemia (CLL) is a malignancy of CD5+ mature B cells that includes two distinct subtypes marked by the differential presence of immunoglobulin gene mutations and a distinct clinical course. The entire spectrum of CLL-associated genetic lesions is currently unknown. To determine the extent of somatic genetic alterations present in the CLL coding genome, we have integrated next generation whole-exome sequencing analysis and genome-wide high-density single nucleotide polymorphism (SNP) array analysis in 5 cases representative of the two CLL immunogenetic subgroups and their paired normal DNAs. Candidate tumor-specific non-synonymous mutations were verified by Sanger sequencing in the same tumor/normal pairs, and all the identified mutated genes were screened by PCR amplification/direct sequencing of their entire coding region and by SNP Array analysis in an independent panel of 48 CLL cases. A total of 38 somatic non-silent mutations, corresponding to 38 distinct genes were identified. SNP array analysis in the same 5 CLL discovery cases confirmed the presence of known cytogenetic abnormalities in 4 cases, and identified 10 additional regions of copy number (CN) changes, mostly represented by deletions. The overall load of potentially relevant lesions per case was 10. When screened in the panel of 48 CLLs the majority of the identified genes were not recurrently altered in CLL, while 5 genes showed either recurrent mutations or CN aberrations at a low frequency (≤ 6% of CLL cases).

Overall, our integrated approach revealed that the coding genome of CLL appears to be relatively stable, containing on average 10 potential relevant genetic alterations per case. The low frequency of alterations in individual genes may reflect a high recurrence of alterations affecting common pathways important for the pathogenesis of the disease.

Introduction

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia in the Western world. It is characterized by the accumulation of CD5+ mature B lymphocytes in blood, bone marrow, lymph nodes or other lymphoid tissues.

Presence or absence of somatic mutations in the IgV-H gene sequences defines two disease subtypes associated with a distinct clinical outcome. CLL cases carrying IgV-H genes with <98% of similarity to the closest germline gene ("mutated CLL") have a more indolent clinical course than those with ≥ 98% similarity ("unmutated CLL")[1-2].

Independent of the IgV mutational status, CLL cases display a largely homogeneous phenotype, that shares significant similarities with that of antigen-experienced B cells, suggesting an origin from the oncogenic transformation of a common cellular precursor[3-4]. Moreover, all CLL cells express B-cell receptors that show evidence of antigen encounter and that can be classified into subsets of stereotypy based on the homology in the antigen-binding regions[5-8].

CD5+ monoclonal B cell lymphocytosis (MBL), defined by the presence of clonal B cell populations with a CLL immunophenotype in the peripheral blood of healthy adults, can progress to CLL in ~1% of cases, while there is evidence that CLL is preceded by MBL in most of the cases[9-10]. In a small percentage of cases CLL can transform to an aggressive form of clonally related lymphoma, most commonly of the diffuse large B cell type, termed Richter Syndrome[11].

The molecular pathogenesis of CLL is largely unknown: in contrast to other types of germinal center-derived B cell malignancies, CLL is not associated with recurrent balanced chromosomal translocations. So far no genes have been found that are specifically targeted by somatic mutations, with the exception of *ATM* and *TP53*, mutated in 6-12% of cases[12-15]. Instead, more than 80% of CLL patients carry recurrent chromosomal abnormalities, including mono and biallelic deletions of chromosomal regions 13q14, 11q22-23, and 17p13, and trisomy 12. These genetic lesions are independent predictors of disease progression and survival in CLL patients[16]. Of these aberrations, deletion 13q14, encompassing the *DLEU2/miR-15a/miR-16-1* cluster[17-18], has been shown to promote the development of a variety of lymphoproliferative disorders including MBL, CLL and B-cell non-Hodgkin lymphoma (B-NHL) in mice[19], demonstrating its pathogenetic role in the human disease. However, the known CLL-associated lesions do not account for all CLL cases, and therefore they most likely represent only a subset of those required for neoplastic initiation. Genome-wide methods aimed at the characterization of the entire spectrum of genetic lesions present in the CLL genome are required not only to provide insights into its pathogenesis, but also to develop novel targeted therapies.

In order to determine the extent of somatic genetic lesions (non-silent coding mutations and gene copy number changes) associated with CLL, we have integrated next generation whole-exome (WES) sequencing analysis (Nimblegen/Roche FLX454)[20-21] and genome-wide high-density single nucleotide polymorphism (SNP) array analysis (Affymetrix SNP 6.0)[22] of 5 CLL cases representative of distinct immunogenetic subgroups and their paired normal DNAs.

Materials and methods

Primary tumor and normal samples, DNA extraction and Whole Genome Amplification

Samples from 5 newly diagnosed and previously untreated CLL patients (Discovery panel) were obtained from the Department of Clinical and Experimental Medicine, University of Novara, as frozen peripheral blood mononuclear cells isolated by Ficoll-Paque gradient centrifugation. The study was approved by the institutional review board of Columbia University. In all cases, the fraction of tumor cells corresponded to >80%, as assessed by FACS analysis of CD19 expression. The matched normal samples were obtained as frozen purified granulocytes. High molecular weight genomic DNA was extracted by standard methods and the absence of contaminating tumor cells in the normal samples was verified by analysis of the clonal immunoglobulin gene rearrangement observed at diagnosis (see below).

For the Screening phase, 48 additional CLL diagnostic samples and matched normal DNAs (Screening panel) were provided by the Department of Clinical and Experimental Medicine, University of Novara. Whole-genome amplification was performed on two independent aliquots of high molecular weight genomic DNA (12ng each) with the use of the Repli-G Midi kit (Qiagen, Valencia, CA). Quantitation and purity of the whole-genome amplified DNA were assessed by Nanodrop (Thermo Scientific®, Wilmington, DE). The quality of the extracted DNA was verified by gel electrophoresis and by PCR amplification of two housekeeping genes.

IgV-H mutational status

IgV-H mutational status was performed as previously reported[23]. PCR products were directly sequenced with the ABI PRISM BigDye Terminator v1.1 Ready Reaction Cycle Sequencing kit (Applied Biosystems) using the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Sequences were aligned to the ImMunoGeneTics sequence directory[24] and considered mutated if homology to the corresponding germline gene was <98%[1-2].

Interphase fluorescence in situ hybridization analysis

The following probes were used for fluorescence in situ hybridization (FISH) analysis: i) LSID13S319, CEP12, LSIp53, LSIATM, LSI IGH/BCL2, LSI IGH/CCND1, LSI BCL6, LSI, IGH/c-MYC/CEP8, c-MYC break-apart, LSI N-MYC, CEP2, CEP3, CEP18, CEP19 (Abbott, Rome, Italy); ii) a BCL-3 split signal probe (Dako); iii) 6q21/alpha-satellite (Kreatech Biotechnology, Amsterdam, The Netherlands); iv) BAC clones 373L24-REL and 440P05-BCL11A. The detailed protocol is described elsewhere[23].

Exome Capture and whole exome sequencing approach

The CLL tumor/normal matched DNA samples were used for exon-capture and next-generation whole-exome sequencing through the Roche NimbleGen Sequence Capture 2.1M Human Exome Array and the Roche FLX454 Sequencer, according to the manufacturer's instructions.

The Roche NimbleGen Array contains probes for capturing ~15,000 coding genes and 551 miRNAs (~34Mb of target region spanning ~1% of the human genome). 1,343 coding genes are not covered or covered in <30% of their coding sequence. Briefly, 5ug of high molecular weight genomic DNA were sonicated and linker-ligated. After removing of small fragments, the 500-700 bp whole genome fragment pool was hybridized to the array at 42°C for 64 hours. After hybridization the array was washed at high stringency at 47.5°C to remove background unbound DNA. The bound genomic DNA was eluted, purified and amplified by linker-mediated PCR (LM-PCR). qPCR SYBR Green assays were performed to estimate relative fold-enrichment and to estimate the efficiency of the capture prior to sequencing. The successfully captured samples were ready for high-throughput sequencing with the Genome Sequencer FLX System.

The 454 sequencing is based on an emulsion PCR (emPCR) approach in a context of a water-in-oil emulsion. One of the PCR primers is 5'-attached on 28uM beads that are included in the reaction. Every bead is deposited to a microfabricated array of picoliter-scale wells together with DNA polymerase, ATP sulfurylase, luciferase and apyrase. The sequencing process is based on a reaction in which each cycle consists of the introduction of a single nucleotide (dNTP), followed by the addition of substrate (luciferin-adenosine-5'-phosphosulfate). If the dNTP is complementary to the base in the template strand there is release of pyrophosphate, which is converted by the enzyme ATP sulfurylase to ATP in the presence of adenosine-5'-phosphosulfate. The ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin, with subsequent light production at wells where polymerase-driven incorporation of that nucleotide took place. This is followed by wash mediated by the enzyme apyrase, a nucleotide degrading enzyme that continuously degrades ATP and unincorporated dNTPs.

454 Sequencing data analysis

The computational pipeline used for the analysis of the deep sequencing data is reported in Fig. S1.

The sequencing reads obtained from the Roche FLX454 Sequencer were aligned to the human reference genome (NCBI build 36; hg18 assembly) with the software Genome Sequencer Reference Mapper, version 2.3 (Roche). Sequencing reads aligning in multiple positions of the reference sequence were discarded, and only reads mapping uniquely to the human genome were considered for further analysis.

Duplicate reads of a single DNA input are a known potential artifact of the emulsion PCR step of 454 sequencing. Groups of duplicate reads, starting at the same base of the reference sequence, were therefore counted as single reads for the subsequent computation of variants.

The alignments between the uniquely mapped reads and the reference genome were used to define the genetic variants in the CLL tumor and normal DNAs as single base-pair substitutions, small insertions and small deletions (indels).

The software uses a combination of informations to identify "High-Confidence" variants. There must be at least 3 non-duplicate reads harboring the variant (two in the sense and one in the antisense orientation). Considering that the genome is diploid, more than $m=3$ reads are required in one allele to make a call. The effective coverage ("High Confidence Coverage", HCC) to discover new, clonally represented variants in a 100% pure diploid genome can be approximated to the portion of the exome with ≥6 fold depth (with at least 3 reads/allele), according to the following formula:

$$HCC = \sum_{m}^{\infty} \frac{N_k}{N_T} \sum_{n=m}^{k} \binom{k}{n}$$

in which N_k corresponds to the number of base-pairs covered by k sequencing reads and N_T corresponds to the total number of base-pairs covered.

Only non-silent mutations spanning the targeted exomic sequence or changing one of the 4 nucleotides flanking each consensus splice site were considered.

Germline polymorphisms, including SNPs reported in publicly available databases (dbSNP, Ensembl Database, UCSC Genome Browser) and variants identified in the paired normal DNAs were removed. Systematic errors of the technique (small indels affecting homo-polymer repeats and stretches) and variants below the sensitivity of direct sequencing (observed in less than 20% of the reads), were also discarded.

Each of the candidate tumor-specific non-synonymous variants was then amplified from matched tumor and normal CLL samples by polymerase chain reaction (PCR) and the obtained products were used for direct sequencing by Sanger method (see below).

PCR amplification and Sanger sequencing analysis

The sequences surrounding the genomic locations of the candidate tumor-specific non-synonymous mutations identified through whole exome sequencing were obtained from the UCSC Human Genome database (Validation phase). The sequences for all annotated exons and flanking introns of the genes confirmed to be somatically mutated in CLL (listed in Table S1) were also obtained from the UCSC Human Genome database, using the corresponding mRNA accession number as a reference (Screening phase).

The oligonucleotides used for PCR amplification and direct sequencing of each candidate tumor-specific non-synonymous mutation and of each coding exon of the somatically mutated genes (including ~50bp of adjacent introns) were designed using the Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

PCR was performed in a 25uL reaction containing 1X Buffer 3 (Expanded Long Template PCR System, Roche), 200uM dNTPs, 0.2uM of forward and 0.2uM of reverse oligonucleotides, 5% DMSO, 0.06U of Taq DNA Polymerase (Invitrogen) and 25ng of template DNA. The PCR reactions were carried out in 96-well thermocyclers (Eppendorf Mastercycler) using a touch-down protocol previously reported[25].

The oligonucleotides used for the Validation phase produced PCR products and chromatograms evaluable in ~99.3% of the target regions in the 5 discovery cases. The vast majority of the oligonucleotides used for the Screening phase (~98%) yielded PCR products and high quality sequencing in 40 or more of the 48 screening samples.

Purified amplicons were sequenced directly from both strands, and compared to the corresponding germline sequences, using the Mutation Surveyor Version 2.41 software package (Soft Genetics, LLC).

For the Validation phase high molecular genomic, while for the Screening phase whole genome amplified DNAs were used for the reactions. In the Screening phase synonymous mutations, changes due to previously reported polymorphisms (Human dbSNP Database at NCBI, Build 132, and Ensembl Database) and changes present in normal genomic DNA from the same patient were excluded. Somatic mutations were confirmed on independent PCR products using the corresponding genomic DNA to exclude false positives arising from PCR or sequencing artifacts.

Copy number analysis by high-density SNP array analysis

Genome-wide DNA profiles were obtained from high molecular weight tumor and normal genomic DNA of 53 CLL patients using the Affymetrix Genome-Wide Human SNP

Array 6.0 (Affymetrix, Santa Clara, CA). Briefly, DNA was restriction enzyme digested, ligated, PCR-amplified, purified, labeled, fragmented and hybridized to the arrays according to the manufacturer's instructions.

Image data analysis and quality control for the hybridized samples were performed using the Affymetrix Genotyping Console 3.0.1 software, and only samples passing the Affymetrix recommended contrast QC and SNP call rates threshold (in the Birdseed v2.0 algorithm) were considered for analysis. Affymetrix CEL files and corresponding SNP genotype call files generated by the Affymetrix Genotyping Console tool were then analyzed using the dChip software[26] according to a previously published workflow[27-28]. Candidate genomic regions of amplification and deletion were identified by applying the circular binary segmentation (CBS) algorithm to the SNP array data as described[27,29], with the following criteria: i) mean log2 ratios of ≥ 0.2 or ≤ -0.2; ii) ≥ 8 SNP markers (SNP and/or CNV) within a segment. The results of the CBS algorithm were then compared to those of dChip. To exclude calls of genomic gains or loss arising from inherited genomic CNV, we used a reference set of matched normal DNAs (available for 4 out of 5 "discovery" CLL cases), or a panel of 130 normal DNAs from an independent study[27] as well as 230 normal DNAs from the HapMap project, which were analyzed in parallel using the dChip SNP algorithm. In addition, CNV were excluded if present in the Database of Genomic Variants (http://projects.tcag.ca/cgi-bin/variation/gbrowse/hg18/).

RNA extraction and Gene expression profile analysis

Adequate material for extraction of RNA was available for 3 samples, which were characterized by gene expression profile analysis.

Total RNA extraction was performed using TRIzol (Invitrogen, Carlsbad, CA) following the manufacturer's instruction.

Gene expression profile analysis of normal mature B cells (naïve, memory, centroblasts and centrocytes) and primary CLL cases was performed using Affymetrix HG-U133Plus 2.0 arrays. The cRNA was labeled and hybridized to the array according to the manufacturer's protocol. The data were analyzed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. Probe sets that had a mean (mu) <50, or a coefficient of variation (CV) <0.3, as assessed in a panel of 259 independent B cell related gene expression profiles analyzed on the same array, were filtered out and considered not informative. To assess the expression of the genes found mutated through the whole exome sequencing analysis, we considered expressed genes

whose probe was called “Present” in the CLL case carrying the mutation or whose probe had more than 90% of “Present” calls in an independent panel of 16 CLL primary cases hybridized on the same platform.

Functional categories and Pathways Analysis of the mutated genes

The genes found to be mutated in CLL were assigned to functional categories or to annotated pathways using the gene ontology GO database (http://www.geneontology.org/), and the publicly available bioinformatic tool DAVID 2008 (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov/).

Results

Patients’ features

The biological and clinical characteristics of the 5 CLL samples used in the whole exome sequencing (WES) analysis are shown in Table 1. Patients were representative of the two major immunophenotypic subgroups (IGVH mutated: 2 cases; IGVH unmutated: 3 cases) and harbored common CLL-associated genetic abnormalities, including deletion of 13q14 (2 cases), deletion of 11q23 (1 case), and trisomy 12 (1 case), while one case had normal FISH.

The mean age of the patients was 63.5 years (range, 46-71). Two patients had Rai stage 0 disease, 2 patients stage II and 1 patient stage IV. The median time from diagnosis to treatment (Time to first therapy, TTFT) was 27.3 months (range, 0-85.7). One CLL case (CLL1) underwent clinical transformation to Diffuse Large B cell Lymphoma (DLBCL) after 5.2 years of treatment and subsequently died due to infectious complications.

Results of 454 Sequencing and Mapping after Whole Exome Capture

To identify somatic mutations associated with CLL pathogenesis, we performed next generation pyrosequencing of exonic regions in matched tumor and normal DNAs obtained from 5 CLL patients. The results of the 454 sequencing and mapping after whole exome capture are summarized in Table S2. Overall, we obtained 25,023,143 sequence reads of ~300bp median length (range: 2,175,807-2,715,180/case), with no significant differences between tumor and normal samples (on average, 893Mb/case).

In all experiments >99% of the sequence reads were successfully mapped to the human reference genome (NCBI build 36; hg18 assembly). Of these, 74% corresponded to the target exomic region, resulting in a mean depth of coverage of 9.8-fold in the tumor DNAs (range, 9-11.6/case) and 10.4-fold in the normal DNAs (range, 9.1-11.8/case) (Table S2). While these values are lower than what has been reported in several recent cancer genome studies, it is important to note that because of the longer reads, the relatively low error rate, and the minor fraction of duplicate reads produced by the FLX454 Sequencer, as compared to other approaches, the cutoff used for the identification of candidate mutations can be very low (n=3 variant sequence reads, of which one in the opposite direction). Thus, the effective depth of coverage required for discovering new heterozygous variants in a highly enriched diploid organism can be approximated to the portion of the exome with depth equal or greater than 5, which corresponded to ~80% of the sequenced target exome (see Methods and Table S2). On average, ~96% of the target exome was covered by at least 1 read, indicating that the exomic enrichment was successful.

After filtering for germline SNPs, systematic errors and variants present in less than 20% of the sequencing reads (i.e. below the sensitivity of Sanger sequencing), 277 candidate tumor-specific non-synonymous variants were identified in the 5 tumor cases and subjected to validation by PCR amplification and direct sequencing of the paired tumor/normal DNAs (Fig. S1 and Table S3-S4).

Mutation Load in CLL

Thirty-eight sequence variants, corresponding to 38 distinct protein coding genes, were confirmed to be somatic in origin, at an average of 7.6 mutations/case (range, 5-10) (Fig. 1a and Table S3-S4). Of the remaining variants, 170 were also found in the matched normal DNAs, thus representing non previously annotated germline polymorphisms that escaped detection during the high throughput sequencing analysis, possibly due to the relatively low depth of coverage at their relative genomic position in the normal DNA (<5 reads in 55% of the cases).

The majority of the observed mutations were represented by single base-pair substitutions (N=36, 94.7%) which introduced aminoacid changes (N=32) or premature stop codons (N=4) in the sequence of the affected genes, while frameshift insertions/deletions were rare (N=2 deletions of 4 and 32 base pairs, respectively; 5.3%) (Fig. 1b). No mutations were found to affect consensus splice sites. Analysis of the mutation context showed a prevalence of transitions versus transversions (69% vs 31%) and an elevated G+C over A+T

ratio (69% vs 31%), analogous to the mutation spectrum observed in the genome of epithelial tumors, such as colorectal, pancreatic and brain cancer, and of diffuse large B cell lymphoma (Fig. 1c-d and Table S5)[30-31].

Analysis of DNA Copy number aberrations

Genome-Wide SNP array analysis of the same 5 cases revealed a total of 14 copy number aberrations (CNAs), with on average 2.8 lesions per case (range, 0-5) (Table S6). Eleven of the affected regions harbored genetic elements, including protein coding genes and/or non coding RNA genes, and six of them encompassed less than 3 genes (Fig. 2a).

Consistent with previous reports[32-33], the majority of the observed lesions were represented by deletions (11/14 CNAs, 70%), ranging in size from 32 kb to the loss of an entire chromosomal arm, as in CLL1 which carried an 11q deletion. The three regions of gain included a gain of chromosome 12 in a CLL patient #4, a region encompassing an intron of the *TMEM117* gene and a region apparently devoid of genetic elements.

Importantly, the SNP array approach successfully identified all of the 4 lesions previously detected by FISH analysis, including two 13q14 deletions (CLL2 and CLL5), one trisomy 12 (CLL4) and one 11q deletion (CLL1) (Fig. 2b). Both 13q14 deletions encompassed the *miR-15a/16-1* locus along with 50 to 58 additional genetic elements, including the *RB1* gene. The 11q deletion spanned more than one hundred genes, including the tumor-suppressor *ATM* and the negative regulator of the non-canonical NF-κB pathway *BIRC3* (see below).

Overall load of genetic lesions in CLL

The integration between whole exome sequencing and copy number data in the 5 CLL discovery cases revealed a total of 52 genetic lesions, ranging from 5 to 13 lesions per case (Fig. 3).

Therefore, on average, each CLL case had ~10 potentially relevant genomic alterations (considering that we may be missing ~20% of the mutations). This load is 5 to 20 times lower than that observed in solid tumor genomes sequenced to date[30,34-35], while it is comparable to the one reported in acute myeloid leukemia[36] and in pediatric medulloblastoma[31], suggesting that the CLL genome is relatively stable (see discussion).

CLL mutated genes

The detailed list of mutations identified by WES analysis is reported in Table 2. While none of the identified genes had been previously found mutated in CLL, fourteen of them are listed in The Catalogue of Somatic Mutations in Cancer (COSMIC) v50[37] as targets of somatic mutations in other cancer types.

The assignment of the 38 mutated genes to annotated molecular functions or biological process groups, reported in the Gene Ontology (GO) Database, revealed a wide variety of categories of interest, the most common being represented by transcriptional regulation and chromatin remodeling processes (N=12/38 mutated genes, 32%). This functional group included, among others, the *ACTL6A* gene, encoding for a component of the SWI/SNF complex, that has been shown to be essential in the promoter-directed assembly of *TCR-β* genes and plays multiple roles in IgH gene assembly[38], and *WHSC1*, encoding for a SET domain histone methyltransferase protein, affected by chromosomal translocations in multiple myeloma[39].

Genes involved in diverse signal transduction pathways were also highly represented (N=11/38 mutated genes), and included components of the NF-κB (*BIRC3*, *MYD88* and *PLEKHG5*), Wnt (*FAM123B*) and TGF-β (*GDF2*) signaling pathways. *BIRC3*, a negative regulator of the non-canonical NF-κB pathway that is affected by chromosomal translocations in MALT lymphomas and by focal deletions in multiple myeloma[40-42], harbored a frameshift deletion located between the third Baculovirus IAP Repeat (BIR) and the Caspase Activation and Recruitment Domain (CARD) in one case. This mutation is predicted to remove the RING domain, which is required for its inhibitory function, and may therefore lead to its inactivation. Notably, the second allele of this gene was also lost in the mutated CLL case due to the presence of a deletion affecting the entire long arm of chromosome 11, suggesting a tumor suppressor role. MYD88, an adapter protein involved in Toll-like receptor pathway and in NF-κB activation, harbored a missense mutation (L265P) in the Toll/IL-1 receptor (TIR) domain, which was recently found in a large fraction of ABC-DLBCL, where it was shown to be oncogenically active[43]. *FAM123B*, a negative regulator of the Wnt pathway, which is frequently deleted and/or mutated in Wilms tumors[44-45], carried a missense mutation predicted by PolyPhen-2 to have deleterious consequences for the function of the encoded protein, as assessed by the PolyPhen-2 algorithm (see below).

Genes coding for proteins involved in cytoskeleton organization, cell motion and cellular transport processes were found mutated in the majority of the analyzed CLL cases

(four out of five). Among these: *ABCA7* and *SLC1A1*, involved in lipid and glutamate transport[46-47]; *ACTN1*, encoding for a bundling cytoskeleton protein; *GOLM1*, *GPRASP1* and *SNX9*, involved in protein transport through the Golgi apparatus, lysosomal sorting and intracellular trafficking[48-50]; and *AP1G2*, encoding for an adaptor protein, that plays an important role in clathrin-coated transport of ligand-receptor complexes from the plasma membrane or the trans-Golgi network to the lysosomes[51].

The PolyPhen algorithm, a sequence/structure-based method that computes the functional consequence of point mutations (http://genetics.bwh.harvard.edu/pph2/)[52], predicted an altered protein function in most of the identified missense events (N=22/32, 69%) were (Table 2). Sixteen out of 38 (42%) genes appear to be expressed at detectable levels in CLL (Table 2).

The identified genes are rarely mutated in CLL

To evaluate the prevalence of genetic alterations affecting the newly identified mutated genes, we screened an independent panel of 48 CLL cases by PCR amplification and direct sequencing of their entire coding region and by genome-wide SNP Array analysis. Two out of the 38 genes were not further investigated, since the mutations found in the discovery panel most likely represent a subclonal event. In addition, we screened for the presence of mutations those genes that had been found targeted by copy number aberrations, if located in regions of recurrent focal loss or gain.

This analysis revealed the presence of additional lesions (mutations and/or copy number changes) in 5 genes, including *PLEKHG5*, *TGM7*, *BIRC3*, *ACTN1* and *INHBC*, all at very low frequency (≤ 6%) (Fig. 4).

PLEKHG5, a family G-member protein that plays a role in NF-κB activation, was found mutated in one case (overall, 2/53 cases). *TGM7*, a tranglutaminase that mediates protein-crosslinks harbored a somatic missense mutation (V103L) which was predicted to potentially alter the protein function in one additional patient.

Interestingly, *BIRC3* was biallelically inactivated by a nonsense mutation (E424X) and a deletion truncating the corresponding protein between the third BIR domain and the CARD domain, similarly to what observed in the discovery case CLL1 (overall frequency N=2/53 cases).

ACTN1 and *INHBC* were affected by non focal (i.e. encompassing >100 genes) CNAs in the screening panel.

The targeted resequencing of *FGFR2*, the only gene mapping in a recurrent region of focal CN loss (overall frequency, 5.4%) did not reveal any somatic mutations in the 48 patients of the screening panel.

Additional 28 sequence changes identified in 15 genes resulted to be "private" polymorphisms not previously reported in publicly available databases.

Overall, these data show that none of the identified genes is affected at high frequency in CLL.

Discussion

The data reported here provide a comprehensive analysis of the full spectrum of somatic lesions (non-silent mutations and copy number aberrations) that are clonally represented in the CLL genome at diagnosis, and may have thus contributed to the malignant transformation process.

Our data show that the CLL coding genome harbors on average ~10 potentially relevant genetic lesions. The observed load is up to 20 times lower than the one reported in melanoma and lung cancer, where factors such as sun light and tobacco smoke cause additional mutations on top of the ones introduced by intrinsic and endogenous cellular processes. This number is also up to 10 times lower than the one reported for other solid tumors, such as pancreatic, brain, breast and colorectal cancers, and is comparable to the mutation load of acute myeloid leukemia. These findings are consistent with the notion that the genome of lymphoid malignancies is relatively stable and lacks the generalized random instability typical of many epithelial cancers.

We identified somatic non-synonymous mutations in thirty-eight genes that had not been previously linked to CLL, and copy number aberrations in 10 regions, besides the ones already known to be associated with the disease. Although none of the newly identified lesions appeared to be recurrent at high frequencies when screened in a larger panel of cases, a number of the involved genes converge on common pathways and/or biological processes that may play a broader role in CLL pathogenesis.

Mutations affecting chromatin remodeling genes were found in 3 CLL cases, suggesting that deregulation of this process may be a relevant contributor to neoplastic transformation. Among them, *ACTL6A* is an actin-like member of the SWI/SNF complex, which acts as a chromatin-remodeling machine by providing proper nucleosome positioning and density at genes and other loci. Several components of this complex have been recently

shown to be recurrently mutated in other cancer types[53-59]. *HIST1H2AM*, which encodes for the nucleosome component Histone H2A type 1, was found to be mutated in the histone-fold domain in another CLL case. One CLL case carried a potentially deleterious missense mutation in the SET domain of the histone-methyltransferase gene *WHSC1*, which is known to undergo translocations[39] and rare somatic mutations (Golub T, personal communication) in multiple myeloma. Genes involved in epigenetic regulation of gene expression through methylation- and demethylation-based processes have been recently linked to various cancer types, being affected by recurrent mutations[31,60-68].

We also found mutations in several genes that play a role in signaling cascades previously linked to CLL, such as the NF-κB and the Wnt pathways. In particular, *BIRC3*, a negative regulator of the non-canonical NF-κB pathway, was biallelically inactivated in two patients, consistent with a tumor suppressor role. One additional patient harbored an oncogenic activating mutation in *MYD88*, which is known to enhance NF-κB in ABC-DLBCL[43]. To date there are multiple evidences of constitutive NF-κB activation in CLL (reviewed in Ref [69]). While a major role for sustained NF-κB activity in CLL is likely mediated by the contact between CLL cells and microenvironment[70], future studies should include the mutational screening of additional genes within this pathway. The cytosolic adapter MYD88 also plays a central role in innate and adaptive immune responses, functioning as essential signal transducer of the Toll-like receptor signaling pathway. This pathway is of particular interest as it has been shown to participate in memory B cell development and in T cell dependent immune responses[71-72].

The identification of a missense mutation in the negative Wnt regulator *FAM123B* (*WTX*) also warrants further investigations. Wnt signaling genes have been shown to be overexpressed and active in CLL[73] and, more recently, somatic mutations activating this pathway have been reported at a low frequency in CLL[74].

Finally, several of the CLL-mutated genes play a role in cytoskeleton organization, cell motion and transport processes. Several studies have shown that CLL cells have an altered cytoskeletal organization[75-77]. More recently, it has been shown that HS1 is a central regulator of cytoskeleton remodeling by controlling lymphocyte trafficking and homing and by influencing the tissue invasion and infiltration in CLL[78]. The hyperphosphorylated form of this protein is associated with poor outcome[79]. The cytoskeleton has an important role in the B cell receptor (BCR)-mediated process of Ag-recognition and cell activation, implying the need of actin polymerization and cytoskeleton reorganization. Therefore, cytoskeletal defects

could partially explain the attenuated BCR response to Ag stimulation observed in CLL (reviewed in Ref.[80]).

The prevalence of mutations in the identified genes in an extended panel of CLL cases was overall low, with no gene found to be affected in more than 6% of the cases. There are several possible explanations to this finding. It is possible that the most common CLL-associated lesion have already been identified (i.e. 13q14, 11q and 17p deletions and trisomy 12, and *TP53* and *ATM* mutations). However, at least some of the identified genetic lesions may play an important role in the pathogenesis of CLL, given their proven oncogenic activity in other disease (e.g. *MYD88* in DLBCL[43]). Indeed, dysregulation of specific cellular processes may be determined by alterations of distinct components in individual patients. Therefore, genomic analysis of other pathway members is required to define the relative contribution of individual lesions and the roles of combined alterations.

Finally, we cannot rule out that frequent events affecting protein coding genes escaped detection because located in the portion of the exome that was not captured or that was sequenced at low depth of coverage. Analogously, other types of genetic alteration such as mutations in non-coding genes, mutations in the non-coding region of coding genes, and gene rearrangements, would not be detected by the method used.

Future efforts will be required to assess the definitive landscape of the disease and to establish accurate estimates of mutation frequencies by sequencing many additional CLL cases, possibly extending the analysis to non-coding genomic regions.

References

1. Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G. & Stevenson, F.K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848-1854 (1999).
2. Damle, R.N., *et al.* Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94**, 1840-1847 (1999).
3. Klein, U., *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* **194**, 1625-1638. (2001).
4. Rosenwald, A., *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* **194**, 1639-1647. (2001).
5. Fais, F., *et al.* Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* **102**, 1515-1525 (1998).

6. Tobin, G., *et al.* Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood* **104**, 2879-2885 (2004).
7. Messmer, B.T., *et al.* Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J Exp Med* **200**, 519-525 (2004).
8. Stamatopoulos, K., *et al.* Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* **109**, 259-270 (2007).
9. Rawstron, A.C., *et al.* Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med* **359**, 575-583 (2008).
10. Landgren, O., *et al.* B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med* **360**, 659-667 (2009).
11. Rossi, D., *et al.* Biological and clinical risk factors of chronic lymphocytic leukaemia transformation to Richter syndrome. *Br J Haematol* **142**, 202-215 (2008).
12. Zenz, T., *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* **112**, 3322-3329 (2008).
13. Zenz, T., *et al.* TP53 mutation profile in chronic lymphocytic leukemia: evidence for a disease specific profile from a comprehensive analysis of 268 mutations. *Leukemia* (2010).
14. Austen, B., *et al.* Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood* **106**, 3175-3182 (2005).
15. Austen, B., *et al.* Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. *J Clin Oncol* **25**, 5448-5457 (2007).
16. Dohner, H., *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* **343**, 1910-1916 (2000).
17. Migliazza, A., *et al.* Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood* **97**, 2098-2104. (2001).
18. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-866 (2006).
19. Klein, U., *et al.* The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell* **17**, 28-40 (2010).
20. Albert, T.J., *et al.* Direct selection of human genomic loci by microarray hybridization. *Nature methods* **4**, 903-905 (2007).
21. Margulies, M., *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
22. Matsuzaki, H., *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature methods* **1**, 109-111 (2004).
23. Rossi, D., *et al.* Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin Cancer Res* **15**, 4415-4422 (2009).
24. Chiorazzi, N., Rai, K.R. & Ferrarini, M. Chronic lymphocytic leukemia. *N Engl J Med* **352**, 804-815 (2005).
25. Sjoblom, T., *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274 (2006).

26. Lin, M., *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-1240 (2004).
27. Mullighan, C.G., *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764 (2007).
28. Pounds, S., *et al.* Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics* **25**, 315-321 (2009).
29. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-663 (2007).
30. Jones, S., *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806 (2008).
31. Parsons, D.W., *et al.* The Genetic Landscape of the Childhood Cancer Medulloblastoma. *Science* (2010).
32. Kujawski, L., *et al.* Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. *Blood* **112**, 1993-2003 (2008).
33. Grubor, V., *et al.* Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* **113**, 1294-1303 (2009).
34. Pleasance, E.D., *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196.
35. Pleasance, E.D., *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190.
36. Mardis, E.R., *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066 (2009).
37. Forbes, S.A., *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al* **Chapter 10**, Unit 10 11 (2008).
38. Osipovich, O.A., Subrahmanyam, R., Pierce, S., Sen, R. & Oltz, E.M. Cutting edge: SWI/SNF mediates antisense Igh transcription and locus-wide accessibility in B cell precursors. *J Immunol* **183**, 1509-1513 (2009).
39. Martinez-Garcia, E., *et al.* The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* (2010).
40. Zhou, H., Du, M.Q. & Dixit, V.M. Constitutive NF-kappaB activation by the t(11;18)(q21;q21) product in MALT lymphoma is linked to deregulated ubiquitin ligase activity. *Cancer Cell* **7**, 425-431 (2005).
41. Annunziata, C.M., *et al.* Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell* **12**, 115-130 (2007).
42. Keats, J.J., *et al.* Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. *Cancer Cell* **12**, 131-144 (2007).
43. Ngo, V.N., *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* (2010).
44. Rivera, M.N., *et al.* An X chromosome gene, WTX, is commonly inactivated in Wilms tumor. *Science* **315**, 642-645 (2007).
45. Major, M.B., *et al.* Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. *Science* **316**, 1043-1046 (2007).
46. Jehle, A.W., *et al.* ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. *J Cell Biol* **174**, 547-556 (2006).

47. Aoyama, K., *et al.* Neuronal glutathione deficiency and age-dependent neurodegeneration in the EAAC1 deficient mouse. *Nat Neurosci* **9**, 119-126 (2006).
48. Kladney, R.D., *et al.* GP73, a novel Golgi-localized protein upregulated by viral infection. *Gene* **249**, 53-65 (2000).
49. Gray, J.A. & Roth, B.L. Cell biology. A last GASP for GPCRs? *Science* **297**, 529-531 (2002).
50. Yarar, D., Waterman-Storer, C.M. & Schmid, S.L. SNX9 couples actin assembly to phosphoinositide signals and is required for membrane remodeling during endocytosis. *Developmental cell* **13**, 43-56 (2007).
51. Doring, T., Gotthardt, K., Stieler, J. & Prange, R. gamma2-Adaptin is functioning in the late endosomal sorting pathway and interacts with ESCRT-I and -III subunits. *Biochim Biophys Acta* **1803**, 1252-1264 (2010).
52. Adzhubei, I.A., *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249 (2010).
53. Huang, J., Zhao, Y.L., Li, Y., Fletcher, J.A. & Xiao, S. Genomic and functional evidence for an ARID1A tumor suppressor role. *Genes Chromosomes Cancer* **46**, 745-750 (2007).
54. Jones, S., *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231 (2010).
55. Wiegand, K.C., *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* **363**, 1532-1543 (2010).
56. Varela, I., *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* (2011).
57. Schneppenheim, R., *et al.* Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *American journal of human genetics* **86**, 279-284 (2010).
58. Versteege, I., *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206 (1998).
59. Wong, A.K., *et al.* BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer Res* **60**, 6171-6177 (2000).
60. van Haaften, G., *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet* **41**, 521-523 (2009).
61. Mendez-Lago, M. Mutations In MLL2 and MEF2B Genes In Follicular Lymphoma and Diffuse Large B-Cell Lymphoma. *Blood ASH abstracts* (2010).
62. Delhommeau, F., *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301 (2009).
63. Langemeijer, S.M., *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-842 (2009).
64. Dalgliesh, G.L., *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-363 (2010).
65. Morin, R.D., *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-185 (2010).
66. Nikoloski, G., *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42**, 665-667 (2010).
67. Ernst, T., *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722-726 (2010).
68. Ley, T.J., *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433 (2010).

69. Hertlein, E. & Byrd, J.C. Signalling to drug resistance in CLL. *Best Pract Res Clin Haematol* **23**, 121-131 (2010).
70. Herishanu, Y., *et al.* The lymph node microenvironment promotes B-cell receptor signaling, NF-{kappa}B activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* **117**, 563-574 (2011).
71. Pasare, C. & Medzhitov, R. Control of B-cell responses by Toll-like receptors. *Nature* **438**, 364-368 (2005).
72. Meyer-Bahlburg, A., Khim, S. & Rawlings, D.J. B cell intrinsic TLR signals amplify but are not required for humoral immunity. *J Exp Med* **204**, 3095-3101 (2007).
73. Lu, D., *et al.* Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* **101**, 3118-3123 (2004).
74. Wang, L. Whole Genome Sequencing Identifies Functional Mutations In the Wnt Pathway In CLL. *Blood, ASH Abstracts* (2010).
75. Stark, R.S., Liebes, L.F., Shelanski, M.L. & Silber, R. Anomalous function of vimentin in chronic lymphocytic leukemia lymphocytes. *Blood* **63**, 415-420 (1984).
76. Stark, R., Liebes, L.F., Nevrla, D., Conklyn, M. & Silber, R. Decreased actin content of lymphocytes from patients with chronic lymphocytic leukemia. *Blood* **59**, 536-541 (1982).
77. Caligaris-Cappio, F., *et al.* Cytoskeleton organization is aberrantly rearranged in the cells of B chronic lymphocytic leukemia and hairy cell leukemia. *Blood* **67**, 233-239 (1986).
78. Scielzo, C., *et al.* HS1 has a central role in the trafficking and homing of leukemic B cells. *Blood* **116**, 3537-3546 (2010).
79. Scielzo, C., *et al.* HS1 protein is differentially expressed in chronic lymphocytic leukemia patient subsets with good or poor prognoses. *J Clin Invest* **115**, 1644-1650 (2005).
80. Lanasa, M. Novel Insights into the Biology of CLL. *Hematology Am Soc Hematol Educ Program.*, 70-76 (2010).

Figures

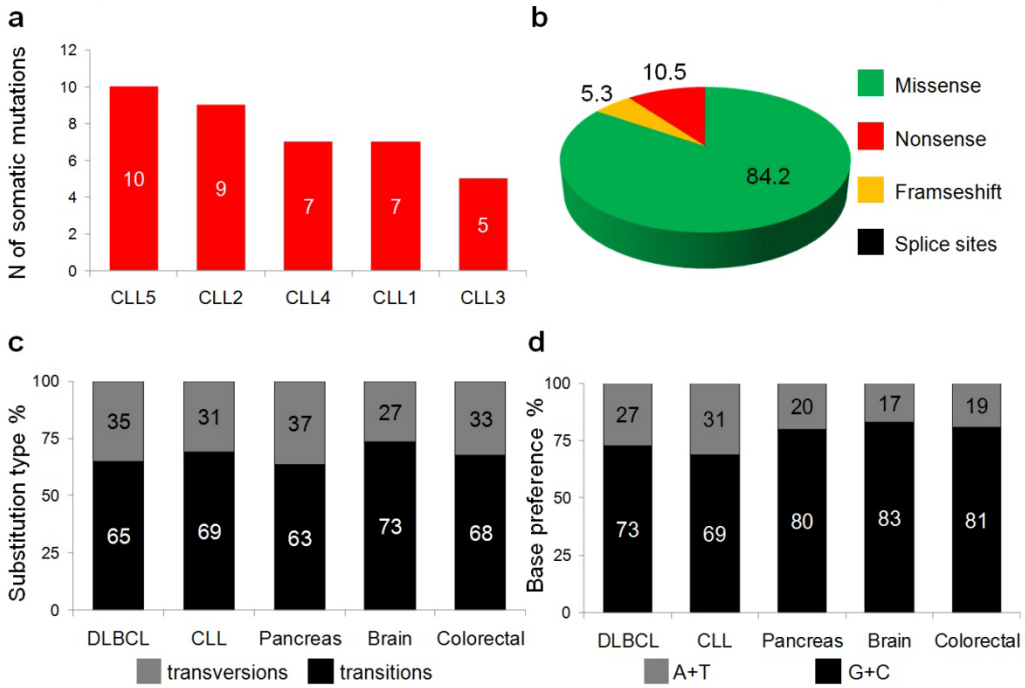


Figure 1. Mutational load and mutation spectrum in CLL a, Histogram representing the number of somatic non-synonymous mutations per CLL case identified through the whole exome sequencing analysis. **b,** Frequency of distinct mutation types identified through the Whole Exome Sequencing analysis. **c and d,** Histograms representing the comparison between the substitution pattern (c) and the nucleotide targeting (d) of the missense mutations observed in the exomes of 5 CLL and 7 DLBCL cases analyzed through the whole Exome sequencing and in the exomes of other solid tumor types analyzed by Sanger sequencing (N=24 pancreatic, N=21 brain and N=11 colorectal cancers)[30]. Only non-synonymous missense mutations are considered for the CLL and DLBCL cases, while both synonymous and non-synonymous missense mutations are considered for the other tumor types analyzed.

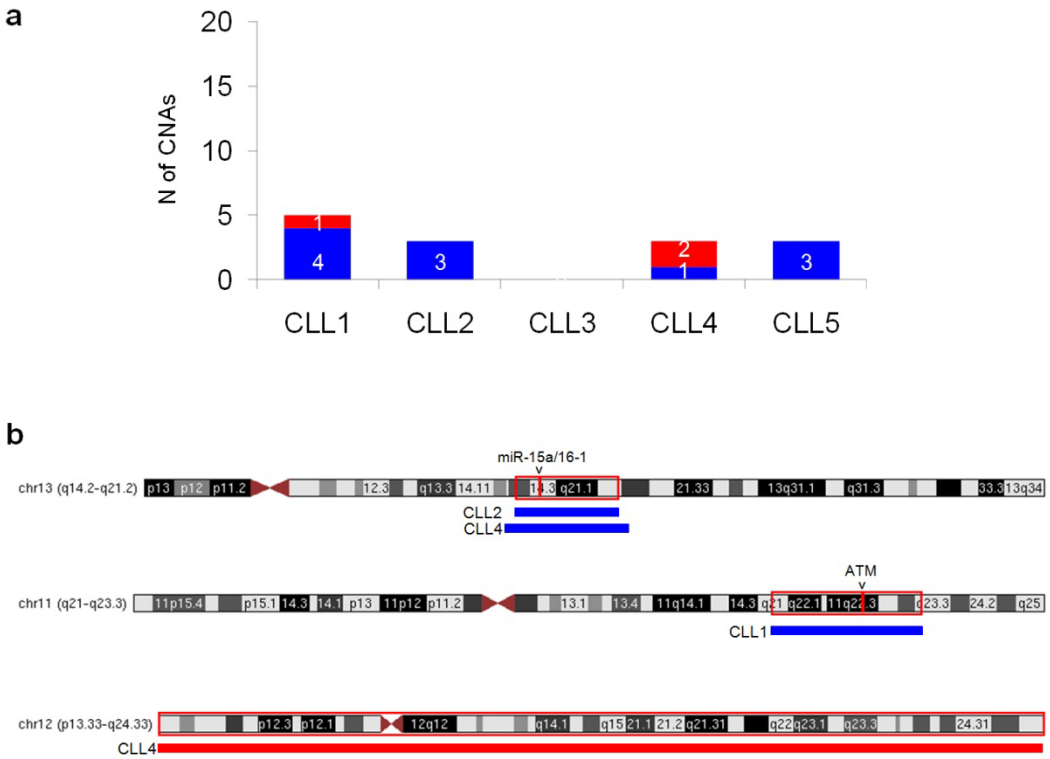


Figure 2. Load of Copy Number Aberrations in CLL a, Histogram representing the number of copy number aberrations (CNAs) per CLL case identified through Genome-Wide SNP Array analysis. Copy number losses are shown in blue, while copy number gains are shown in red. **b,** Schematic representation of the known CLL-associated lesions.

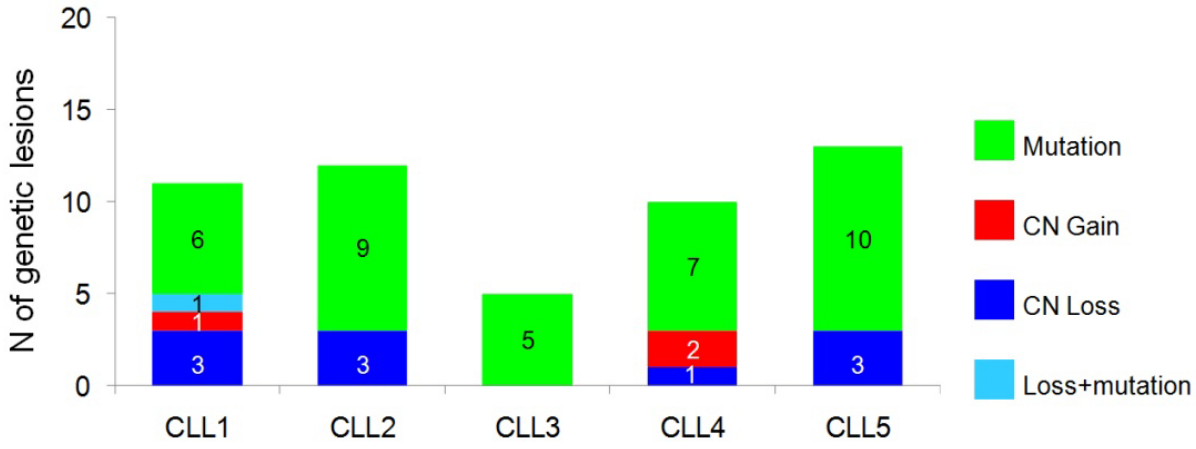


Figure 3. Overall load of genetic lesions in CLL. Histogram representing the number of somatic non-synonymous mutations per CLL case identified through the Whole Exome Sequencing analysis and the number of copy number aberrations identified through the SNP Array analysis. Abbreviations: CN= copy number.

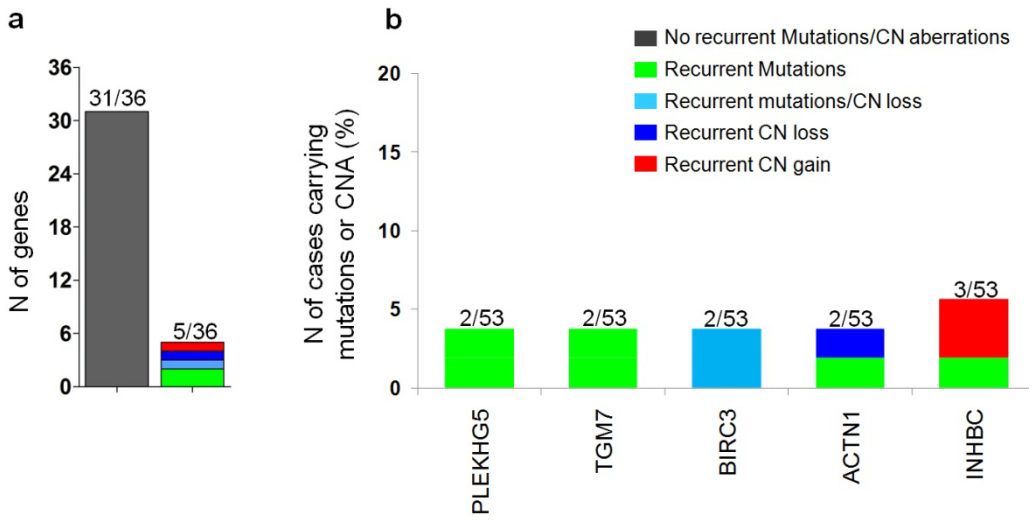


Figure 4. Screening phase: mutational and copy number analysis revealed low recurrence of alterations in the genes identified through whole exome sequencing.

Tables

Table 1. CLL patients' characteristics

Characteristic	CLL patient ID				
	CLL1	CLL2	CLL3	CLL4	CLL5
Age, years	67	71	46	69	63
Sex	F	F	M	F	M
Rai stage	0	IV	II	0	II
Lymphocytes, $x10^9$/Liter	38	124	36	16.7	68.5
Hb, g/dl	13.4	9.5	14.4	13.1	12
Platelets, $x10^9$/Liter	249	159	166	155	135
B2M, mg/l	2.7	6.3	2.2	3.3	3.2
LDH, U/l	551	500	340	480	372
Time to first therapy, Months	87.5	treatment started at diagnosis	33.6	6.1	9.3
Transformation to DLBCL	29/10/07	no	no	no	no
Last follow-up	26/05/09 (death for infection)	6/5/2009 (alive)	6/26/2009 (alive)	6/26/2009 (alive)	6/18/2009 (alive)
IgV_H mutational status	Unmutated	Unmutated	Unmutated	Mutated	Mutated
V gene	1-69*01	3-30*04	1-69*01	4-34*01	3-23*01
D gene	3-3*01	5-24*01	3-16*02	3-22*01	5-5*01
J gene	6*02	6*02	6*02	3*02	4*02
IgVH homology (percentage)	100	100	100	94.38	92.34
Stereotyped HCDR3	Stereotyped	Not stereotyped	Not stereotyped	Not stereotyped	Not stereotyped
FISH^	del11q22-q23	del13q14	Normal	+12	del13q14
ZAP-70 expression*	positive	negative	positive	negative	negative
CD38 expression*	negative	negative	positive	positive	negative
p53 mutational status	Unmutated	Unmutated	Unmutated	Unmutated	Unmutated

* Evaluated by FACS analysis. ZAP-70 expression was considered positive if the percentage of positive leukemic cells was above 20%, CD38 expression was considered positive if the percentage of positive leukemic cells was above 30%. ^based on FISH analysis for del13q, del11q, trisomy12, del17p

Table 2. Somatic mutations CLL specific

Sample ID	Gene ID	Mutation type	Chromosomal position^	Base change	Aminoacid change	Affected domain	Polyphen-2 Prediction	Score	Expression[&]
CLL1-T	ACTN1	missense	chr14:68422007	C>T	A425T	spectrin-2	benign	0	A
CLL1-T	AP1G2	nonsense	chr14:23102859	G>A	R380*	nd	altering	N/A	P
CLL1-T	BBS2	nonsense	chr16:55105893	G>T	Y106*	nd	altering	N/A	P
CLL1-T	BIRC3	frameshift deletion	chr11:101706960	Δ 32bases	E368fs	interdomain (BIR3 and CARD)	altering	N/A	P
CLL1-T	FRMPD1	missense	chr9:37736153	C>T	P1375L	nd	benign	0.025	A
CLL1-T	FSCN3	nonsense	chr7:127025782	C>T	R340*	actin-crosslinking domain	altering	N/A	A
CLL1-T	GOLM1	missense	chr9:87882296	G>A	R54C	lumenal	altering	0.948	A
CLL2-T	ABCA7	missense	chr19:996094	G>A	G437S	extracellular	altering	0.98	P (377.6)
CLL2-T	ALPK3	missense	chr15:83202310	G>A	G1315S	nd	altering	0.949	A (5.3)
CLL2-T	DNAH2[$]	missense	chr17:7618928	T>C[$]	L1543P[$]	interdomain (TPR1-TPR2)	altering	0.999	NA
CLL2-T	GPRASP1	missense	chrX:101798114	A>T	I873F	Glu-rich region	altering	0.569	P (1220.8)
CLL2-T	INHBC	missense	chr12:56115006	G>A	G24S	propeptide (AA19-236)	benign	0.019	P (73.7)
CLL2-T	SCGN[$]	missense	chr6:25809474	G>A	V255M	calcium binding	benign	0.002	A
CLL2-T	WHSC1	missense	chr4:1932599	G>A	E1099K	SET domain	altering	0.982	P (290.5)
CLL2-T	YBX1	missense	chr1:42935456	G>A	A226T	nd	benign	0	P (7900.4)
CLL2-T	ZNF300	missense	chr5:150255232	G>C	Q588E	ZF N.12 (AA 577-599)	altering	0.984	A (15.3)
CLL3-T	EMX2	nonsense	chr10:119297581	C>T	Q203*	homeobox DNA binding	altering	N/A	NA
CLL3-T	MPP2	missense	chr17:39312768	C>T	E447K	guanylate kinase-like	altering	0.418	A (93.5)
CLL3-T	OR4D11	missense	chr11:59028199	C>A	T192N	extracellular	altering	0.924	NA
CLL3-T	SNX9	frameshift deletion	chr6:158250968	Δ ACTG	F290fs	PX domain	altering	N/A	P (829.5)
CLL3-T	TGM7	missense	chr15:41366909	G>T	N242K	nd	altering	0.978	A (9.5)
CLL4-T	ASPHD1	missense	chr16:29820259	C>T	R156C	lumenal	altering	0.974	A (97.3)
CLL4-T	FAM123B	missense	chrX:63329239	T>A	E218V	nd	altering	0.903	A (12.1)
CLL4-T	HIST1H2AM	missense	chr6:27968622	G>C	N95K	histone-fold	altering	0.983	P (180.8)
CLL4-T	MOG	missense	chr6:29735146	C>T	R54C	Ig-like V-type, EC	altering	0.999	A (30.2)
CLL4-T	OR5AN1	missense	chr11:58888757	T>G	S84A	extracellular	altering	0.508	NA
CLL4-T	PLEKHG5	missense	chr1:6453687	A>T	Y525N	nd	altering	1	A (25.3)
CLL4-T	SLC1A1	missense	chr9:4575404	A>G	D474G	extracellular	benign	0.012	A (64.4)
CLL5-T	ACTL6A	missense	chr3:180770679	C>T	P78L	nd	altering	0.979	P
CLL5-T	ADCY8	missense	chr8:131862024	T>C	R1184G	cytoplasmic	benign	0.174	NA
CLL5-T	GDF2	missense	chr10:48033768	G>A	T369M	chain (AA320-429)	altering	0.999	A
CLL5-T	HEXIM1	missense	chr17:40582696	C>T	A119V	nd	benign	0	P
CLL5-T	MYD88	missense	chr3:38157645	T>C	L265P	TIR domain	altering	0.999	P
CLL5-T	NHSL2	missense	chrX:71275588	G>A	G489S	nd	altering	0.906	A
CLL5-T	PUM2	missense	chr2:20317179	T>C	N983S	Pumilio 4	benign	0	P
CLL5-T	SORCS1	missense	chr10:108327217	G>C	P1153R	cytoplasmic	benign	0.001	P
CLL5-T	ST18	missense	chr8:53240291	T>C	T418A	ZF N.2 (AA 409-439)	altering	0.988	NA
CLL5-T	ZNF644	missense	chr1:91176521	T>G	Y993S	ZF N.6 (AA 963-987)	altering	0.995	P

^ Numbering according to NCBI Build 36.1 (hg18).

[$] Subclonal mutation.

& Expression refers to the mutated case or to an independent panel of 16 CLL samples. A gene is considered expressed if the probe is called Present in the mutated case or if it has >90% of Present calls in the 16 CLL cases panel. The value reported in brackets corresponds tot he signal value from the most highly expressed probe in the U133 Plus 2.0 platform.

Abbreviations: nd, not determined; N/A, not available; NA, probe not informative or not present; A, Absent call; P, Present call.

Supplementary Material

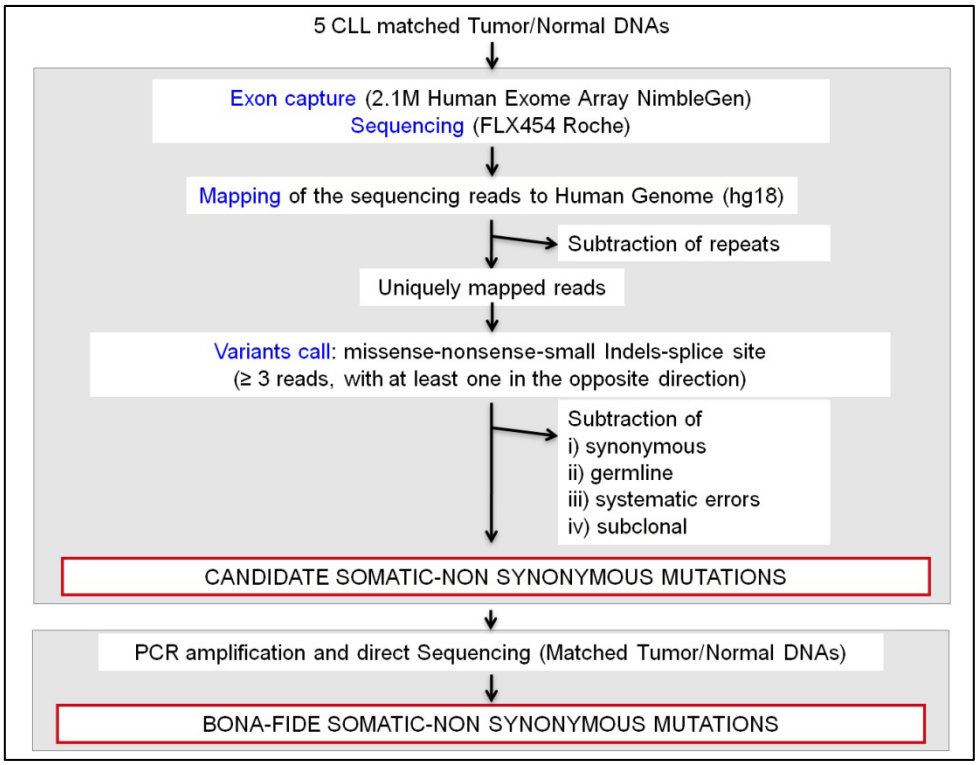


Figure S1. Computational pipeline used for the analysis of 454 Whole Exome Sequencing results and experimental validation

Table S1. List of genes found mutated in the CLL whole exome sequencing analysis

Gene Symbol	Chromosomal location	RefSeq	N. of exons	Coding exons
ABCA7	19p13.3	NM_019112.3	47	Exon 2-47
ACTL6A	3q26.33	NM_004301.3	14	Exon 1-14
ACTN1	14q22-q24	NM_001130004.1	22	Exon 1-22
ADCY8	8q24	NM_001115.2	18	Exon 1-18
ALPK3	15q25.2	NM_020778.4	14	Exon 1-14
AP1G2	14q11.2	NM_003917.2	22	Exon 2-22
ASPHD1	16p11.2	NM_181718.3	3	Exon 1-3
BBS2	16q21	NM_031885.3	18	Exon 1-18
BIRC3	11q22	NM_182962.1	10	Exon 3-10
DNAH2^	17p13.1	NM_020877.2	85	Exon 1-85
EMX2	10q26.1	NM_004098.3	3	Exon 1-3
FAM123B	Xq11.2	NM_152424.3	2	Exon 2
FRMPD1	9p13.2	NM_014907.2	15	Exon 1-15
FSCN3	7q31.3	NM_020369.1	7	Exon 1-6
GDF2	10q11.22	NM_016204.1	2	Exon 1-2
GOLM1	9q21.33	NM_177937.2	10	Exon 2-10
GPRASP1	Xq22.1	NM_014710.4	5	Exon 5
HEXIM1	17q21.31	NM_006460.2	1	Exon 1
HIST1H2AM	6p22-p21.3	NM_003514.2	1	Exon 1
INHBC	12q13.1	NM_005538.2	2	Exon 1-2
MOG	6p22.1	NM_002433.4	10	Exon 1-10
MPP2	17q12-q21	NM_005374.3	13	Exon 2-13
MYD88	3p22	NM_002468.4	5	Exon 1-5
NHSL2	Xq13.1	NM_001013627.2	8	Exon 1-8
OR4D11	11q12.1	NM_001004706.1	1	Exon 1
OR5AN1	11q12.1	NM_001004729.1	1	Exon 1
PLEKHG5	1p36.31	NM_198681.2	22	Exon 1-22
PUM2	2p22-p21	NM_015317.1	20	Exon 1-20
SCGN^	6p22.3-p22.1	NM_006998.3	11	Exon 1-11
SLC1A1	9p24	NM_004170.4	12	Exon 1-12
SNX9	6q25.1-q26	NM_016224.3	18	Exon 1-18
SORCS1	10q23-q25	NM_001013031.1	27	Exon 1-27
ST18	8q11.23	NM_014682.2	26	Exon 7-26
TGM7	15q15.2	NM_052955.2	13	Exon 1-13
WHSC1	4p16.3	NM_001042424.2	22	Exon 1-22
YBX1	1p34	NM_004559.3	8	Exon 1-7
ZNF300	5q33.1	NM_052860.1	6	Exon 3-6
ZNF644	1p22.2	NM_201269.1	6	Exon 2-6

^ DNAH2 and SCGN were affected by a subclonal mutation in the discovery panel, therefore they were not resequenced.

Table S2. Results of 454 sequencing and mapping after whole genome capture

Sample ID	CLL1-T	CLL1-N	CLL2-T	CLL2-N	CLL3-T	CLL3-N	CLL4-T	CLL4-N	CLL5-T	CLL5-N
Target Region Coverage (%)										
1X	95.8	95.9	96.5	95.8	95.7	95.8	95.7	96.2	96.4	96.5
≥5X	74.6	77.2	86.3	80.8	76.5	78.4	75.8	78.6	82.2	83.7
Mean Depth	8.6	9.1	11.8	11.2	9.1	11.1	9.1	9	10.3	11.6
N of bases										
Total N	776458389	797781187	1059708593	1041979809	810937332	978726730	789862934	767859125	896781407	1015206558
Mapped to the human genome (hg18) (%)	773352555 (99.6)	797701409 (99.99)	1057483205 (99.79)	1039062266 (99.72)	808666707 (99.72)	971875643 (99.3)	787256386 (99.67)	765555548 (99.7)	893822028 (99.67)	1012770062 (99.76)
N of reads										
Total N	2404306	2514088	2701957	2664987	2419625	2549794	2175807	2410928	2466472	2715180
Mapped to the human genome (hg18) (%)	2374973 (98.78)	2490456 (99.06)	2684394 (99.35)	2643134 (99.18)	2399542 (99.17)	2531945 (99.30)	2157095 (99.14)	2389230 (99.10)	2440574 (98.95)	2695359 (99.27)
Repeats (%)	102647 (4.27)	102628 (4.08)	102731 (3.8)	110231 (4.14)	104315 (4.31)	104091 (4.08)	91895 (4.22)	110654 (4.59)	119355 (4.84)	111676 (4.11)
Unique reads in target exome (%)	1556726 (68.6)	1679964 (70.4)	2008402 (77.8)	1870633 (73.9)	1664769 (72.6)	1849789 (76.2)	1560948 (75.7)	1633385 (71.7)	1746590 (75.3)	1969017 (76.3)
N of variants										
Total	9730	10162	12457	11392	10459	10925	10201	10956	12459	12367
Known*	9163	9531	11675	10652	9744	10240	9603	10244	11556	11520
Novel**	567	631	782	740	715	685	598	712	903	847
N of tumor-specific non synonymous variants										
Before filtering	86	-	104	-	191	-	62	-	124	-
Filtered out^	43	-	48	-	115	-	19	-	65	-
Candidate Somatic	**43**	-	**56**	-	**76**	-	**43**	-	**59**	-

* Germline SNPs reported the dbSNP_129 database.

** Before cross-comparison with paired normal DNA.

^ N of variants that were excluded because: i) present in <20% of the reads and/or ii) Indels affecting polynucleotide stretches/repeats.

Table S3. Candidate somatic mutations and validation results

Sample ID	gene	Chromosomal position^	Original base	Base change	Variant frequency (%)	Original Aminoacid	Aminoacid change	Normal DNA[N]	Tumor DNA[N]
CLL1	ACTN1	chr14:68422007-68422007	C	T	67%	A	T	NEG	POS
CLL1	AP1G2	chr14:23102859-23102859	G	A	30%	R	*	NEG	POS
CLL1	BBS2	chr16:55105893-55105893	G	T	43%	Y	*	NEG	POS
CLL1	BIRC3	chr11:101706960-101706991	Δ32 bp	-	30%	EDHSEDAIMMN	-	NEG	POS
CLL1	FRMPD1	chr9:37736153-37736153	C	T	75%	P	L	NEG	POS
CLL1	FSCN3	chr7:127025782-127025782	C	T	50%	R	*	NEG	POS
CLL1	GOLM1	chr9:87882296-87882296	G	A	42%	R	C	NEG	POS
CLL1	ACOXL	chr2:111279346-111279346	C	T	57%	S	L	POS	POS
CLL1	ADAMTS18	chr16:75888736-75888736	T	C	57%	K	E	POS	POS
CLL1	ALG10B	chr12:36997068-36997068	A	T	57%	M	L	POS	POS
CLL1	ARSF	chrX:3031931-3031931	G	A	67%	A	T	POS	POS
CLL1	C2orf7	chr2:73310094-73310094	-	22 bp	100%	splice site mutation	-	POS	POS
CLL1	CLCC1	chr1:109278985-109278985	G	C	80%	P	A	POS	POS
CLL1	CLDN19	chr1:42974163-42974163	C	T	80%	R	Q	POS	POS
CLL1	COL11A1	chr1:103200900-103200900	G	T	67%	P	Q	POS	POS
CLL1	CPA5	chr7:129777170-129777170	T	C	50%	M	T	POS	POS
CLL1	CSF3R	chr1:36704634-36704634	C	T	44%	E	K	POS	POS
CLL1	DNAH10	chr12:122854217-122854217	G	A	36%	V	I	POS	POS
CLL1	ESF1	chr20:13643559-13643560	CT	-	57%	TE	NX	POS	POS
CLL1	FBXW10	chr17:18616527-18616527	C	T	60%	T	M	POS	POS
CLL1	FGGY	chr1:59846076-59846076	C	T	100%	P	L	POS	POS
CLL1	GLI1	chr12:56150964-56150964	C	T	64%	P	L	POS	POS
CLL1	HFE	chr6:26199144-26199144	T	C	50%	F	S	POS	POS
CLL1	IL10RA	chr11:117371606-117371606	C	T	78%	R	W	POS	POS
CLL1	OR6C76	chr12:54107044-54107044	T	A	40%	I	N	POS	POS
CLL1	PSD2	chr5:139173992-139173992	C	T	83%	S	L	POS	POS
CLL1	PSG6	chr19:48103063-48103063	G	A	50%	S	F	POS	POS
CLL1	RNF122	chr8:33526501-33526501	C	T	50%	V	M	POS	POS
CLL1	SIGLEC11	chr19:55155482-55155482	T	G	50%	K	Q	POS	POS
CLL1	SMPD3	chr16:66962580-66962580	G	A	50%	R	C	POS	POS
CLL1	SOAT2	chr12:51799214-51799214	C	T	83%	R	C	POS	POS
CLL1	TBX18	chr6:85503844-85503844	T	G	67%	N	H	POS	POS
CLL1	TNC	chr9:116875721-116875721	G	A	43%	R	C	POS	POS
CLL1	TNIP3	chr4:122287766-122287766	T	G	60%	E	A	POS	POS
CLL1	UGT2A1	chr4:70547376-70547376	A	T	80%	Y	*	POS	POS
CLL1	AQP7	chr9:33376144-33376146	GAC	AAA	27%	V	F	NEG	NEG
CLL1	AQP7	chr9:33376167-33376167	G	C	27%	Q	E	NEG	NEG
CLL1	C5orf23	chr5:32825589-32825590	GT	-	33%	V	X	NEG	NEG
CLL1	DYRK4	chr12:4578484-4578484	A	G	23%	N	S	NEG	NEG
CLL1	RASL10B	chr17:31091617-31091617	A	G	37%	D	G	NEG	NEG
CLL1	SFRS4	chr1:29347895-29347895	G	A	62%	R	C	NEG	NEG
CLL1	SLC16A6	chr17:63778676-63778676	A	G	21%	L	P	NEG	NEG
CLL1	DFFB	chr1:3778921-3778921	T	C	80%	L	S	NA	NA
CLL2	ABCA7	chr19:996094-996094	G	A	40%	G	S	NEG	POS
CLL2	ALPK3	chr15:83202310-83202310	G	A	33%	G	S	NEG	POS
CLL2	DNAH2	chr17:7618928-7618928	T	C	20%	L	P	NEG	POS
CLL2	GPRASP1	chrX:101798114-101798114	A	T	23%	I	F	NEG	POS
CLL2	INHBC	chr12:56115006-56115006	G	A	47%	G	S	NEG	POS
CLL2	SCGN	chr6:25809474-25809474	G	A	18%	V	M	NEG	POS
CLL2	WHSC1	chr4:1932599-1932599	G	A	33%	E	K	NEG	POS
CLL2	YBX1	chr1:42935456-42935456	G	A	43%	A	T	NEG	POS
CLL2	ZNF300	chr5:150255232-150255232	G	C	50%	Q	E	NEG	POS
CLL2	ABCB10	chr1:227720676-227720676	T	C	50%	N	S	POS	POS
CLL2	AHNAK	chr11:62043242-62043242	C	T	57%	V	M	POS	POS
CLL2	ATAD3B	chr1:1420848-1420848	C	T	50%	R	C	POS	POS
CLL2	C1orf9	chr1:170838515-170838515	T	C	100%	M	T	POS	POS
CLL2	CELSR2	chr1:109597265-109597265	T	A	87%	L	H	POS	POS
CLL2	CHD1L	chr1:145232003-145232003	A	G	30%	I	V	POS	POS
CLL2	CSPG4	chr15:73756390-73756390	C	T	43%	R	Q	POS	POS
CLL2	DCC	chr18:48846529-48846529	A	G	71%	K	R	POS	POS
CLL2	DMPK	chr19:50972561-50972561	G	C	58%	P	R	POS	POS
CLL2	FN1	chr2:215957834-215957834	C	G	43%	E	Q	POS	POS
CLL2	FOLH1B	chr11:89053442-89053442	G	A	56%	V	I	POS	POS
CLL2	IL23R	chr1:67421184-67421184	G	A	67%	G	R	POS	POS
CLL2	ITGA9	chr3:37670282-37670282	C	T	37%	S	F	POS	POS
CLL2	LGI2	chr4:24641233-24641233	C	T	60%	G	S	POS	POS
CLL2	NMNAT1	chr1:9954836-9954836	A	G	100%	splice site mutation		POS	POS
CLL2	OR5P2	chr11:7774963-7774963	-	21 bp	67%	VILFMIIL	VILFMIIY	POS	POS
CLL2	PAPSS1	chr4:108772261-108772261	G	A	37%	R	C	POS	POS
CLL2	SCRIB	chr8:144945463-144945463	G	A	80%	P	L	POS	POS
CLL2	SIGLEC12	chr19:56696608-56696608	-	T	100%	F	L	POS	POS
CLL2	SLC26A9	chr1:204167649-204167649	C	T	71%	V	M	POS	POS
CLL2	STK17A	chr7:43630955-43630957	TTT	-	62%	F	X	POS	POS
CLL2	SULT1C3	chr2:108247722-108247722	C	T	54%	R	W	POS	POS
CLL2	TARS2	chr1:148736685-148736685	C	T	50%	T	M	POS	POS
CLL2	TBP	chr6:170712977-170712978	GC	-	54%	QQ	XS	POS	POS
CLL2	TNFSF11	chr13:42046546-42046546	C	G	75%	P	R	POS	POS
CLL2	TTN	chr2:179322369-179322369	T	C	62%	R	G	POS	POS
CLL2	WASF2	chr1:27611737-27611737	G	A	60%	P	L	POS	POS
CLL2	ZNF528	chr19:57601039-57601039	C	T	80%	A	V	POS	POS
CLL2	ZNF610	chr19:57561243-57561243	A	G	36%	K	R	POS	POS
CLL2	BRUNOL5	chr19:3244421-3244421	-	CA	32%	K	Q	NEG	NEG
CLL2	C18orf32	chr18:45263949-45263950	TA	GCAC	21%	splice site mutation		NEG	NEG
CLL2	CCDC144NL	chr17:20710594-20710595	CC	AG	33%	QA	HS	NEG	NEG
CLL2	CELA3A	chr1:22204832-22204832	G	C	31%	E	D	NEG	NEG
CLL2	CELA3A	chr1:22204828-22204828	A	G	29%	E	G	NEG	NEG
CLL2	CELA3A	chr1:22204779-22204779	A	G	25%	N	D	NEG	NEG
CLL2	CELA3A	chr1:22204783-22204783	T	G	25%	L	R	NEG	NEG
CLL2	EXOC7	chr17:71605658-71605658	A	G	21%	S	P	NEG	NEG
CLL2	FANCD2	chr3:10063408-10063412	GTAAG	T	69%	splice site mutation		NEG	NEG
CLL2	HMCN1	chr1:184323935-184323935	A	G	21%	I	V	NEG	NEG
CLL2	HPS5	chr11:18269933-18269935	TCT	C	25%	KE	KR	NEG	NEG
CLL2	HTR3B	chr11:113321872-113321872	A	G	30%	T	A	NEG	NEG
CLL2	IGLL1	chr22:22245610-22245610	A	T	62%	M	K	NEG	NEG
CLL2	MADCAM1	chr19:452695-452695	C	T	60%	P	S	NEG	NEG
CLL2	PCDHB16	chr5:140543897-140543897	-	GCT	46%	F	A	NEG	NEG
CLL2	SPRR3	chr1:151242375-151242377	CCC	-	33%	VP	X*	NEG	NEG
CLL2	SPRR3	chr1:151242382-151242383	CC	-	33%	P	X	NEG	NEG
CLL2	SPRR3	chr1:151242368-151242369	CC	-	29%	T	X	NEG	NEG
CLL2	PRKRA	chr2:179016360-179016360	A	G	60%	splice site mutation		NA	NA
CLL3	EMX2	chr10:119297581-119297581	C	T	38%	Q	*	NEG	POS
CLL3	MPP2	chr17:39312768-39312768	C	T	33%	E	K	NEG	POS

CLL3	AADAT	chr4:171222536-171222536	T	C	57%	H	R	POS	POS
CLL3	ADAD1	chr4:123553204-123553204	A	C	71%	I	L	POS	POS
CLL3	AHNAK	chr11:62060140-62060140	-	CCT	50%	EK	ER	POS	POS
CLL3	ANXA7	chr10:74818073-74818073	T	C	87%	splice site mutation	-	POS	POS
CLL3	ARHGAP20	chr11:109956316-109956316	C	A	67%	R	L	POS	POS
CLL3	ASB15	chr7:123054423-123054423	G	-	75%	D	X	POS	POS
CLL3	ASZ1	chr7:116810286-116810286	G	A	50%	H	Y	POS	POS
CLL3	ATP13A5	chr3:194475681-194475681	G	C	75%	D	E	POS	POS
CLL3	B4GALNT2	chr17:44565468-44565468	G	A	83%	E	K	POS	POS
CLL3	C1orf146	chr1:92483823-92483823	A	G	44%	splice site mutation	-	POS	POS
CLL3	C1orf189	chr1:152444707-152444707	G	A	67%	T	M	POS	POS
CLL3	C2orf64	chr2:98583644-98583644	G	C	56%	splice site mutation	-	POS	POS
CLL3	C2orf67	chr2:210727242-210727242	G	A	50%	P	S	POS	POS
CLL3	CASQ2	chr1:116112577-116112577	G	A	25%	L	F	POS	POS
CLL3	CCDC69	chr5:150547178-150547178	G	T	62%	L	I	POS	POS
CLL3	CCDC90B	chr11:82674650-82674650	T	C	67%	Q	R	POS	POS
CLL3	CCL28	chr5:43417717-43417717	C	G	60%	splice site mutation	-	POS	POS
CLL3	CCNL1	chr3:158353642-158353642	A	G	33%	splice site mutation	-	POS	POS
CLL3	CEP192	chr18:13082508-13082508	C	G	50%	A	G	POS	POS
CLL3	DGKG	chr3:187520909-187520909	C	T	62%	E	K	POS	POS
CLL3	DNAH5	chr5:13955177-13955177	A	C	23%	L	W	POS	POS
CLL3	DNAH6	chr2:84658044-84658044	G	C	80%	splice site mutation	-	POS	POS
CLL3	GUCY2C	chr12:14670016-14670016	G	T	67%	Q	K	POS	POS
CLL3	HNRNPU	chr1:243088036-243088036	G	A	83%	P	L	POS	POS
CLL3	IL25	chr14:22914916-22914916	G	A	100%	R	H	POS	POS
CLL3	INTS7	chr1:210222664-210222664	A	G	37%	I	T	POS	POS
CLL3	KCNJ12	chr17:21259680-21259680	G	A	33%	G	S	POS	POS
CLL3	KCNJ12	chr17:21259672-21259672	C	A	33%	T	N	POS	POS
CLL3	KRI1	chr19:10526974-10526974	T	C	60%	T	A	POS	POS
CLL3	KRT28	chr17:36209533-36209533	A	G	43%	C	R	POS	POS
CLL3	LAPTM5	chr1:30979333-30979333	C	T	75%	E	K	POS	POS
CLL3	LIMS2	chr2:128117121-128117121	C	T	60%	splice site mutation	-	POS	POS
CLL3	LRRC37B	chr17:27372319-27372319	C	T	100%	T	M	POS	POS
CLL3	NDRG2	chr14:20557950-20557950	G	A	44%	P	L	POS	POS
CLL3	NPHS1	chr19:41033151-41033151	T	A	62%	N	I	POS	POS
CLL3	PDCD5	chr19:37764006-37764006	C	T	80%	splice site mutation	-	POS	POS
CLL3	PDE4DIP	chr1:143732507-143732507	T	C	22%	D	G	POS	POS
CLL3	PKDREJ	chr22:45036545-45036545	T	C	39%	R	G	POS	POS
CLL3	PSG1	chr19:48067758-48067758	C	G	71%	splice site mutation	-	POS	POS
CLL3	PSG8	chr19:47959982-47959982	G	-	56%	S	X	POS	POS
CLL3	RANGAP1	chr22:39980284-39980284	G	A	80%	R	W	POS	POS
CLL3	SAFB2	chr19:5549801-5549801	C	T	100%	splice site mutation	-	POS	POS
CLL3	SCRIB	chr8:144945463-144945463	G	A	80%	P	L	POS	POS
CLL3	SFPQ	chr1:35425233-35425233	C	T	29%	V	I	POS	POS
CLL3	SFXN1	chr5:174851845-174851845	G	A	67%	E	K	POS	POS
CLL3	SGIP1	chr1:66898791-66898791	G	C	56%	S	T	POS	POS
CLL3	SNUPN	chr15:73700321-73700321	G	A	80%	R	C	POS	POS
CLL3	SPOCK2	chr10:73497374-73497374	T	C	33%	splice site mutation	-	POS	POS
CLL3	STAB1	chr3:52512397-52512397	C	A	80%	splice site mutation	-	POS	POS
CLL3	TCHHL1	chr1:150326644-150326644	C	T	67%	splice site mutation	-	POS	POS
CLL3	TLE1	chr9:83416697-83416697	G	A	75%	splice site mutation	-	POS	POS
CLL3	TM4SF18	chr3:150525483-150525483	A	T	57%	I	N	POS	POS
CLL3	TTC12	chr11:112739835-112739835	G	A	89%	E	K	POS	POS
CLL3	ANP32B	chr9:99817538-99817539	CC	TG	37%	splice site mutation	-	NEG	NEG
CLL3	ASPN	chr9:94276845-94276845	-	TCA	100%	E	D	NEG	NEG
CLL3	C11orf1	chr11:111255504-111255504	T	C	36%	S	P	NEG	NEG
CLL3	CCBL2	chr1:89221886-89221886	C	G	44%	G	A	NEG	NEG
CLL3	CCBL2	chr1:89221978-89221978	T	C	44%	I	M	NEG	NEG
CLL3	FANCD2	chr3:10063410-10063412	AAG	-	46%	splice site mutation	-	NEG	NEG
CLL3	OR4C3	chr11:48303934-48303934	C	T	43%	T	I	NEG	NEG
CLL3	RBMX	chrX:135784074-135784074	A	G	33%	Y	H	NEG	NEG
CLL3	RBMX	chrX:135784128-135784128	G	C	40%	R	G	NEG	NEG
CLL3	RBMX	chrX:135784172-135784172	C	G	37%	R	P	NEG	NEG
CLL3	SIRPA	chr20:1844059-1844059	G	A	44%	V	M	NEG	NEG
CLL3	SIRPA	chr20:1844060-1844060	T	C	44%	V	A	NEG	NEG
CLL3	SLC25A18	chr22:16443859-16443859	T	C	25%	V	A	NEG	NEG
CLL3	SLCO1C1	chr12:20745546-20745546	T	-	82%	Y	X	NEG	NEG
CLL3	TTN	chr2:179339364-179339364	A	G	27%	L	P	NEG	NEG
CLL3	WHSC1	chr4:1872433-1872433	-	CG	30%	R	P	NEG	NEG
CLL3	XAGE3	23:52908370-52908370	A	G	44%	splice site mutation	-	NEG	NEG
CLL3	ZNF468	chr19:58036689-58036689	A	G	45%	F	L	NEG	NEG
CLL4	ASPHD1	chr16:29820259-29820259	C	T	33%	R	C	NEG	POS
CLL4	FAM123B	chrX:63329239-63329239	T	A	54%	E	V	NEG	POS
CLL4	HIST1H2AM	chr6:27968622-27968622	G	C	33%	N	K	NEG	POS
CLL4	MOG	chr6:29735146-29735146	C	T	71%	R	C	NEG	POS
CLL4	OR5AN1	chr11:58888757-58888757	T	G	45%	S	A	NEG	POS
CLL4	PLEKHG5	chr1:6453687-6453687	A	T	33%	Y	N	NEG	POS
CLL4	SLC1A1	chr9:4575404-4575404	A	G	56%	D	G	NEG	POS
CLL4	CROCC	chr1:17144624-17144624	G	A	80%	R	H	POS	POS
CLL4	LBP	chr20:36422824-36422824	A	G	67%	Q	R	POS	POS
CLL4	PITPNM1	chr11:67024205-67024205	C	T	100%	D	N	POS	POS
CLL4	PRMT10	chr4:148801643-148801643	G	T	75%	splice site mutation	-	POS	POS
CLL4	ACE2	chrX:15490022-15490022	G	A	57%	A	V	POS	POS
CLL4	ACSM5	chr16:20356096-20356096	G	A	50%	R	Q	POS	POS
CLL4	ACTN4	chr19:43906961-43906961	C	G	60%	S	R	POS	POS
CLL4	ASPA	chr17:3349031-3349031	G	A	50%	V	M	POS	POS
CLL4	ATRIP	chr3:48476528-48476528	G	A	43%	M	I	POS	POS
CLL4	C5orf42	chr5:37218659-37218659	G	A	100%	P	L	POS	POS
CLL4	C8orf45	chr8:67949187-67949187	C	T	56%	A	V	POS	POS
CLL4	CLN5	chr13:76464091-76464091	C	T	71%	R	C	POS	POS
CLL4	CSPG4	chr15:73768482-73768482	C	T	57%	R	Q	POS	POS
CLL4	CSPG4	chr15:73768492-73768492	C	T	62%	V	M	POS	POS
CLL4	CST2	chr20:23752663-23752663	T	A	60%	E	D	POS	POS
CLL4	GDAP2	chr1:118256217-118256217	T	C	27%	M	V	POS	POS
CLL4	GDPD4	chr11:76657159-76657159	A	G	60%	I	T	POS	POS
CLL4	KEL	chr7:142368301-142368301	G	A	60%	T	I	POS	POS
CLL4	LUZP1	chr1:23291849-23291849	-	TCCA	50%	ES	EM	POS	POS
CLL4	MAGEB6	chrX:26122398-26122398	G	C	50%	G	R	POS	POS
CLL4	MOCOS	chr18:32021566-32021566	C	A	75%	S	R	POS	POS
CLL4	OR4C3	chr11:48303537-48303538	AA	TG	36%	N	C	POS	POS
CLL4	PCDH15	chr10:55251935-55251935	T	G	75%	M	L	POS	POS
CLL4	PWP2	chr21:44365390-44365390	A	G	50%	T	A	POS	POS

CLL4	REPS1	chr6:139308350-139308350	C	T	57%	R	H	POS	POS
CLL4	RNF213	chr17:75964825-75964825	C	G	43%	L	V	POS	POS
CLL4	TMEM102	chr17:7280627-7280627	C	G	50%	A	G	POS	POS
CLL4	TMEM119	chr12:107510105-107510105	G	A	75%	L	F	POS	POS
CLL4	USP54	chr10:74934731-74934731	G	T	87%	S	R	POS	POS
CLL4	VPS13A	chr9:79210694-79210694	C	A	50%	H	N	POS	POS
CLL4	ZNF768	chr16:30444829-30444829	C	T	60%	G	S	POS	POS
CLL4	FANCD2	chr3:10063408-10063412	GTAAG	T	62%	splice site mutation	-	NEG	NEG
CLL4	RBMX	chrX:135784172-135784172	C	G	36%	R	P	NEG	NEG
CLL4	RBMX	chrX:135784074-135784074	A	G	44%	Y	H	NEG	NEG
CLL4	RBMX	chrX:135784128-135784128	G	C	44%	R	G	NEG	NEG
CLL4	ZXDA	chrX:57953176-57953176	C	G	37%	C	S	NEG	NEG
CLL5	ACTL6A	chr3:180770679-180770679	C	T	57%	P	L	NEG	POS
CLL5	ADCY8	chr8:131862024-131862024	T	C	75%	R	G	NEG	POS
CLL5	GDF2	chr10:48033768-48033768	G	A	60%	T	M	NEG	POS
CLL5	HEXIM1	chr17:40582696-40582696	C	T	56%	A	V	NEG	POS
CLL5	MYD88	chr3:38157645-38157645	T	C	50%	L	P	NEG	POS
CLL5	NHSL2	chrX:71275588-71275588	G	A	100%	G	S	NEG	POS
CLL5	PUM2	chr2:20317179-20317179	T	C	67%	N	S	NEG	POS
CLL5	SORCS1	chr10:108327217-108327217	G	C	44%	P	R	NEG	POS
CLL5	ST18	chr8:53240291-53240291	T	C	50%	T	A	NEG	POS
CLL5	ZNF644	chr1:91176521-91176521	T	G	55%	Y	S	NEG	POS
CLL5	ABHD14B	chr3:51980705-51980705	G	A	67%	R	C	POS	POS
CLL5	ALG1L	chr3:127131026-127131026	C	G	30%	Q	H	POS	POS
CLL5	C5orf42	chr5:37218659-37218659	G	A	36%	P	L	POS	POS
CLL5	CCDC48	chr3:130241246-130241246	G	C	57%	E	D	POS	POS
CLL5	CD1A	chr1:156493314-156493314	G	A	90%	G	D	POS	POS
CLL5	CORO7	chr16:4352706-4352706	G	A	80%	S	L	POS	POS
CLL5	DDHD1	chr14:52689250-52689250	C	G	44%	G	A	POS	POS
CLL5	DENND4C	chr9:19331008-19331008	G	C	50%	G	A	POS	POS
CLL5	DNAH3	chr16:20873669-20873669	C	A	67%	V	F	POS	POS
CLL5	EGFLAM	chr5:38442008-38442008	C	T	50%	R	C	POS	POS
CLL5	FAM65A	chr16:66131909-66131909	T	C	57%	splice site mutation	-	POS	POS
CLL5	GJB4	chr1:34999949-34999949	C	G	37%	C	W	POS	POS
CLL5	GNPDA1	chr5:141364785-141364785	T	A	50%	M	L	POS	POS
CLL5	GPC1	chr2:241050472-241050472	C	T	58%	R	C	POS	POS
CLL5	GSTM5	chr1:110059377-110059377	C	T	30%	R	C	POS	POS
CLL5	KRTAP10-10	chr21:44881845-44881846	AG	CC	50%	E	A	POS	POS
CLL5	LILRA1	chr19:59799035-59799039	GGAGA	TGGGG	50%	GE	WG	POS	POS
CLL5	LOXL3	chr2:74633191-74633191	G	C	100%	P	A	POS	POS
CLL5	NRG2	chr5:139211439-139211439	C	T	56%	R	H	POS	POS
CLL5	OR2T35	chr1:246868233-246868233	G	A	100%	A	V	POS	POS
CLL5	OR2T35	chr1:246868234-246868234	C	T	100%	A	T	POS	POS
CLL5	OR9G1	chr11:56224625-56224625	-	GT	79%	TG	TW	POS	POS
CLL5	POTEB	chr15:19307240-19307240	C	T	33%	G	E	POS	POS
CLL5	QRICH2	chr17:71784862-71784862	C	A	43%	E	*	POS	POS
CLL5	RAB42	chr1:28793016-28793016	A	T	50%	M	L	POS	POS
CLL5	SPNS3	chr17:4337934-4337934	C	G	36%	P	R	POS	POS
CLL5	TUBB8	chr10:83694-83694	T	C	29%	K	R	POS	POS
CLL5	ZNF434	chr16:3380052-3380054	TAG	-	50%	LY	PX	POS	POS
CLL5	ZNF502	chr3:44737863-44737863	A	G	54%	T	A	POS	POS
CLL5	CCDC144NL	chr17:20710458-20710459	GG	AGC	62%	GP	GA	NEG	NEG
CLL5	CDC27	chr17:42569599-42569599	A	G	67%	F	L	NEG	NEG
CLL5	CDC27	chr17:42569632-42569632	A	C	67%	Y	D	NEG	NEG
CLL5	CDC27	chr17:42569647-42569647	G	C	62%	Q	E	NEG	NEG
CLL5	CDC27	chr17:42569650-42569650	T	G	62%	I	L	NEG	NEG
CLL5	CDC27	chr17:42569653-42569653	C	T	62%	A	T	NEG	NEG
CLL5	FANCD2	chr3:10063410-10063412	AAG	-	33%	splice site mutation	-	NEG	NEG
CLL5	KCNN3	chr1:153108868-153108868	A	T	52%	L	H	NEG	NEG
CLL5	LAMB4	chr7:107530730-107530737	CACGCGTA	-	44%	YAC	RSX	NEG	NEG
CLL5	OPA1	chr3:194866847-194866847	-	TA	25%	LR	FK	NEG	NEG
CLL5	OR2T34	chr1:246804287-246804287	C	T	30%	C	Y	NEG	NEG
CLL5	OR2T35	chr1:246868215-246868215	C	T	100%	G	D	NEG	NEG
CLL5	OR51B6	chr11:5329888-5329888	T	-	100%	F	X	NEG	NEG
CLL5	PARP4	chr13:23914762-23914762	G	A	40%	T	I	NEG	NEG
CLL5	PER3	chr1:7812611-7812613	TGA	GGG	29%	MK	RE	NEG	NEG
CLL5	RBMX	chrX:135784074-135784074	A	G	40%	Y	H	NEG	NEG
CLL5	RBMX	chrX:135784128-135784128	G	C	38%	R	G	NEG	NEG
CLL5	RBMX	chrX:135784172-135784172	C	G	31%	R	P	NEG	NEG
CLL5	SLC25A5	chrX:118487672-118487672	G	T	40%	Q	H	NEG	NEG
CLL5	SLC25A5	chr23:118489051-118489051	T	G	75%	splice site mutation	-	NEG	NEG

^ Numbering according to NCBI Build 36.1 (hg18).
# Results after PCR amplification and Sanger sequencing. Abbreviations: POS, mutation confirmed; NEG, mutation not confirmed; NA, not amplified.

Table S4. Validation by PCR amplification and Sanger sequencing of 454 Sequencing results

Sample ID	Validation by PCR amplification and Sanger Sequencing				
	Candidate	Somatic (%)^	Germline (%)^^	False positive (%)^^^	NA (%)˜
CLL1	43	7 (16.3)	28 (65.1)	7 (16.3)	1 (2.3)
CLL2	56	8 (14.3)	29 (51.8)	18 (32.1)	1 (1.8)
CLL3	76	5 (6.6)	53 (69.78)	18 (23.7)	0 (0)
CLL4	43	7 (16.3)	31 (72.1)	5 (11.6)	0 (0)
CLL5	59	10 (16.9)	29 (51.7)	20 (33.9)	0 (0)

^ Positve in the tumoral DNA but not in the paired normal
^^ Positve in paired tumoral and normal DNA
^^^ Negative in both the tumoral and the normal DNA
˜ NA: not amplified

Table S5. Mutation Features

Mutation type	Number
Single bp Substitutions	36
Transitions (%)	25 (69%)
Transversions (%)	11 (31%)
Deletions	2
Insertions	-
Transitions	
G□ A	10
A□ G	1
T□ C	5
C□ T	9
Transversions	
C□ G	0
C□ A	1
G□ T	2
G□ C	3
A□ T	2
A□ C	0
T□ G	2
T□ A	1
Totals for each nucleotide	
G□ N	13
C□ N	10
A□ N	3
T□ N	8

Table S6. Segmentation data from Genome-Wide SNP Analysis in the CLL Discovery panel

			Cytoband			Position			N of genes	N of non-coding RNAs	Copy number
Sample ID	value	chr	start	stop	N SNP	start	stop	kB			
CLL1-T	LOSS	3	p21.1	p21.1	105	52967887	53045061	77	1, SFMBT1	2, CU-1130 mir-575	1.62
CLL1-T	LOSS	4	q21.1	q21.23	5361	77636110	86867769	9,232	41		1.62
CLL1-T	LOSS	11	q21	q23.3	15457	94305511	116417742	22,112	115	7, including mir-34b mir-34c mir-34c-3p mir-34c-5p	1.75
CLL1-T	GAIN	12	q12	q12	46	42696125	42772579	76	1, TMEM117		2.65
CLL1-T	LOSS	16	q13	q13	25	55871973	55918847	47	1, PLPP	1, CU-1244	1.31
CLL2-T	LOSS	1	q25.2	q25.2	84	177354757	177398977	44	1, ABL2		1.61
CLL2-T	LOSS	10	q26.13	q26.13	121	123262361	123357137	95	1, FGFR2		1.67
CLL2-T	LOSS	13	q14.2	q21.2	7450	47169132	60087954	12,919	46	6, including mir-15a mir-16 mir-16-1	1.75
CLL4-T	LOSS	3	p24.3	p24.3	24	23623120	23655331	32	0		1.20
CLL4-T	GAIN	6	q14.1	q14.1	127	78276195	78459151	183	0		2.40
CLL4-T	GAIN	12	p13.33	q24.33	84371	20691	132287718	132,267	983	49	2.30
CLL5-T	LOSS	8	q24.3	q24.3	418	140701232	141264456	563	2, KCNK9 NIBP	1, CU-1339	1.31
CLL5-T	LOSS	12	q23.3	q23.3	75	106929454	107010341	81	0		1.27
CLL5-T	LOSS	13	q14.11	q14.3	6518	40468098	50307044	9,839	56	4, including mir-15a mir-16-1	1.35

Abstracts and publications

- The genetic landscape of CLL genome. Giulia Fabbri, Vladimir Trifonov, Davide Rossi, Oliver T Elliot, Joseph M Chan, Andrea Rinaldi, Ivo Kwee, Francesco Bertoni, Gianluca Gaidano, Laura Pasqualucci, Raul Rabadan, and Riccardo Dalla-Favera (*Paper in preparation*).

- Molecular relapse is highly predictive of clinical recurrence in children with acute lymphoblastic leukemia enrolled in the AIEOP-BFM ALL 2000 protocol. Giulia Fabbri, Maddalena Paganin, Katia Polato, Giarin Emanuela, Nardo Daniel, Elena Barisone, Gianni Cazzaniga, Andrea Biondi , Franco Locatelli , Giuseppe Basso (*Paper in preparation*).

- Inactivating mutations of acetyltransferase genes in B cell lymphoma. Laura Pasqualucci, David Dominguez-Sola, Annalisa Chiarenza, Giulia Fabbri, Adina Grunn, Vladimir Trifonov, Lawryn H. Kasper, Stephanie Lerach, Honjyan Tang, Jing Ma, Davide Rossi, A. Chadburn, Vundavalli V. Murty, Charles G. Mullighan, Gianluca Gaidano, Raul Rabadan, Paul K. Brindle and Riccardo Dalla-Favera (*Nature, paper in press*).

- Minimal residual disease is an important predictive factor of outcome in children with relapsed 'high-risk' acute lymphoblastic leukemia. Maddalena Paganin, Marco Zecca, Giulia Fabbri, Katia Polato, Andrea Biondi, Carmelo Rizzari, Franco Locatelli, Giuseppe Basso (Leukemia. 2008; 22:2193-2000).

- ASH Congress 2010, Abstract and oral communication. **The Genome of Chronic Lymphocytic Leukemia.** Giulia Fabbri, Vladimir Trifonov, Davide Rossi, Oliver T Elliot, Joseph M Chan, Andrea Rinaldi, Ivo Kwee, Francesco Bertoni, Gianluca Gaidano, Laura Pasqualucci, Raul Rabadan, and Riccardo Dalla-Favera. Blood (ASH Annual Meeting Abstracts).

- ASH Congress 2010, Abstract and oral communication. **Genome-Wide Analysis Reveals Frequent Inactivating Mutations of Acetyltransferase Genes In B-Cell Lymphoma** Laura Pasqualucci, David Dominguez-Sola, Annalisa Chiarenza, Giulia Fabbri, Adina Grunn, Vladimir Trifonov, Lawryn H. Kasper, Stephanie Lerach, Ma Jing, Davide

Rossi, Charles Mullighan, Gianluca Gaidano, Raul Rabadan, Paul K. Brindle, and Riccardo Dalla-Favera. Blood (ASH Annual Meeting Abstracts).

- ASH Congress 2008, Abstract. Clonal Profile Analysis of Leukemic Progenitors and Diagnosis Blast Cells in Pediatric B-Cell Precursor Acute Lymphoblastic Leukemia Marinella Veltroni, Maddalena Paganin, Chiara Frasson, Giulia Fabbri, Antonio Marzollo, Elena Seganfreddo, Emanuela Giarin, Elena Fortunato, Maurizio Aricó, and Giuseppe Basso. Blood (ASH Annual Meeting Abstracts).

- ASH Congress 2008, Abstract and poster. A New Subtype of Excellent Prognosis Identified by GEP in Childhood BCP-ALL Patients. Elena Vendramini, Marco Giordan, Silvia Bresolin, Luca Trentin, Andrea Zangrando, Giulia Fabbri, Geertruy Kronnie, and Giuseppe Basso. Blood (ASH Annual Meeting Abstracts).

- XXXIV National congress AIEOP 2007, Abstract. Maddalena Paganin, Marco Zecca , Giulia Fabbri, Katia Polato, Andrea Biondi, Carmelo Rizzari, Franco Locatelli, Giuseppe Basso. MRD monitoring in S3-S4 ALL relapsed patients.

Oral communications

- ASH Congress 2010, Abstract and oral communication. **The Genome of Chronic Lymphocytic Leukemia.** Giulia Fabbri, Vladimir Trifonov, Davide Rossi, Oliver T Elliot, Joseph M Chan, Andrea Rinaldi, Ivo Kwee, Francesco Bertoni, Gianluca Gaidano, Laura Pasqualucci, Raul Rabadan, and Riccardo Dalla-Favera. Blood (ASH Annual Meeting Abstracts).

- XXXIV National congress AIEOP 2007, Abstract and oral communication. Giulia Fabbri, Maddalena Paganin, Katia Polato, Daniel Nardo, Gianni Cazzaniga, Andrea Biondi, Giuseppe Basso. Significance of MRD monitoring during the follow-up of ALL pediatrics patients.

Ringraziamenti

Un grazie di cuore

Alla mia famiglia, per aver sostenuto le mie scelte, per avermi sempre dato fiducia e per essere sempre presente

Ai miei amici più cari, in particolare Chicca, Martina, Francesca e Maddalena, per essermi sempre vicine, anche a distanza e nei momenti più difficili

A tutti i colleghi che ho incontrato nel Laboratorio del Professor Basso, in particolare Maddalena, Martina, Emanuela, Katia e Elena, con cui ho condiviso molte tappe importanti della mia crescita professionale ed individuale

Al Professor Basso che con entusiasmo e calore mi ha accolto nel suo fantastico gruppo di lavoro e mi ha incoraggiato a perseguire i miei sogni, offrendomi la libertà e gli strumenti per intraprendere il meraviglioso cammino della ricerca

Ai nuovi colleghi che ho incontrato nel Laboratorio del Professor Riccardo Dalla-Favera, in particolare Adina e Christof, per condividere giorno dopo giorno ogni aspetto di questa meravigliosa avventura

Ai Professori Riccardo Dalla-Favera e Laura Pasqualucci per guidarmi con estrema passione e dedizione e con straordinaria competenza in ogni passo della mia formazione scientifica