Transposon mutagenesis reveals fludarabine-resistance mechanisms in chronic lymphocytic leukemia

Tatjana Pandzic1*, Jimmy Larsson1,2*, Liqun He1, Snehangshu Kundu1, Kenneth Ban1,3, Muhammad Akhtar-Ali1, Anders R Hellström1, Anna Schuh4, Ruth Clifford4, Stuart J. Blakemore5, Jonathan C. Strefford6, Tycho S Bauman6, Armando Lopez-Guillermo6, Elias Campo7, Viktor Ljungström1, Larry Mansouri1, Richard Rosenquist1, Tobias Sjöblom1 and Mats Hellström1§

1. Dept. of Immunology, Genetics and Pathology, Science for Life laboratory, Uppsala University, 75185 Uppsala, Sweden.

2. Present address: Dept. of Cell and Molecular Biology, Uppsala University, Sweden.

3. Dept. of Biochemistry, Yong Loo Lin School of Medicine, NUS, Inst. of Molecular and Cell Biology, A*STAR, Singapore.

4. Radcliffe Department of Medicine, Oxford University, UK.

5. Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK.

6. Servicio de Hematología, Hospital Clínico, IDIBAPS, 08036 Barcelona, Spain.

7. Unitat de Hematologia, Hospital Clínic, IDIBAPS, Universitat de Barcelona, 08036 Barcelona, Spain.

* These authors contributed equally

Running title: Forward genetic screen of fludarabine resistance

§ Corresponding author: mats.hellstrom@igp.uu.se
Abstract

**Purpose:** To identify resistance mechanisms for the chemotherapeutic drug fludarabine in chronic lymphocytic leukemia (CLL), as innate and acquired resistance to fludarabine-based chemotherapy represents a major challenge for long-term disease control.

**Experimental design:** We employed *piggyBac* transposon-mediated mutagenesis, combined with next-generation sequencing, to identify genes that confer resistance to fludarabine in a human CLL cell line.

**Results:** In total, this screen identified 782 genes with transposon integrations in fludarabine-resistant pools of cells. One of the identified genes is a known resistance mediator *DCK* (deoxycytidine kinase), which encodes an enzyme that is essential for the phosphorylation of the pro-drug to the active metabolite. *BMP2K*, a gene not previously linked to CLL, was also identified as a modulator of response to fludarabine. In addition, ten out of 782 transposon-targeted genes had previously been implicated in treatment resistance based on somatic mutations seen in patients refractory to fludarabine-based therapy. Functional characterization of these genes supported a significant role for *ARID5B* and *BRAF* in fludarabine sensitivity. Finally, pathway analysis of transposon-targeted genes and RNA-seq profiling of fludarabine-resistant cells suggested deregulated MAPK signaling as involved in mediating drug resistance in CLL.

**Conclusions:** To our knowledge, this is the first forward genetic screen for chemotherapy resistance in CLL. The screen pinpointed novel genes and pathways involved in fludarabine resistance along with previously known resistance mechanisms. Transposon screens can therefore aid interpretation of cancer genome sequencing data in the identification of genes modifying sensitivity to chemotherapy.
Translational relevance

We performed a *piggyBac* transposon mutagenesis screen in a human CLL cell line to identify genes and pathways involved in fludarabine resistance. This screen revealed both known and novel genes mediating drug resistance, in contrast to previous transposon screens that only identified already known genes. Three new candidate genes for drug resistance to fludarabine were identified, i.e. *BMP2K*, *ARID5B* and *BRAF*, and a key role for the mitogen-activated protein kinases (MAPK) pathway was implicated. Hence, the study provides a proof-of-principle that transposon mutagenesis can generate important functional data regarding drug response, thus complementing ongoing large-scale mutational analyses in clinical materials.
Introduction

Chronic lymphocytic leukemia (CLL) is a clinically and biologically heterogeneous disease. While most patients are asymptomatic at diagnosis and only a minor proportion of patients will experience rapid disease progression and early treatment, the majority of patients will develop a more aggressive disease requiring therapy over the coming years (1). To identify patients with an expected inferior outcome, detection of certain recurrent cytogenetic aberrations (i.e. deletion of 11q, 13q and 17p and trisomy 12) and TP53 mutational screening is currently recommended for all patients before initiation of therapy (1-3). In fact, TP53 aberrations, i.e. del(17p) and/or TP53 mutations, are associated with a particularly short time to progression, poor response to therapy and the worst overall survival of all subtypes of CLL (1).

The gold-standard first-line chemoimmunotherapy for medically-fit CLL patients is a combination treatment, i.e. fludarabine, cyclophosphamide and the anti-CD20 antibody rituximab (FCR), with overall response rates of 90-95% (4, 5). However, most patients will eventually relapse and 10% of patients are also primary fludarabine-refractory (FR). Approximately 40% of the primary FR patients have somatic TP53 mutations, further underscoring the key role of p53 in drug response (6, 7). ATM dysfunction, due to del(11q) and ATM mutations, is also associated with reduced response to FC, but, in contrast to TP53 aberrations, patients with del(11q) have an improved response rate with the addition of rituximab (4, 8).

The recent development of next-generation sequencing (NGS) technologies has been successful in identifying not only genes driving CLL ontogeny and evolution, but also in revealing mutations that are predictive for therapy response (reviewed in (9)). So far, mutations in SF3B1, BIRC3, NOTCH1, POT1, FAT1, SAMD1 and IKZF3, have been
identified as enriched in patients relapsing after FCR or with FR disease (10-17). Regardless of this, there is still a significant fraction of CLL patients relapsing after FCR where the genetic cause of resistance remains elusive. Similar to other cancer types, NGS efforts in CLL require large patient cohorts and longitudinally collected samples to pinpoint new genetic alterations as causally involved in drug resistance and formal proof also requires experimental validation. Therefore, alternative approaches complementing the existing sequencing efforts in patients could greatly improve the interpretation of NGS data.

In this study, we performed a piggyBac transposon mutagenesis screen, which is the first forward genetic screen to identify genes modifying drug response in CLL. piggyBac transposon mutagenesis has been applied to identify mechanisms of chemotherapy resistance to purine analogues (6-thioguanine) and taxanes (paclitaxel), while also aiding the identification of novel second site mutations in BRAF in response to vemurafenib (18-20). Our analysis revealed both novel genes, not previously implicated in chemotherapy resistance in CLL, as well as genes already linked to refractory disease thus validating our approach.
Materials and Methods

Cell culture

The human EBV-transformed, lymphoblastoid cell line HG3, established from a poor-prognostic CLL patient with unmutated IGHV genes (further characterized in (21)) was propagated in RPMI 1640 medium supplemented with 0.5 × Glutamax, 10% heat inactivated FBS and 1% PEST (all from Gibco/Life Technologies). The human colorectal cancer cell lines RKO, its isogenic derivative RKO $BRAF^{wt/-}$ (both from Horizon Discovery) and HCT116 (ATCC) were maintained in McCoy’s 5A medium supplemented with 10% FBS and 1% PEST (all from Gibco/Life Technologies). HCT116 $TP53^{-/-}$ and its parental HCT116 $TP53^{+/+}$ cell line (kindly provided by B. Vogelstein, Johns Hopkins University) were cultured in DMEM medium (Gibco/Life Technologies) supplemented with 10% FBS and 1% PEST.

piggyBac transposon screen

Transposon and transposase (HyPBase) constructs are schematically illustrated in Supplementary Figure S1. 15 - 30 × 10^6 HG3 cells were co-electroporated with 3 μg of the transposase and transposon plasmids using the Amaxa Nucleofector kit R (Lonza) with the program W-01. Following a 48-hour recovery, the cells were grown either under selection with 3 μM F-ara-A (Sigma Aldrich), the nucleoside pro-drug of fludarabine, or no selection for 6-12 weeks. During this time the cells were passaged twice per week and each time seeded at a density of 100,000 cells/mL in a t75 culture flask. Following this, F-ara-A resistant cells were treated with 250 μM hygromycin B (Life Technologies) to select only those cells with active transposon insertions. The treatment continued until no viable cells remained in a control HG3 cell culture (not transfected with transposon/transposase).
Cell viability assay

Following hygromycin selection of cells with transposon insertions, cell pools that were either treated or untreated with F-ara-A were tested for F-ara-A sensitivity using a resazurin viability assay. Briefly, 100,000 cells were seeded in 1 ml in triplicate wells in a 24-well plate and treated with increasing concentrations of F-ara-A (ranging from 1 to 64 μM) for 72h. To assess growth suppression, 44 μM resazurin (Sigma Aldrich) was added to the medium and fluorescence was measured in a Wallac Victor Multilabel Counter (Wallac, Turku, Finland) with excitation and emission wavelengths of 530 and 590 nm, respectively. The relative number of viable cells was expressed as percentage of untreated cells.

Splinkerette PCR

Genomic DNA was isolated from cell pools with the NucleoSpin DNA extraction kit (Machery Nagel). Splinkerette PCR was performed as described (22). Briefly, following digestion of genomic DNA with Sau3AI, DNA was ligated to adaptors and a primary PCR reaction was performed with a Splink1- 5’-CGAAGAGTAACCGTTGCTAGGAGAGACC-3’ and HMsp1-5’- CGAAGAGTAACCGTTGCTAGGAGAGACC-3’ primer pair (20). Secondary PCR was carried out with barcoded primers (Supplementary Material and Methods) and the products were treated with ExoSap. The barcoded samples were mixed in equimolar concentrations and sequenced on an Illumina HiSeq 2000 instrument.

Sequence processing and identification of genes with transposon integrations

To map the transposon integrations, the sequences were trimmed to remove the barcode and adaptor sequences, then aligned to human genome (hg19) using the Smalt software (www.sanger.ac.uk/resources/software/smalt, version 0.7.5.1, with default parameters). The mapped sequences were parsed to identify the genomic integration locations, together with the...
number of the sequences that support the integrations. To identify candidate genes with a significant number of integration sites, the observed and expected integrations for each Ensembl gene were computed (23) and Poisson statistics were used to calculate a p-value for each gene. To correct for multiple hypothesis testing, the false discovery rates were calculated for each candidate gene. Detailed description of data analysis is shown in **Supplementary Material and Methods**.

**Transfection, plasmids and shRNA vectors**

The pGIPZ lentiviral shRNA vectors targeting DCK (RHS4430-20031599, RHS4430-200183211, RHS4430-200189164), BMP2K (RHS4430-200166293, RHS4430-200234236, RHS4430-200237243), UBR5 (RHS4430-200269761, RHS4430-200215630) as well as a non-silencing pGIPZ control vector (RHS4346) were purchased from Thermo Scientific. HG3 cells were electroporated with 5 μg of each vector using the Amaxa Nucleofector kit R (Lonza) with the program W-01. Pools of HG3 cells stably expressing indicated shRNAs were selected in 1 μg/mL puromycin for 6–8 weeks.

For gene overexpression analyses, a Myc tagged ARID5B construct was purchased from Origene (RC211454) while the Flag tagged LARS was from Addgene (plasmid 46341, D. Sabatini) (PMID: 23723238). DCK and NUDCD3 ORFeome clones (ORFeome Collaboration (OC) cDNA Clones: [http://www.orfecomercollaboration.org](http://www.orfecomercollaboration.org)) were obtained from the core facility Karolinska High Throughput Center [http://www.scilifelab.se/facilities/khtc](http://www.scilifelab.se/facilities/khtc) and subsequently cloned into a Vivid Colors™ pcDNA™6.2/C-EmGFP-DEST Mammalian Expression Vector (Life Technologies). The empty pcDNA3 vector was used as a control. Transfection of HCT116 cells was performed with Lipofectamine 2000 (Life Technologies) using 5 μg of recombinant construct.
To assess the sensitivity to F-ara-A following transient or stable transfection cells were seeded in either 96-well plates (HCT116 and RKO; 8,000 cells/well) or 24-well plates (HG3; 100,000 cells/well) and treated with increasing concentrations of the drug (ranging from 2 to 128 μM for HCT116 and RKO cells and from 1 to 64 μM for HG3 cells) for 72h. The growth suppression was evaluated using a viability assay as described.

**Western blot**

Cells were lysed in NP40 buffer in the presence of protease inhibitors (Complete Mini; Roche) and phosphatase inhibitors (Halt Phosphatase Inhibitor Cocktail; Thermo Scientific). An aliquot of 50 μg total protein was run on an SDS–PAGE gel and transferred to a nitrocellulose membrane (GE Healthcare). Primary antibodies used were anti-Myc (Cell Signaling; 1:1000), anti-pERK1/2 (Cell Signaling #4377; 1:1000), anti-ERK1/2 (Cell Signaling #4696; 1:1000), anti-α-Tubulin (T8203, Sigma Aldrich; 1:1000), anti-Flag (Sigma Aldrich F3165, 1:1000) and anti-β-actin (Abcam ab6276, 1:2000). Immunoreactive proteins were visualized with SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) on the ImageQuant LAS 4000 imaging system (GE Healthcare).

**Real-time quantitative PCR**

Total RNA was extracted using an RNeasy Plus Mini kit (Qiagen) and cDNA was synthesized using a Maxima H Minus First Strand cDNA Synthesis kit (Thermo Scientific). Real-time quantitative PCR was performed with Maxima SYBR Green /Rox qPCR Master Mix (Thermo Scientific) in a StepOne qPCR instrument (Applied Biosystems) (primers are listed in Supplementary Material and Methods). Data was analyzed by employing the ΔΔCt method with the software provided by Applied Biosystems Inc. using ACTB as the reference gene.
RNA-sequencing

RNA samples were prepared as described. The RNA integrity and concentration was determined on a Bioanalyzer instrument (Agilent) using an RNA 6000 Nano kit. Samples were sequenced on the Ion Proton system (Ion Torrent/Life Technologies). In each sample, approx. 30–50 million reads were sequenced. The reads were aligned to the human genome hg19 assembly ([https://genome.ucsc.edu](https://genome.ucsc.edu)) using Tophat2 software (version 2.0.4) (24). The RNAseq read count quantification was performed using the HTSeq tool (version 0.6.1) (25). To identify the gene expression differences between the three pairs of HG3 pools, DESeq2 software was used. (26)

BRAF screening of CLL patient samples

Using whole-exome/genome sequencing, the mutation status of BRAF was assessed in 506 Spanish patients as recently reported (15). In addition, samples from 247 untreated patients enrolled in the UK LRF CLL4 clinical trial (27) were investigated using targeted NGS including the BRAF gene.

Statistical analysis

Graphpad Prism 6.0b was used for statistical analysis. Dose-response data was analyzed using non-linear regression and log(inhibitor) vs response - variable slope (four parameters) to calculate IC₅₀ values from viability tests. Statistical significance was calculated using students T-test for IC₅₀ values for specific doses of F-ara-A. All data shown are based on at least three independent experiments. For the Spanish cohort, time-to-first-treatment (TTFT) was measured from time of sampling until date of first treatment, and overall survival (OS) from time of sampling until the last follow-up or death. For the UK LRF CLL4 trial patients, OS was calculated from randomization date until the last follow-up or death, while progression-
free survival (PFS) was measured from randomization until event. Survival curves were constructed using the Kaplan-Meier method and the log-rank test was applied to determine differences between survival proportions using the Statistica Software 12.0 (Stat Soft Inc, Tulsa, OK, USA).
Results

A forward genetic screen to identify fludarabine resistance genes in CLL

Three independent pools of the transposon mutagenized human CLL cell line (HG3) were generated and selected in 3 µM of F-ara-A (the nucleoside pro-drug of fludarabine) for 6-12 weeks (Figure 1 and Supplementary Figure S1). Each HG3 pool was estimated to contain 2-4 × 10^5 independent insertions after transfection, which corresponds to the potential activation of 95% of the genome (18). After selection in F-ara-A, each pool exhibited 2–5 fold reduced sensitivity to F-ara-A, as compared to parental HG3 (Figure 1 and Supplementary figure S1B and S1C). No difference in F-ara-A sensitivity was seen in mutagenized HG3 cells that were not cultured in F-ara-A or in non-mutagenized HG3 cells that were cultured in F-ara-A (Figure 1, Supplementary Figure S1 and data not shown). Splinkerette PCR was applied to amplify the flanking regions of the transposon insertions and high-throughput sequencing of the PCR products were performed to identify the location of the transposons (Figure 1).

Validation of genes mediating fludarabine resistance

Collectively, this screen identified 782 genes with transposon integrations not overlapping with genes identified in the non-selected pools (Supplementary Table S1). Among the top 10 most significantly targeted genes by the transposon insertions was deoxycytidine kinase (DCK), which encodes an essential step in the activation of the purine nucleoside analogue fludarabine, where it phosphorylates the nucleoside (F-ara-A) to nucleoside-monophosphate (Table 1). This indicates that the screen indeed can identify genes modifying fludarabine sensitivity. While performing splinkerette PCR a single strong band was observed on the gel in addition to the expected smear. This band was isolated and Sanger sequenced and identified as an insertion within BMP2K, one of the top 10 transposon targeted genes out of the 782 identified (Table 1). We next performed shRNA mediated knock-downs of BMP2K and DCK.
which led to a robust down-regulation of the BMP2K and DCK transcripts (Supplementary Figure S2) and protein (data not shown). Titration of F-Ara-A resulted in a 2.5- and 3.3-fold increase in IC50 when BMP2K and DCK had been knocked-down as compared to HG3 cells expressing control shRNA (Figure 2A).

To identify potentially clinically relevant mechanisms for acquired resistance to F-ara-A, the gene sets identified by piggyBac transposon mutagenesis were intersected with 148 genes found somatically mutated in CLL patients refractory to fludarabine-based therapy (12). Overall, 10 genes were found in common between the transposon and the genomic screen, which is a significant enrichment compared to the expected enrichment occurring by chance (hypergeometric probability P=0.034) (Table 2). ARID5B, BRAF, LARS, NUDDCD3, and UBR5 were selected for functional validation, primarily based on their gene expression in CLL and in the HG3 cell line. The choice of validation using cDNA over-expression or shRNA mediated knock-down was based on transposon insertion, where piggyBac insertion upstream or in first or second intron in a sense orientation suggested an activation event that was validated by gene overexpression. In contrast, piggyBac insertion within the gene and in an antisense orientation indicated gene silencing.

shRNA mediated knock-down of UBR5 led to robust down-regulation of the mRNA, but did not result in any alteration in the IC50-value for F-ara-A (data not shown). Due to the poor transfection efficiency of HG3 cells, we decided to use adherent cell lines to allow for over-expression of plasmids as well as the use of available isogenic cell lines. To test the validity of the colorectal cancer cell line HCT116, we titrated F-ara-A in parental and TP53 knock-out cells. In agreement with the known role of p53 in response to fludarabine-based treatment, HCT116 TP53 knock-out cells showed a 3-fold reduced sensitivity for F-ara-A, suggesting that HCT116 cells recapitulate some key features of fludarabine response and resistance (Supplementary Figure S3). Transfection of GFP-tagged expression vectors in HCT116
cells resulted in approximately 20-30% of the cells being GFP-positive (Supplementary Figure S4). As predicted, transient transfection of DCK resulted in an increased sensitivity to F-ara-A, with IC₅₀ of 30 µM for DCK expressing cells versus 35 µM for control cells (Figure 2B). Given the low frequency of cells expressing the plasmids it is still a significant and sizable change. Over-expression of ARID5B resulted in a modest but significant increase in F-ara-A IC₅₀, 41 µM for ARID5B expressing cells versus 35 µM for control cells (t-test P=0.0064), while no difference was observed between LARS- or NUDCD3-expressing and control HCT116 cells (Figure 2C and data not shown).

The case of BRAF

A transposon insertion was identified in the first intron of the serine/threonine kinase BRAF, which is compatible with gene activation (Figure 2E). Noteworthy, an activating L597Q BRAF mutation was recently identified in one patient reported to be refractory to fludarabine-based therapy (12). To test if BRAF activation confers a reduced sensitivity to fludarabine, we used the colorectal RKO cell line that harbors the activating V600E mutation in BRAF and its isogenic derivative where the mutant alleles have been knocked-out. The presence of V600E was associated with significantly reduced sensitivity to higher doses of F-ara-A, as revealed by cell viability assay (Figure 2D). The difference in response to F-ara-A between BRAF mutant and wt cells was normalized by the addition of the BRAF V600E inhibitor vemurafinib, (Supplementary Figure S5).

To investigate the significance of BRAF mutations in CLL patients, we analysed whole-exome/genome data available from a ‘general practice’ cohort. Ten of 506 patients carried non-synonymous BRAF mutations (median variant allele frequency, 36%, range, 17-58%). Notably, patients with BRAF mutation had a significantly shorter TTFT compared to wildtype (median 0.5 vs 64.5 months, p<0.001), while no significant difference was seen for OS (median 128.0 vs 122.9 months, p=0.98), Figure 3 and Supplementary Table S2. Nine out
of 10 BRAF-mutant patients required treatment within 2 years. We also investigated 247 cases from the UK CLL4 trial and detected 8 patients with BRAF mutations (median variant allele frequency, 32%, range, 12-50%, Supplementary Table S2), however, no significant difference in OS or PFS was observed (OS 79.0 vs 75.0 months, p=0.97, PFS, 39.0 vs 24.0 months, p=0.53).

The MAPK signaling pathway and fludarabine resistance in CLL

Pathway analysis of the 782 genes identified in the transposon screen using gene-set enrichment analysis (GSEA) (28, 29) showed an enrichment of genes in cancer-related pathways and mitogen-activated protein kinase (MAPK) signaling (Table 3). In fact, one third of the genes in the cancer-related pathways belonged to the MAPK-pathway (Supplementary Table S3).

To address if the fludarabine-resistant HG3 cells were transcriptionally different at a global level compared to their parental cells, we performed RNA-sequencing of three pairs of HG3 pools. This analysis revealed 302 genes that were significantly differentially expressed (more than 2-fold) and hierarchical clustering demonstrated distinct and stable changes in the RNA expression patterns in the fludarabine resistant HG3 pools versus the parental cell lines (Supplementary Figure S6 and Supplementary Table S4). Comparing the mRNA expression profiles of the HG3 resistant pools to reported profiles using GSEA (“oncogenic signatures”) revealed significant similarities to 35 other mRNA expression signatures (Supplementary Table S5). Nine were related to MAPK-signaling, due to activation of EGFR, KRAS, CRAF, MEK1 or inactivation of STK33.

Taken together, the experimental validation of BRAF, the bioinformatics analysis of transposon insertions and the transcriptional deregulation of the MAPK-pathway indicate a role for the MAPK pathway in fludarabine resistance. In line with this, phosphorylated ERK
(extracellular signal-regulated kinase) was also increased in a fludarabine refractory HG3 pool 
(Supplementary Figure S7).
Discussion

Interpretation of data from large-scale cancer genome sequencing studies is often hampered by the high frequency of passenger mutations, sequencing artifacts and small cohort sizes. Complementary approaches are therefore needed to assign relevant functions to genes identified as mutated in cancer in order to understand the genetics and biology driving cancer development and to provide insights into mechanisms of resistance to therapy. piggyBac transposon mutagenesis is a powerful forward genetics tool that can facilitate the identification of genes mediating resistance to chemotherapy. This technique has been successfully applied in transgenic mice as well as in cell culture systems for the discovery of genes driving cancer development and drug resistance (18-20, 30, 31).

To determine genetic mediators of resistance to fludarabine, we performed a piggyBac transposon screen in the HG3 CLL cell line and identified 782 transposon-targeted genes in pools of cells exhibiting fludarabine resistance. The identification of DCK, a known determinant of cellular response to fludarabine, underscores that the screen has the power to detect biologically relevant genes. DCK encodes a rate-limiting kinase in the activation of the pro-drug F-ara-A to the nucleoside-monophosphate. Low levels of DCK have been shown to correlate with impaired response to fludarabine in CLL (32). In addition, reduced levels of DCK were also observed in mantle cell lymphoma patients exhibiting refractoriness to cytarabine, a member of the same class of drugs as fludarabine (33).

The piggyBac screen also identified BMP2K as a novel gene conferring resistance to fludarabine. BMP2K was first identified as a BMP2 inducible kinase that plays a role in osteoblast differentiation (34) and, together with BMP2, it was also shown to regulate eye development (35). The role of BMP signaling in B-cell malignancies is less clear, but it was reported that BMP6 inhibits the growth of mature B cells and that epigenetic inactivation of
BMP6 was associated with more aggressive disease in diffuse large cell B-cell lymphomas (36, 37). Since BMP receptors are found upregulated in CLL cells, but not in normal B cells, it is plausible that BMP2K might act as an essential down-stream target of BMP signaling in CLL (38). The fact that BMP2K was the only transposon-targeted gene in our screen that was detected by traditional Sanger sequencing following splinkerette PCR, indicates a possible enrichment of cell clones with this integration within the pool of fludarabine resistant cells. Until now, BMP2K has not been linked to either CLL pathophysiology or the cellular response to fludarabine and elucidation of its role in these processes hence warrants further investigation.

To identify genes that could mediate fludarabine resistance of potential direct relevance to CLL patients, we compared the gene set identified in pools of piggyBac mutagenized HG3 cells after selection in fludarabine with genes identified in a set of patients with FC-refractory CLL (12). The combined approach efficiently narrowed down the number of candidate genes from hundreds to ten and their contribution to fludarabine resistance was functionally assessed in vitro using gene overexpression or gene knock-down approach. ARID5B was one of the genes identified and validated in the intersection between the piggyBac screen and genes found mutated in a FC-refractory patient. ARID5B is a member of the AT-rich interaction domain containing protein (ARID) family. This family consists of DNA-binding proteins that promote transcription via association with the SWI/SNF chromatin-remodeling complex (39). Interestingly, genome-wide association studies have revealed an association of a variant allele of ARID5B with an increased risk for childhood B-cell ALL and treatment outcome in ALL (40, 41). ARID5B is mutated in 6–23% of endometrial carcinoma, and Arid5b knockout mice exhibit defects in normal B cell development (39, 42). In addition, the homologue ARID1A is a driver mutation in 1.6% of CLL patients and was also found mutated in splenic marginal zone lymphoma and germinal-center derived lymphomas (15, 43, 44).
Thus, ARID5B is a plausible novel candidate for affecting CLL progression and response to treatment.

Another gene that was common to both the piggyBac screen and the analysis of patients with FC-refractory CLL was BRAF, where our validation experiments using cancer cell lines with an activating V600E mutation in BRAF confirmed that its activation may lead to fludarabine resistance. BRAF mutations have been recurrently reported in CLL patients with a frequency of approximately 3% (12, 14, 15, 45, 46). While the most prevalent BRAF mutation in cancer in general is V600E, which leads to a constitutive activation of BRAF, mutations other than V600E are more frequently seen in CLL without any obvious hot-spot (Supplementary Figure S8 and Table S6). However, most of them cluster within or near the activation loop where V600E resides and are known to confer variable, but increased signaling and oncogenic capacity (47). Given that most of the mutations may lead to increased BRAF activity, we speculate that the BRAF mutations observed in the activation loop in CLL patients act in a similar way with regards to fludarabine sensitivity, even though we cannot rule out that mutations other than V600E might behave differently. Of note, activating mutations in other members of the MAPK pathway have also been reported in CLL, e.g. NRAS, KRAS, MAPK1 (ERK2) and ERG2 (14, 15, 45). Hence, if one considers the entire MAPK pathway it is possible that its contribution to CLL pathobiology is larger than previously thought (45, 46).

Transcriptional profiling of the fludarabine-refractory HG3 pools revealed distinct changes in mRNA expression that were similar to established “oncogenic” signatures observed after activation of EGFR, KRAS, CRAF and MEK1. Three genes that were common to both fludarabine-resistant HG3 cells and activated EGFR/KRAS/CRAF/MEK1 pathways had previously been linked to B-cell signaling and function, e.g. IGFBP4, PTPRE, and VAV3. IGFBP4 is an insulin growth factor (IGF) binding protein that modulates cell growth via IGF.
It is upregulated in CLL and acts downstream of oncogenic KRAS and EGR1 (48, 49). PTPRE is a phosphatase expressed in leukocytes that is induced by antigen stimulation and regulates MAPK signaling (50). VAV proteins are critical for T- and B-cell receptor-induced Ca$^{2+}$ signaling and NF-κB-mediated survival signaling in B cells (51, 52). In line with our findings from the piggyBac screen, other studies have also suggested that RAF/MEK/ERK signaling has pro-survival effects under fludarabine treatment (53-55).

CLL cells are dependent on stromal cell interaction in order to prevent apoptosis and this protective effect is partially due to CXCL12-CXCR4 interaction that leads to induction of pERK and STAT3 and protects the CLL cells from fludarabine-induced apoptosis (53). In addition, inhibition of BRAF or MEK1 was shown to abrogate the protective effect of stromal cells (54), while cladribine, another purine nucleoside analogue, induces ERK phosphorylation and together with inhibition of MEK protein kinase acting upstream of ERK led to synergistic cell killing (55).

In our screening of CLL patient samples, the mutation status of BRAF was associated with shorter TTFT in the ‘general practise’ cohort, while no difference was observed for OS. In line with the latter finding, investigation of patients requiring treatment from the UK CLL4 trial did not show any difference in PFS and OS. Hence, further studies on larger number of cases are imperative to determine the potential prognostic/predictive role of BRAF mutations in CLL.

Given the known importance of TP53, ATM and SF3B1 in clinical response to fludarabine-based therapy, it might at first sight appear surprising that these genes were not identified in this screen. However, considering the biological heterogeneity of the disease, one CLL cell line (HG3), though established from a CLL patient with aggressive disease, may not capture
all types of resistance patterns, and further investigation of additional CLL cell lines will be necessary to get different model systems representative of major CLL subtypes.

In conclusion, to our knowledge, this is the first forward genetic screen for genes mediating fludarabine resistance in an experimental system. Our data supports the notion that the combination of genetic screens in experimental systems and NGS-based mutational analyses of human samples complement each other in an efficient way to identify functionally relevant genes not only for CLL but also for other cancer types. In particular, we identified both known and novel mediators of fludarabine resistance, hence underscoring the broad applicability of this method and highlighting activation of BRAF/MAPK signaling and downstream targets as a central mechanism behind reduced fludarabine sensitivity in CLL.

Acknowledgement:

This research project was supported by the Swedish Cancer Society, the Swedish Research Council, The Beijer Foundation, the Lion’s Cancer Research Foundation, and Selander’s Foundation, Uppsala. Sequencing was performed by the SNP&SEQ Technology Platform, Science for Life Laboratory at Uppsala University, a national infrastructure supported by the Swedish Research Council (VRRFI) and the Knut and Alice Wallenberg Foundation. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2014071. JCS is supported by the Bloodwise (11052, 12036, 14027), the Kay Kendall Leukaemia Fund (873) and the Bournemouth Leukaemia Fund, with infrastructure support from a Cancer Research-UK centre grant (C34999/A18087).
References


Figure Legends

Figure 1. A piggyBac transposon-based screen generates pools of HG3 cells with reduced sensitivity to fludarabine. A, HG3 cells (white and grey circles) were electroporated (lightning) with piggyBac transposon and transposase (Supplementary Figure S1) and subsequently selected in the presence of 3 μM (+F-ara-A) or absence of fludarabine (- F-ara-A). Cells were passaged twice weekly and cultured for 6 – 12 weeks and thereafter selected in Hygromycin B to allow survival of only the transposon-containing HG3 cells (black circles). This removed cells that had potentially become resistant to F-ara-A by means other than transposon insertion. Subsequent cell viability tests with increasing concentrations of F-ara-A revealed resistant pools of HG3 cells. Cell pools were harvested for DNA and RNA extraction and processed for splinkerette PCR to identify insertion sites for piggyBac transposon. B, Titration of F-ara-A showing a pool of piggyBac mutagenized HG3 cells with 3-fold reduced sensitivity to fludarabine. Parental HG3 (black line, IC₅₀: 6.4μM, a pool of HG3 cells that had been mutagenized with piggyBac and selected in F-ara-A (piggybac + F-ara-A) IC₅₀: 20.9μM and a pool of HG3 cells that had been mutagenized with piggyBac but not selected in F-ara-A (piggybac - F-ara-A) IC₅₀: 6.2μM). The individual doses with statistically significant differences are indicated with stars (* P<0.05, ** P<0.01), the error bars indicate standard error of mean (SEM).

Figure 2. The putative fludarabine resistance genes DCK, BMP2K, ARID5B and BRAF confer acquired resistance to fludarabine. A, Summary of F-ara-A dose-response curves of HG3 cells expressing control shRNA (black line and boxes) or shRNA targeting DCK (grey line and triangles) and BMP2K (grey dashed line and boxes). Knock-down of either gene confers reduced sensitivity to F-ara-A as compared to control cells. B, Summary of dose-
response curves from HCT116 cells of overexpression of \textit{DCK} (grey line and triangles) resulted in an increase in sensitivity to F-ara-A as compared to control vector (black line and boxes). \textbf{C}, Overexpression of \textit{ARID5B} (grey line and triangles) in HCT116 cells led to reduced sensitivity to F-ara-A as compared to control vector (black line and boxes). \textbf{D}, Summary of dose-response experiments of F-ara-A in human RKO colorectal cancer cells (\textit{BRAF} V600E, black line and boxes) and an isogenic derivative where the mutant \textit{BRAF} allele had been knocked out (\textit{BRAF} V600E KO, grey line and triangles). \textit{BRAF} V600E confers reduced sensitivity to F-ara-A. The individual doses with statistically significant differences are indicated with black stars (*=P<0.05, **=P<0.01, ***=P<0.001), the error bars indicate standard error of mean (SEM) \textbf{E}, Schematic representation of the transposon integration within the intron 1 of the \textit{BRAF} gene. Transposon element is shown as a black arrow where the direction of the arrow indicates orientation of integration.

**Figure 3. Time to first treatment in CLL patients with and without \textit{BRAF} mutation.** \textit{BRAF}-mutant patients exhibit significantly shorter time to first treatment compared to \textit{BRAF} wildtype patients (log-rank test, p<0.001). Median time from diagnosis to sampling in BRAF-mutated cases was 0.65 years (0 - 3.46 years).
Table 1. The top ranking genes mediating resistance to fludarabine identified in the piggyBac screen.

<table>
<thead>
<tr>
<th>Gene (HGNC symbol)</th>
<th>Description</th>
<th>TTAA sites</th>
<th>Expected insertions</th>
<th>Observed insertions</th>
<th>Cell pool</th>
<th>P value</th>
<th>P value FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNX5</td>
<td>sorting nexin 5</td>
<td>235</td>
<td>0.02</td>
<td>8</td>
<td>17, 28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MGME1</td>
<td>mitochondrial genome maintenance exonuclease 1</td>
<td>172</td>
<td>0.02</td>
<td>8</td>
<td>17, 28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EIF2AK3</td>
<td>eukaryotic translation initiation factor 2-alpha kinase 3</td>
<td>634</td>
<td>0.04</td>
<td>6</td>
<td>17, 28</td>
<td>6.86E-14</td>
<td>1.57E-11</td>
</tr>
<tr>
<td>DCK</td>
<td>deoxycytidine kinase</td>
<td>348</td>
<td>0.03</td>
<td>4</td>
<td>17</td>
<td>4.24E-10</td>
<td>3.88E-08</td>
</tr>
<tr>
<td>SLC1A3</td>
<td>solute carrier family 1 (glial high affinity glutamate transporter), member 3</td>
<td>609</td>
<td>0.04</td>
<td>4</td>
<td>17</td>
<td>1.43E-09</td>
<td>1.19E-07</td>
</tr>
<tr>
<td>RBM34</td>
<td>RNA binding motif protein 34</td>
<td>184</td>
<td>0.02</td>
<td>3</td>
<td>17</td>
<td>6.68E-09</td>
<td>5.11E-07</td>
</tr>
<tr>
<td>TRIM31</td>
<td>tripartite motif containing 31</td>
<td>76</td>
<td>0.01</td>
<td>2</td>
<td>28</td>
<td>4.00E-08</td>
<td>2.62E-06</td>
</tr>
<tr>
<td>NUDT8</td>
<td>nudix (nucleoside diphosphate linked moiety X)-type motif 8</td>
<td>5</td>
<td>0.00</td>
<td>1</td>
<td>28</td>
<td>1.00E-07</td>
<td>6.11E-06</td>
</tr>
<tr>
<td>BMP2K</td>
<td>BMP2 inducible kinase</td>
<td>1415</td>
<td>0.14</td>
<td>4</td>
<td>28</td>
<td>4.31E-07</td>
<td>1.96E-05</td>
</tr>
<tr>
<td>MOB1A</td>
<td>MOB kinase activator 1A</td>
<td>201</td>
<td>0.014</td>
<td>2</td>
<td>28</td>
<td>4.71E-07</td>
<td>1.96E-05</td>
</tr>
</tbody>
</table>

The ranking was based on frequency of observed transposon integrations in relation to number of total potential piggyBac insertion sites (4 nt recognition site, TTAA) within the genes as well as 5 kb up-stream of the genes. Only protein encoding genes expressed in HG3 cells, according to the RNA-sequencing data (normalized FPKM value greater than 0.1), were considered as plausible candidates. The P values were calculated using Poisson statistics, with False Discovery Rate (FDR) correction for multiple tests and a corrected P value <0.05 as cutoff for selection of significant genes.

Abbreviations: HGNC, HUGO Gene Nomenclature Committee;
Table 2. Genes overlapping between piggyBac mutagenesis and genes found somatically mutated in CLL patients refractory to fludarabine-based therapy.

<table>
<thead>
<tr>
<th>Gene (HGNC symbol)</th>
<th>Mutation (Messina et al.)</th>
<th>Location</th>
<th>Strand</th>
<th>Expression in CLL ( ^a )</th>
<th>Expression in HG3 ( ^b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACSS3</td>
<td>L305V</td>
<td>upstream</td>
<td>-</td>
<td>low</td>
<td>present</td>
</tr>
<tr>
<td>ANO2</td>
<td>F614fs</td>
<td>intron 7</td>
<td>+</td>
<td>low</td>
<td>absent</td>
</tr>
<tr>
<td>ARID5B</td>
<td>N297T</td>
<td>upstream, intron 1</td>
<td>+</td>
<td>high</td>
<td>present</td>
</tr>
<tr>
<td>BRAF</td>
<td>L597Q</td>
<td>intron 1</td>
<td>+</td>
<td>moderate</td>
<td>present</td>
</tr>
<tr>
<td>CCDC80</td>
<td>D872E</td>
<td>intron 2</td>
<td>-</td>
<td>low</td>
<td>absent</td>
</tr>
<tr>
<td>LARS</td>
<td>I852V</td>
<td>intron 1</td>
<td>-</td>
<td>moderate to high</td>
<td>present</td>
</tr>
<tr>
<td>NUDCD3</td>
<td>H144Y</td>
<td>intron 2</td>
<td>+</td>
<td>moderate</td>
<td>present</td>
</tr>
<tr>
<td>PDE7B</td>
<td>E350K</td>
<td>intron 2</td>
<td>-</td>
<td>low</td>
<td>absent</td>
</tr>
<tr>
<td>TMEM132B</td>
<td>A911S</td>
<td>intron 1, intron 4</td>
<td>+</td>
<td>low</td>
<td>present</td>
</tr>
<tr>
<td>UBR5</td>
<td>A2154V</td>
<td>intron 1</td>
<td>-</td>
<td>high</td>
<td>present</td>
</tr>
</tbody>
</table>

a) Based on Genesapiens
low: < 300
moderate: 300 - 1000
high: > 1000

b) Based on RNA sequencing

The mutation identified by Messina et al. is presented as an amino acid location in the protein. Location refers to the transposon insertion in relation to the targeted gene. Strand, + indicates the same direction of the piggyBac cassette and gene transcription; - indicates the opposite direction of the piggyBac cassette relative to gene transcription. Genes shown in bold were selected for functional validation.

* According to the GeneSapiens database: expression value below 300 = low, expression value between 300 and 1000 = moderate, expression value above 1000 = high.

** Based on RNA sequencing data in HG3 cells. Transcripts with a normalized FPKM (fragment per kilobase of transcript per million fragments mapped) value less than 0.1 were considered “absent”.

Abbreviations: HGNC, HUGO Gene Nomenclature Committee.
Table 3. Top 10 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways identified by gene set enrichment analysis (GSEA).

<table>
<thead>
<tr>
<th>Gene Set Name</th>
<th># Genes in Gene Set (K)</th>
<th>Description</th>
<th># Genes in Overlap (k)</th>
<th>% of MAPK genes</th>
<th>FDR q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG PATHWAYS IN CANCER</td>
<td>328</td>
<td>Pathways in cancer</td>
<td>25</td>
<td>28 (7/25)</td>
<td>4.63E-9</td>
</tr>
<tr>
<td>KEGG NON SMALL CELL LUNG CANCER</td>
<td>54</td>
<td>Non-small cell lung cancer</td>
<td>9</td>
<td>44 (4/9)</td>
<td>7.68E-6</td>
</tr>
<tr>
<td>KEGG SMALL CELL LUNG CANCER</td>
<td>84</td>
<td>Small cell lung cancer</td>
<td>10</td>
<td>20 (2/10)</td>
<td>2.6E-5</td>
</tr>
<tr>
<td>KEGG ENDOCYTOSIS</td>
<td>183</td>
<td>Endocytosis</td>
<td>14</td>
<td>14 (2/4)</td>
<td>2.66E-5</td>
</tr>
<tr>
<td>KEGG UBIQUITIN MEDIATED PROTEOLYSIS</td>
<td>138</td>
<td>Ubiquitin mediated proteolysis</td>
<td>12</td>
<td>8 (1/12)</td>
<td>3.51E-5</td>
</tr>
<tr>
<td>KEGG MAPK SIGNALING PATHWAY</td>
<td>267</td>
<td>MAPK signaling pathway</td>
<td>16</td>
<td>100 (16/16)</td>
<td>7.34E-5</td>
</tr>
<tr>
<td>KEGG GLIOMA</td>
<td>65</td>
<td>Glioma</td>
<td>8</td>
<td>38 (3/8)</td>
<td>1.25E-4</td>
</tr>
<tr>
<td>KEGG MELANOMA</td>
<td>71</td>
<td>Melanoma</td>
<td>8</td>
<td>38 (3/8)</td>
<td>2.14E-4</td>
</tr>
<tr>
<td>KEGG ADHERENS JUNCTION</td>
<td>75</td>
<td>Adherens junction</td>
<td>8</td>
<td>0</td>
<td>2.86E-4</td>
</tr>
<tr>
<td>KEGG SNARE INTERACTIONS IN VESICULAR TRANSPORT</td>
<td>38</td>
<td>SNARE interactions in vesicular transport</td>
<td>6</td>
<td>0</td>
<td>3.21E-4</td>
</tr>
</tbody>
</table>

K, number of genes in the pathway; k, number of genes overlapping with the KEGG pathway.

The fraction of genes overlapping with MAPK pathway was based on the genes identified and is presented in (Supplementary Table S2). Statistical significance is given as False Discovery Rate (FDR) q-value with q<0.05 used as cutoff for selection of significant pathways.
Figure 1. Pandzic et al.

A

Transposon mutagenesis

HG3 cells

+ F-ara-A

- F-ara-A

Hygromycin B

Viability assay

DNA extraction and splinkerette PCR

Illumina sequencing of the splinkerette PCR libraries

94,362,550 reads from CLL libraries

58,419,251 reads mapped to Hg19

3,832 unique integration sites

844 significant genes (FDR<0.05) in F-ara-A treated group

744 significant genes (FDR<0.05) in F-ara-A untreated group

782 F-ara-A specific genes

B

Relative survival (%)

-6.0

-5.5

-5.0

-4.5

-4.0

-3.5

-3.0

-2.5

-2.0

-1.5

-1.0

0

F-ara-A, M

Downloaded from clinicanoncologies.cancerres.aacrjournals.org on 2017-09-23.
Figure 2. Pandzic et al.
Figure 3. Pandzic et al.

The figure shows a Kaplan-Meier survival curve comparing the percentage of untreated patients over time from sampling. Two groups are compared:

- WT, n=485
- BRAF mut, n=10

The curve indicates a significant difference between the two groups, with p<0.001.
Clinical Cancer Research

Transposon mutagenesis reveals fludarabine-resistance mechanisms in chronic lymphocytic leukemia

Tatjana Pandzic, Jimmy Larsson, Liquin He, et al.

*Clin Cancer Res* Published OnlineFirst March 8, 2016.

Updated version
Access the most recent version of this article at:

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2016/03/08/1078-0432.CCR-15-2903.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.