Supporting Information

Uckun et al. 10.1073/pnas.1209828109

SI Text

Cells. Surplus leukemia cells from four pediatric patients with newly diagnosed B-lineage acute lymphoblastic leukemia (ALL) were used in the described experiments. The secondary use of leukemic cells for subsequent molecular studies did not meet the definition of human subject research per 45 CFR 46.102 (d and f) because it does not include identifiable private information, as confirmed by the Committee on Clinical Investigations (CCI) Institutional Review Board at Children's Hospital Los Angeles (Human Subject Assurance Number: FWA0001914). The Institutional Review Board approved project numbers were CCI-09–00304 (CCI review date: 12/21/2009; approval date: 12/29/09) and CCI-10–00141 (CCI review date: 7/27/2010; approval date: 7/27/2010). We also used the following human cell lines: the human embryonic kidney cell line 293T (ATCC; CRL-11268), human glioblastoma cell line U373 (ATCC; HTB-17), ALL-1 [Ph^+ ALL, B-lineage, spleen tyrosine kinase (SYK^+)], LOUCY (T-lineage ALL; ATCC CRL-2629), N6-1 (NALM-6 subclone) (pre-B ALL; DSMZ No. ACC-128), and DAUDI (Burkitt's leukemia/lymphoma; ATCC, CCL-213) (1–3). In addition, the wild-type DT40 chicken B-lymphoma cell line, its SYK-deficient subclone established by homologous recombination knockout of the *syk* gene, and SYK-deficient DT40 cells reconstituted with wild-type *syk* were used as components of a well-established genetic model for B-lineage lymphoid cells (2, 3).

Reagents. The rabbit polyclonal C-20 (sc-929) and mouse monoclonal 4D10 (sc-1240) antibodies against human SYK were obtained from Santa Cruz Biotechnology. The rabbit polyclonal antibody for Ikaros (IK) (H-100, sc-13039) was purchased from Santa-Cruz Biotechnology for Western blot analysis and fluorescence staining of IK using previously reported procedures (4). The mouse monoclonal anti-IK antibody was prepared in our laboratory. Goat anti-mouse IgG:Horseradish peroxidase (HRPO) (M15345) and goat anti-rabbit IgG: HRPO (R14745) antibodies were purchased from Transduction Labs. Green-fluorescent Alexa Fluor 488 dye-labeled secondary antibody Alexa Fluor 488 goat anti-mouse IgG (A-11001) and red-fluorescent Alexa Fluor 568 dye-labeled secondary antibody Alexa Fluor 568 $F(ab')_2$ fragment of goat anti-rabbit IgG (A-21069) for confocal microscopy were purchased from Invitrogen. UltraCruz Mounting Medium containing 1.5 μg/mL of DAPI was purchased from Santa-Cruz Biotechnology (sc-24941). The mouse monoclonal anti-HA antibody (HA-probe F-7, sc-7392) recognizing an internal region of the influenza HA protein was purchased from Santa-Cruz Biotechnology. TOTO-3 iodide was obtained from Molecular Probes. Mouse antiphosphoserine (α-PS) antibody (PSR-45) (ab6639) was obtained from Abcam. Molecular-weight markers were purchased from Amersham Pharmacia Biotech. All chemicals used were reagent grade or higher. Protein A-Sepharose was purchased from Repligen. Restriction enzymes and proteinase inhibitors were purchased from Roche. The siRNA pool used for knock-down of the *syk* gene (Gene ID: 6850) (Catalog No. L-003176) included 5′-GCUGCGCAAUUACUACUAU-3′, 5′-GAGCAAAUUGUCCUGAUAG-3′, 5′-CGGAAUGCAUCAACUACUA-3′, and 5′-AGAAAUGUGUUGCUAGUUA-3′. The siRNA pool used for knock-down of the *btk* gene (Gene ID: 695) (Catalog No. L-003107) included 5′-UGAGCAAUAUUCUAGAUGU-3′, 5′-CAACUCUGCAGGACUCAUA-3′, 5′-GGUGAUACGUCAUUAUGUU-3′, and 5′-GCGGAAGGGUGAUGAAUAU-3′. The siRNA pool used for knock-down of the *Ikaros* gene (Gene ID: 10320) (Catalog No. L-019092) included 5′-GCUCAUGGUUCACAAAAGA-3′, 5′-CAAGUAACGUCGCCAAACG-3′, 5′-GCGCAGCGGUCUCAUCUAC-3′, and 5′-GGACGCACUCCGUUGGUAA-3′. Controls included the ON-TARGET*plus* Non-Targeting siRNA scrambled control pool (Catalog No. D-001810). All siRNA were purchased from Thermo Scientific Dharmacon). The SYK inhibitor Piceattanol (PCT) was obtained from Sigma-Aldrich. The phycoerythrin (PE)-labeled anti-CD19 monoclonal antibody (Cat. # 555413) and FITC-labeled anti-CD20 monoclonal antibody (Cat. # 556632) and their controls (MsIgG) were obtained from BD Biosciences.

Recombinant IK1 Protein. The cDNAs encoding murine IK1 was cloned into the *Eschereichia coli* expression vector pMAL-C2 with the isopropyl-1-thio-β-galactopyranoside–inducible Ptac promoter to create an in frame fusion between IK1 coding sequence and the 3′end of the *E. coli* maIE gene, which codes for maltose-binding protein (MBP). *E. coli* strain DH5a was transformed with the generated recombinant plasmid and single transformants were expanded in 5 mL of LB medium (10.0 g/L tryptone, 10.0 g/L NaCl, 5.0 g/L yeast extract) containing ampicillin (1,000 μg/mL) by overnight culture at 37 °C. Expression of the MBP-tagged IK1 was induced with 10 mM isopropyl-1-thio-β-galactopyranoside. The cells were harvested by centrifugation at 4,500 × *g* in a Sorvall RC5B centrifuge for 10 min at 4 °C, lysed in sucrose-lysozyme buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 10% sucrose, 1 mM EDTA, 20 mM lysozyme), and further disrupted by sonication. After removal of the cell pellets by centrifugation at 35,000 × *g* for 1 h at 4 °C, MBP-tagged IK1 protein was purified from the respective culture supernatant by amylose affinity chromatography, as previously described (5–7).

Recombinant SYK Protein. Recombinant murine SYK was produced in a baculovirus expression system and purified as previously described in detail (2, 5).

Standard Biochemical Assays. Immunoprecipitations, kinase assays, phospho-amino acid analyses, and Western blot analysis using the enhanced chemiluminescence detection system (Amersham Pharmacia Biotech) were performed, as described previously (1–9). In phospho-amino acid analyses, the positions of ninhydrin-stained phosphoamino acid standards [phosphoserine (S), phosphothreonine (T), and phosphotyrosine (Y)] are indicated with circles.

Mass Spectrometry. After the in vitro kinase reaction with recombinant murine SYK, phosphorylated recombinant murine IK1 samples were digested in solution overnight at 37 °C with trypsin. Peptides were concentrated and desalted using the Millipore C18 reverse-phase Zip-Tips column (ZTC 18S096; Millipore). Peptide mixtures were separated and analyzed in the University of Southern California (USC) Proteomics Core by using the NanoLC system from Eksigent, a nano–LC-MS/MS proteomics system that employs Eksigent's Microfluidic Flow Control technology, which creates rapid, reproducible, low flow-rate binary gradients using continuous flow measurement in combination with a controlled pressure source (10). Reverse-phase chromatography was performed using an in-house packed C18 column (OD 360 μm, ID 75 μm, 10 cm in length, packed with 5-μm reverse-phase C18 packing material from Varian. Columns were washed with 95% mobile phase A [0.1% formic acid in H_2O (vol/vol) and 5% mobile phase B (ACN with 0.1% formic acid) (vol/vol)]. Separation of peptides was carried out using a short gradient elution from 5% mobile phase B to 60% B in 15 min followed by a gradient from 60% B to 70% B in 35 min. Eluted peptides were analyzed with an LTQ-

Orbitrap XL hybrid FTMS (Fourier Transform Mass Spectrometer) (Thermo-Fischer Scientific) in positive ionization mode (10). The 10 most intense ions were selected for collision-induced dissociation MS/MS scans in parallel in the linear ion trap with a normalized collision energy setting of 35% and an activation time of 30 ms. Proteome Discoverer software (Thermo Scientific) with the SEQUEST database search algorithm was used for protein identification. MS/MS spectra were searched against the National Center for Biotechnology Information (NCBI) Database and a custom-made mouse IK database. Search parameters included two missed cleavages from trypsin proteolysis. Modifications selected were set as fixed carbamidomethylation of cysteines, variable deamidation of asparagine and glutamine, variable oxidation of methionine, and phosphorylation of serine, threonine, and tyrosine residues. All peptides with high-quality tandem mass spectra were manually interpreted and confirmed by the USC Proteomics Core. The sequences and phosphorylation sites of all identified phosphopeptides were carefully assigned accordingly. A MS analysis was performed on trypsin-digested recombinant IK after an in vitro kinase reaction with purified recombinant SYK. We identified Ser^{358} (S358) and Ser^{361} (S361) as two unique SYK phosphorylation sites within the mouse IK peptide ^{358}SNHSAQDAVDNLLLLSK^{374} corresponding to a consensus sequence encoded by exon 8 and found in IK from mouse (NM_001025597), human (NM_006060.4), and chicken (NM_205088). S358 corresponds to S361 and S361 corresponds to S364 of human IK1. No phosphorylation was detected at any of the other 15 serines and 6 threonines present in eight additional IK peptides, including ^{16}E-^{42}G, ^{39}S-^{58}K, ^{97}D-^{110}R, ^{114}S-^{120}K, ^{186}D-^{193}R, ^{260}S-^{272}K, ^{260}S-^{266}R, and ^{274}S-^{284}K.

Molecular Model of SYK-Phosphorylated IK Segment. We constructed a molecular model for the SYK-phosphorylated segment of human IK1 by comparative homology modeling. The amino acid sequence containing S361 and S364 of human Ikaros (accession number: Q13422) was blasted against the Research Collaboratory for Structural Bioinformatics Protein Data Bank. The C-terminal segment of adenylate kinase 5 was retrieved to match the IK sequence 341–375 (PDB ID 2BWJ). The query and the subject sequence were further aligned using the global multiple alignement program ClustalW (11). The penalties for gap opening and extension were set at 10 and 0.1, respectively. Blosum62 was used as comparison matrix. The secondary structures were identified along with the template defining the α-helices, β-sheets, coils, and loops. The homology model was built by comparative homology modeling and subsequently refined using the Prime module within the molecular modeling suite Maestro from Schrodinger (www.schrodinger.com). The final model was energy-minimized. Molecular mechanics simulation used OPLS2005 force field and distance dependent dielectric constant within the MacroModel module.

We also constructed a structural complex model of SYK and the IK peptide (residues 341–375) to investigate the phosphorylation mechanism of SYK on S361 and S364 residues of IK at the atomic level. This model was constructed based on the X-ray crystal structure of SYK kinase domain (PDB ID 1XBA). The structure of IK peptide was built by Modeler (12) and homologous crystal templates (PDB ID 2BWJ and 3LQN). Conformation and the primary position of ATP were selected by homology modeling through structural superimposition between the crystal structures of protein kinase B (PDB ID 1O6K) and SYK using Sybyl6.8 (Tripos). Then, the target IK peptide was "grafted" into the catalytic site of SYK by using the computational protein-protein docking Web server GRAMM (13). After obtaining the initial structure of the SYK-ATP-IK peptide complex, the position of the peptide within a 6.5 Å radius around the catalytic site (14) was optimized by energy minimization using the AMBER force field implemented in the Sybyl with the following parameters: (*i*) a distance-dependent dielectric function, (*ii*) nonbonded cutoff 8 Å, (*iii*) Gasterger–Hückel charges for ATP, and (*iv*) amber charges for the protein and the peptide. The structure was minimized first by applying the simplex method, followed by the use of the Powell method to the energy gradient < 0.05 kcal/(mol·Å).

EMSAs. EMSAs were performed on purified recombinant proteins as well as nuclear extracts from 293T cells, as previously described (2, 6, 7). Preparation of nuclear extracts was carried out using CHEMICON's Nuclear Extraction Kit (Catalog No. 2900) (Millipore) with some modifications. Five million cells were spun down and washed with 1× PBS. Cells were resuspended in cytoplasmic buffer containing 0.5 mM DTT and a 1/1,000 dilution of protease inhibitor mixture by gentle inversion and were then incubated on ice for 15 min. The cell suspension was centrifuged at 250 × *g* for 5 min at 4 °C. The cell pellet was resuspended in ice-cold cytoplasmic lysis buffer and disrupted using a 27-gauge needle. The disrupted cell suspension was centrifuged at 8,000 × *g* for 20 min at 4 °C. The supernatant containing the cytosolic portion of the cell lysate was removed. The remaining pellet containing the nuclear portion of the cell lysate was resuspended in ice-cold nuclear extraction buffer containing 0.5 mM DTT and 1/1,000 dilution of protease inhibitor mixture. The nuclear suspension was gently agitated using an orbital shaker at 4 °C for 1 h and then centrifuged at 16,000 × *g* for 5 min at 4 °C to isolate the extracted nuclear protein fraction. The protein concentration of extracts was determined by using the Quick Start Bradford Protein Assay Kit (Catalog No. 500–0202) (BioRad). Oligonucleotide probes for EMSA were purchased from Integrated DNA Technologies and included *IK-BS1* (5′-TCAGCTTTT-GGGAATACCCTGTCA-3′) and *IK-BS5* (5′-TCAGCTTTT-GAGAATACCCTGTCA-3′) (4, 15). The *IK-BS1* oligonucleotide probe containing a high-affinity IK1 binding site was end-labeled with [γ-^{32}P]ATP (3,000 Ci/mmol) using T4 polynucleotide kinase and purified using a Nuctrap probe purification column (Stratagene). Three-microgram samples of the nuclear extracts and 1 ng of labeled *IK-BS1* probe (1×10^5 cpm/ng) were used in the DNA binding reaction. For competition reactions, a 60-fold excess of unlabeled *IK-BS1* probe was added before the addition of the labeled IK-BS1 probe. The IK-BS5 oligonucleotide has a single base pair (G > A) substitution at position 3 within the core consensus and does not bind IK (4). In addition, we used the γ-satellite probe A (5′-TATGGC GAGGAAAAC TGAAAAAG-GTGGAAAATTTAGAAATGT-3′ and 5′-ACATTTCTAAATT-TTCCACCTTTTTCAGTTTTCCTCGCCATA-3′) derived from the centromeric γ-satellite repeat sequences (16, 17). The γ-satellite A DNA probe contains two consensus IK binding sites in close proximity to each other and is a target for high-affinity binding of wild-type IK and the binding affinity to this probe shows an excellent correlation with the homing of IK to pericentromeric heterochromatin (PC-HC). EMSAs were also performed using the Thermo Scientific LightShift Chemiluminescent EMSA Kit (Catalog No. 20148) (Pierce) following the manufacturer's protocol (18). In these experiments, single-stranded *IK-BS1* and *IK-BS5* oligonucleotides were biotin-labeled using the Biotin 3′ End Labeling Kit (Catalog No. 89818) (Pierce). One-hundred nanomols of unlabeled oligonucleotides were incubated for 30 min at 37 °C in a reaction mixture containing 1× TdT Reaction Buffer (500 mM cacodylic acid, 10 mM $CoCl_2$, 1 mM DTT, pH 7.2), 0.5 μM Biotin-11-UTP, and 0.2 Units of TdT. 0.2 M EDTA was added to stop the reaction and biotin-labeled oligonucleotides were extracted using chloroform: isoamyl alcohol (24:1). Single-stranded biotinylated oligonucleotides were duplexed by mixing together equal amounts and incubating for 1 h at room temperature. Each binding reaction for EMSA included 10X Binding Buffer (100 mM Tris, 500 mM KCl, 10 mM DTT), 2.5% glycerol, 5 mM $MgCl_2$, 50 ng Poly (dI*dC), 0.05% Nonidet P-40, 4 μg nuclear protein extract, and 20 fmols of the biotin-labeled duplexed probe in a total

volume of 20 μL. Binding reactions were performed at room temperature for 20 min. A 6% nondenaturing polyacrylamide gel was prerun during the 20-min incubation time at 200 V in prechilled 0.5× TBE buffer (89 mM Tris base, 89 mM boric acid, 1 mM EDTA, pH ~8.0). Loading buffer was added five-times to each reaction sample and samples were loaded onto the polyacrylamide gel. Samples were electrophoresed at 100V and transferred at 380 mA (~50 V) for 30 min to a Biodyne B Nylon Membrane (Catalog No. 77016) (Thermo Scientific) soaked in 0.5× TBE buffer. When the transfer was complete, biotin-labeled DNA was cross-linked to the membrane at 120 mJ/cm^2 using a Spectrolinker XL-1000 UV cross-linker with 254-nm UV light bulbs. The biotin-labeled DNA was detected using a stabilized streptavidin- HRP conjugate and a highly sensitive chemiluminescent substrate according to the manufacturer's instructions (18). The membrane was exposed to X-ray film and developed with a film processor.

RT-PCR Analysis of IK Target Genes. RT and PCR were used to evaluate the expression levels of previously published and validated IK target genes (19). Total cellular RNA was extracted from cells using the QIAamp RNA Blood Mini Kit (Catalog No. 52304) (Qiagen) following the manufacturer's instructions. Oligonucleotide primers were ordered from Integrated DNA Technologies to amplify a 236-bp region of the *KIF23* transcript (Gene ID 9493: 5′-CGGAAACCTACCGTGAAAAA-3′ and 5′-AGTTCCTTCTGGGTGGTGTG-3′), a 168-bp region of the *TNFAIP8L2* transcript (Gene ID 79626: 5′-GGCACTTAGC-TTTGGTGAGG-3′ and 5′-AGCAGACCTGGGTCAGAGAA -3′), a 245-bp region of the *PREP* transcript (Gene ID 5550: 5′-TGAGCAGTGTCCCATCAGAG-3′ and 5′-CATCTTCGCT-GAACGCATAA-3′), a 232-bp region of the *DNAJC6* transcript (Gene ID 9829: 5′-GTCCTTCGCCCACAGTACAT-3′ and 5′-TTGCTGGCAAAGGAAGAACT-3′), and a 209-bp region of the *EIF4E3* transcript (Gene ID 317649: 5′-CCGCAGCAG-ATGATGAAGTA-3′ and 5′-GTGTTTTCCACGTCCACCTT-3′). The Qiagen One-Step RT-PCR Kit (Catalog No. 210212) (Qiagen) was used following manufacturer's instructions to amplify the target PCR products. The conditions were 1 cycle × (30 min 50 °C, 15 min 95 °C) and 35 cycles × (45 s 94 °C, 1 min 60 °C, 1 min 72 °C). PCR products were separated on a 1.2% agarose gel and visualized after ethidium bromide staining using a UVP Epi Chemi II Darkroom Transilluminator.

Confocal Laser Scanning Microscopy. Subcellular IK localization and colocalization studies using immunofluorescence and spinning disk confocal microscopy were performed as previously described (2, 4, 5, 17). In experiments using primary lymphocyte precursors from SYK-abundant vs. SYK-deficient ALL patients, confocal microscopy was performed using a BioRad MRC-1024 Laser Scanning Confocal Microscope (BioRad) equipped with a krypton/argon mixed gas laser and mounted on a Nikon Eclipse E800 series upright microscope equipped with high numerical objectives. TOTO-3 iodide from Molecular Probes was used for nuclear staining. Confocal images were obtained using a Nikon 40× or 60× (NA 1.4) objective and Kalman collection filter. Digitized images were saved on a Jaz disk (Iomega) and processed with the Adobe Photoshop software (Adobe Systems). During confocal imaging in other experiments, slides were imaged using the PerkinElmer Spinning Disk Confocal Microscope and the PerkinElmer UltraView ERS software or the Volocity V5.4 imaging software (PerkinElmer). The cover-slips were fixed with ice-cold MeOH at −20 °C for 10 min. The fixed cells were permeabilized and their nonspecific antibody binding sites blocked with 0.1% Triton X-100 and 10% goat serum in PBS for 30 min, respectively. To detect and localize the SYK and IK proteins, cells were stained with a mouse monoclonal anti-SYK antibody (sc-1240; Santa Cruz), an in house mouse monoclonal anti-IK antibody or a mouse monoclonal anti-HA antibody (HA-probe F-7, sc-7392) recognizing an internal region of the influenza HA protein as primary antibodies for 1 h at room temperature. Cells were washed with PBS and incubated with appropriate secondary antibodies conjugated with either Alexa Fluor 488 (Cat #: A11001; Invitrogen) or Alexa Fluor 568 (Cat #: A21069; Invitrogen) for 1 h. Cells were then washed with PBS and counterstained with the DNA-specific nuclear dye DAPI. The cover-slips were inverted, mounted onto slides in Vectashield (Vector Labs) to prevent photobleaching, and sealed with nail varnish. UltraCruz Mounting Medium containing 1.5 μg/mL of DAPI was purchased from Santa Cruz Biotechnology.

Bioinformatics and Statistical Analysis of Gene-Expression Profiles. In the analyses of gene-expression profiles of lymphocyte precursors with high vs. low SYK/Ikaros expression levels, we focused our analysis on validated IK target genes. The publicly available archived GSE32311 database (19) was used to compare gene expression changes in $CD4^+CD8^+$ double-positive wild-type ($n = 3$; GSM800500, GSM800501, GSM800502) vs. *Ikaros* null mouse thymocytes ($n = 3$, GSM800503, GSM800504, GSM800505) from the same genetic background of (C57BL/6 x129S4/SvJae). Gene-expression changes were screened using probe-level RMA signal intensity values from the mouse 430_2.0 Genome Array. To identify the gene signatures for up-regulated and down-regulated transcripts in *ikaros* knockout mice, we filtered changes greater than twofold and *t*-test *P* values less than 0.05 (*t*-test, unequal variances, Excel formula). Application of this filter identified 1,158 transcripts representing 924 genes that were down-regulated in *ikaros* null mice with a subset of 201 transcripts representing 137 genes exhibiting >twofold decreased expression levels. By cross-referencing this IK-regulated gene set with the archived CHiPseq data (GSM803110) using the Integrative Genomics Browser (20), we identified 45 IK target genes that harbored Ikaros/IKZF1 binding sites (Table S1). The Gene Pattern Web-based software (www.broadinstitute.org/cancer/software/genepattern) was used to extract expression values for human lymphocyte precursors from the NCBI Gene Expression Omnibus (GEO) database. We compiled 1,104 primary leukemia specimens from pediatric ALL patients (GSE3912, $n = 113$; GSE18497, $n = 82$; GSE4698, $n = 60$; GSE7440, $n = 99$; GSE13159, $n = 750$) to focus our analysis on 45 validated mouse IKZF1 target genes (Table S1) as well as 20 lymphoid-priming genes (21) (Fig. S1). Human orthologs of the mouse genes were identified by interrogating the gene symbols of the mouse genes using the curated online repository of HUGO Gene Nomenclature Committee-approved gene nomenclature (www.genenames.org).

For each study, the gene-expression values were transformed into SD units calculated from the mean and SD expression values for all of the samples in each study. Standardized values compiled from the five studies were rank-ordered according to the mean expression of SYK transcripts [207540_s_at, 209269_s_at, 226068_at and 244023_at (207540_s_at and 209269_s_at were common transcripts in all five Affymetrix platforms)]. Prospective power analysis was used to determine the SD cutoff for "high SYK/Ikaros expression" and "low SYK/Ikaros expression" samples in the datasets. We set the unadjusted critical *P* value at 2.5×10^{-6} to control for false-positive rate at 0.05 to detect significant differences in any one of the SYK/IK transcripts of ~20,000 transcripts common across the five Affymetrix platforms. At this critical *P* value, a total sample size greater than 254 would be sufficient to detect a difference of 1 SD unit with 99.9% power for SYK/IK transcripts. Therefore, samples were assigned to the "high SYK expression" group if their expression level was >0.5 SD unit higher than the mean expression level ($n = 285$) and to the "low SYK expression" group if their expression level was >0.5 SD unit lower than the mean expression level ($n = 270$). Similar calculation for IKZF1 rank-ordered expression levels three highly

correlated transcripts for IK/*IKZF1* (205038_at, 205039_s_at, 216901_s_at, 227344_at and 227346_at) resulted in characterization of 302 ALL samples with high IKZF1 expression and 318 samples with low IKZF1 expression. *t*-Tests were performed using standardized expression values combined from five datasets (two-sample, unequal variance correction, P values < 0.05 deemed significant) revealing an intersect of 34 transcripts representing 22 IK target genes (Table S1) and 14 transcripts representing 11 genes for lymphoid-priming genes (Fig. S1) that were significantly up-regulated in both high *SYK* and high IK/*IKZF1* expression groups.

We used a one-way agglomerative hierarchical clustering technique to organize expression patterns using the average distance linkage method such that genes (rows) having similar expression across patients were grouped together (average distance metric). Dendrograms were drawn to illustrate similar gene-expression profiles by joining pairs of closely related gene-expression profiles, whereby genes joined by short branch lengths showed the most similarity in expression profile across all samples (1, 2). The same analyses were also performed for the B-lineage subset of the ALL patients ($n = 884$) and gene-expression profiles of leukemic B-cell precursors from 264 B-lineage ALL patients with high SYK expression levels were compared with those of leukemic B-cell precursors from 151 B-lineage ALL patients with low SYK expression levels.

Site-Directed Mutagenesis. The full-length mouse IK cDNA (NM_001025597) was subcloned into the pCMV6-Entry precision shuttle vector (Cat# PS100001; Origene) at the restriction sites SgfI and MluI to generate the pCMV6-mIK mammalian cell expression vector. The pCMV6-mIK construct was then used as a backbone vector to generate the mIKS358A_S361A and mIKS358D_S361D mutant vectors encoding IK proteins with alanine or phosphomimetic aspartate mutations at the SYK phosphorylation sites S358 and S361 using the QuikChange II Site-Directed Mutagenesis Kit from Agilent Technologies (9). The synthetic oligonucleotide primers that were used to mutate the SYK phosphorylation sites of Ikaros were (*i*) for alanine mutations: Forward primer 5′-CCCCCCACGGGCCAACCATGCAGCACAGGAC-3′/Reverse primer 5′-GTCCTGTGCTGCATGGTTGGCCCGTGGGGGG-3′, and (*ii*) for aspartate mutations: Forward primer 5′-AGATGGCCCCCCACGGGACAACCATGATGCACAGGACGCCGTGG-3′/Reverse primer 5′-CCACGGCGTCCTGTGCATCATGGTTGTCCCGTGGGGGGCCATCT-3′. Similarly, we used our recently published plasmids for: (*i*) the 12A mutant mouse IK protein containing alanine mutations at the 11 CK2 target sites along with the protein phosphatase 1 (PP1) recognition motif (A465/467) (pcDNA3.1-HA-mIK-13A, 21–23A, 63A, 101A, 294A, 393–394A, 396A, 398A, 465–467A) and (*ii*) the 6D mutant mouse IK protein with combined aspartate phosphomimetic mutations of six N-terminal CK2 phosphorylation sites (pcDNA-HA-mIK-13D+21D+23D+63D+101D+294D) (9, 22) as backbone vectors to generate the mIK12A+S358A_S361A, mIK6D+S358A_S361A, mIK12A+S358D_S361D, and mIK6D+S358D_S361D vectors using the QuikChange II Site-Directed Mutagenesis Kit. The final vectors were subjected to DNA sequencing with a T7 forward primer 5′-TAATACGACTCACTAT-3′ (Invitrogen) at the DNA Sequencing Core Facility of the USC. The 293T cells were transfected with the IK expression vectors using the BioT transfection reagent (Bioland Scientifics). Functional studies were performed 72 h after transfection.

Site-directed mutagenesis was also performed for four specific tyrosine (Y) residues of IK (Y292, Y409, Y493, and Y499) simultaneously to generate the corresponding phenylalanine (F) quarto-mutant that was used as a control in functional assays. Among the 16 Y-residues found in the IK protein, these selected Y-residues for mutation represent the most likely tyrosine phosphorylation sites because of their high prediction scores (0.873, 0.568, 0.535, and 0.950, respectively) based on NetPhos 2.0, an online software used for sequence- and structure-based prediction of eukaryotic protein phosphorylation sites (www.cbs.dtu.dk/services/NetPhos). The mutagenesis primer sets were: Y292F-sense 5′-CCTGTCAGACATGCCCTTTGACAGTGCCAACTA-3′ Y292F-antisense 5′-TAGTTGGCACTGTCAAAGGGCATGTCTGACAGG-3′; Y409F-sense 5′-GCGCAGCGGCCTTATCTTCCTAACCAACCAC-3′, Y409F-antisense 5′-GTGGTTGGTTAGGAAGATAAGGCCGCTGCGC-3′; Y493F-sense 5′-AGTGTAACATGTGTGGTTTTCACAGCCAGGACAGG-3′, Y493F-antisense 5′-CCTGTCCTGGCTGTGAAAACCACACATGTTACACT-3′; Y499F-sense 5′-CACAGCCAGGACAGGTTCGAGTTCTCATCCC-3′, Y499F-antisense 5′-GGGATGAGAACTCGAACCTGTCCTGGCTGTG-3′. The mutations were confirmed by DNA sequencing using the primer: P262 5′-TCGGGAGAGAAAATGAATGG-3′.

siRNA Transfections. The 293T cells (ATCC; CRL-111268) were cultured in six-well plates with DMEM supplemented with 10% FBS. They were transfected with siRNA after reaching 70–80% confluence using ON-TARGET*plus* SMARTpool siRNA and DharmaFECT Transfection Reagent 4 (Catalog No. T-2004) (Thermo Scientific Dharmacon). Transfections were performed following the manufacturer's instructions. Next, 50 nM of ON-TARGET*plus* SMARTpool siRNA and transfection reagent were mixed in antibiotic-free complete media and added to adherent cells. In experiments aimed at evaluating the effects of *SYK* knockdown on native IK function, cells were transfected with the human SYK specific siRNA (On-Target*Plus* Cat# L-003176). Control siRNA included BTK siRNA (Catalog No. L-003107) and ON-TARGET*plus* Non-Targeting siRNA scrambled control pool (Catalog No. D-001810) as negative controls and IK siRNA as a positive control. Seventy-two hours after transfection with the siRNA, 293T cells were examined for the subcellular localization, DNA binding activity, and transcription factor function of native IK. In some experiments, siRNA-transfected 293T cells (3×10^5 cells per sample) were seeded onto poly-L-lysine–coated glass cover-slips (22×22 mm) after 48 h, allowed to adhere to the cover-slips placed in six-well plates by incubation overnight, and then transfected with expression vectors for the 12A or 6D mutant mouse IK proteins (2 μg per well) using the BioT transfection reagent (Bioland Scientifics). At 48 h after the second transfection, 293T cells were examined for the subcellular localization of the recombinant mutant IK proteins.

Establishment of U373 Cells with Ecdysone-Inducible *SYK* Gene Expression. SYK-deficient U373 cells were transfected with the ecdysone inducible system regulatory vector, pVgRXR, and with a pIND/GS vector containing the cDNA encoding wild-type human *SYK* gene (H-L28824MI) (Invitrogen) using previously published procedures (2, 23). pIND-*SYK* was linearized with AatII and pVgRXR was linearized with MIuI. Ten-million U373 cells were transfected with 12 μg of each linearized vector by electroporation using a BioRad GenePulser II Electroporator (240 V/950 μF). Two days after electroporation, cells were reseeded in 150 × 15-mm tissue culture dishes and stably transfected cells were selected with 500 μg/mL G418-sulfate and 500 μg/mL zeocin. Individual clones were screened for the inducible expression of SYK after 24-h treatment with 10 μM ponasterone (Pon-A) (an analog of ecdysone) (Invitrogen) by Western blot analysis (2).

Experimental Procedures for Gene Expression Analysis in *SYK*-Transfected U373 Cells. We measured levels of RNA transcripts using the Human Genome U95 Av2 GeneChip microarrays (HG-U95Av2) from Affymetrix, which interrogates the expression levels of 9670 genes/EST clusters, as previously reported (2).

Total cellular RNA was extracted from a minimum of 10×10^6 cells with the use of the TriPure isolation reagent (Boehringer Mannheim). DNA was removed from this RNA preparation by DNase treatment (RQ1 RNase-Free DNase; Promega) and then purified by three rounds of organic (phenol-chloroform) extraction followed by ethanol precipitation. Further purification of the RNA was achieved by binding to an RNeasy (Qiagen) column before spectrophotometric quantitation for in vitro transcription reactions. First-strand cDNA was synthesized with a T7- $(dT)_{24}$ primer and second-strand cDNA was synthesized with *E. coli* DNA polymerase I and ligase. The cDNA product was in vitro transcribed and labeled with the Enzo BioArray High Yield RNA transcript labeling kit. The in vitro transcription products were purified on an RNeasy spin column before fragmentation in buffer containing 40 mM Tris-Acetate (pH 8.1), 100 mM KOAc, and 30 mM MgOAC for 35 min at 94 °C. RNA processing and hybridization to the U95Av2 GeneChip oligonucleotide microarray was performed according to the manufacturer's protocol. The test chips were stained and washed on fluidics station as per the manufacturer's recommendations. Arrays were scanned with the Genechip System confocal scanner manufactured for Affymetrix by Agilent. Microarray Suite 5.0 software from Affymetrix was used to determine the relative abundance of each gene based on the average difference of intensities. Each human gene in this platform is represented by at least one probe set composed of multiple probe pairs (16–20 pairs). Each probe pair consists of two sets of 25mer sequences. One set is a perfect match and the other set has a 13th base mismatch to serve as an internal control for the signal produced by the perfect match probe. A quantitative Signal metric was used to measure the level of each transcript on the chip (developed by Affymetrix). The algorithm calculates the signal using a one-step Tukey's Biweight Estimate, which determines a weighted mean that is relatively insensitive to outliers. The estimated real signal is calculated by subtracting the log of the perfect match intensity from the stray signal estimate. The mismatch signal is used to estimate the stray signal where appropriate. The probe pair was weighted more strongly to calculate the signal if the signal is closer to the median value. Using the Affymetrix signal to noise evaluation criteria, we filtered genes that were deemed significantly above background at both days after SYK induction or in both transfected controls for up-regulated genes and down-regulated genes, respectively. Control expression values were calculated by pooling day 1 and day 2 untransfected/uninduced and transfected/uninduced values and the induced values were averaged for transfected/induced SYK-expressing test samples for day 1 and day 2.

A one-way hierarchical clustering technique was used to organize the expression patterns such that genes (in columns) having similar expression on days 1 and 2 after SYK induction were grouped together (JMP Software, SAS). Expression values were analyzed first by joining pairs of genes (in columns) or treatments (in rows) closest in the average expression values (average distance linkage) and then connecting larger groups of probe sets or treatments using the branching structure, whereby larger branch lengths represent larger differences in expression profiles. Raw data were deposited (GEOSeries GSE18798) (2). Mouse IK target genes were also cross-referenced with this dataset. Following normalization by scaling, we compared control versus SYK-induced using $\log_{10}$-transformed mean centered values to the four control samples to calculate significance using planned linear contrasts. We constructed a General Linear Model with three factors [(*i*) treatment (control vs. SYK induction); (*ii*) gene level expression (30 genes represented by 36 transcripts of the 45 IK target genes identified in the Mouse study); (*iii*) a random effect identifying each gene chip] and one interaction term (Gene × Treatment) for the analysis of differential gene expression]. The consistent effect of "treatment" (PonA treatment used to induce SYK expression) for each gene was accounted for by the variance component in the "chip" factor determined using REML (Restricted or Residual Maximum Likelihood) method for fitting mixed models (JMP, SAS). We examined the distribution of the residuals of the model for equal dispersion around the line of best fit. The least-squares best-fit parameters from this model were used to design planned linear contrasts to determine significant gene level treatment effects from mean and SEs calculated in the interaction term. Two-tailed tests for differences between the least-square means with P values less than 0.05 were deemed statistically significant. Contrasts compared the effect sizes such that the combination of linear parameters to be jointly tested sum to zero for each level of the contrast. Least-square mean values were used to construct contrasts between each of the control groups (set to −1) and the SYK induced treatment group (set to 1) for each gene.

Genes Involved in Lymphoid-Lineage Affiliated Transcriptional Program Controlled by IK. Twenty lineage-affiliated genes (*FLT3*, *NOTCH1*, *LTB*, *BTLA*, *CD52*, *CLNK*, *IL7R*, *CCR9*, *DNTT*, *IGJ*, *SATB1*, *SOX4*, *RUNX2*, *MEF2C*, *RAG1*, *HMGA2*, *CNN3*, *PTGER2*, *ETS1*, *CSF1R*) implicated in lymphoid priming were compiled from mouse studies (21, 24, 25) for evaluation of their expression levels in lymphocyte precursors from primary leukemia specimens of children with ALL (Fig. S1).

Genomic PCR Analysis of the *Ikaros/IKZF1* Gene in Leukemia Cells. DNA sequencing was carried out using the BigDye Terminator v.3.1 cycle sequencing kit (Applied Biosystems), as previously reported (1). Total genomic DNA was extracted from patients' leukemia cells using the Qiagen DNeasy Blood and Tissue kit (Catalog No. 6950) according to the manufacturer's specifications. PCR products encompassing exons 1–8 and their respective exon-intron junctions were amplified using specific genomic PCR primers (Fig. S6*A*). For PCR amplification of exon 8, we used the "primer-walking" strategy. The primer sets [0.7 μL of each primer (50 pmol/μL), 50 μL reaction volume, 150 ng genomic DNA, 0.5 μL of 10 mM dNTP, and 2.5 UTaq polymerase (Invitrogen-Cat. No. 12355–036)] were used with the following thermal cycling conditions: 1 cycle for the initial denaturation (5 min 95 °C), 32 cycles (30 s 95 °C, 30 s 58 °C, 1 min 72 °C); hold at 72 °C for 5 min; indefinite hold at 4 °C. The PCR products were cleaned using a QIAquick PCR purification Kit (Qiagen) and then sequenced using the indicated PCR primers (Fig. S6*B*) in 5-μL reaction volumes containing 0.5 μL BigDye terminator mix v3.1, 1 μL 5× sequencing buffer (Applied Biosystems), 1 μL 3.2 pmol primer, and 25–30 ng template. Sequencing thermal cycling parameters were: 1 cycle (1 min, 96 °C), 35 cycles (10 s at 96 °C, 5 s at 50 °C, 150 s at 60 °C); hold 180 s at 60 °C, and indefinite hold at 4 °C. The sequencing products from each reaction were cleaned using GenScript QuickClean 5M purification kit (GenScript) and analyzed on an ABI 3730XL DNA Analyzer (Applied Biosystems). The PCR products from three cases of ALL are depicted in Fig. S6*C*. Sequences obtained from the genomic PCR products were analyzed using the SeqMan II contiguous alignment software in the LaserGene suite from DNASTAR and the MegAlign multisequence alignment software in comparison with the wild-type *Ikaros* sequence (NCBI Reference Sequence: Gene ID 10320, www.ncbi.nih.gov), as previously described (1).

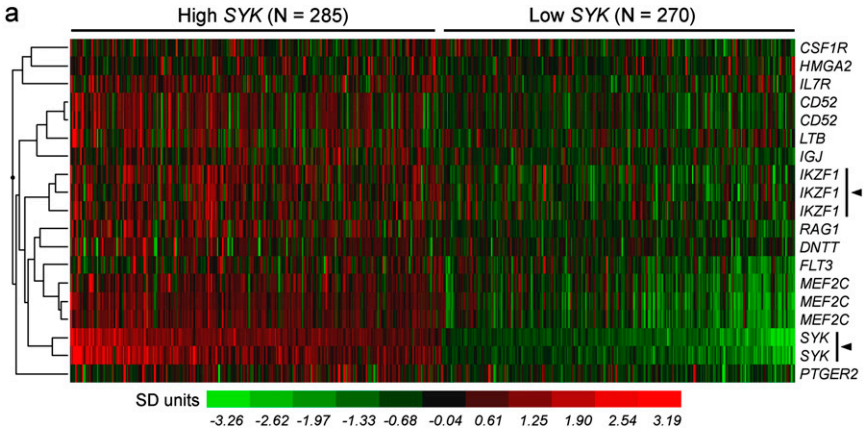
Signaling Experiments with the FL112 Human Pro-B Cell Line and SYK Inhibitor PCT. CD19 plays a critical role in initiation of SYK dependent signaling events in both immature and mature B-lineage lymphoid cells (8, 26–29). FL112 is a $CD19^+CD10^+CD34^+BCR^-$ human pro–B-cell line with germ-line Ig genes (8). Treatment of FL112 cells with an anti-CD19 monoclonal antibody conjugate (CD19×CD19) triggers activation of the Src family tyrosine ki-

nase LYN and both serine as well as tyrosine phosphorylation of the CD19 receptor (8). CD19-receptor engagement with monoclonal antibodies or recombinant human CD19 ligand has also been shown to trigger SYK activation (26, 27). CD19 was also shown to play a critical role in B-cell activation in part by recruiting SYK to membrane-bound microclusters (28). We used FL112 cells in stimulation experiments with CD19×CD19 (1 μg/mL × 15 min) to examine the effects of the SYK inhibitor PCT on SYK-dependent IK serine phosphorylation using standard biochemical assays (8, 29).

Differentiation Experiments with the ALL-1 Human Prepre-B Cell Line. In differentiation experiments, we examined the immediate effects of wild-type and mutant murine IK proteins on the maturation stage of the human prepre-B ALL cell line ALL-1 by using the transcript and protein expression levels of the mature B-cell antigen CD20 as a differentiation marker. Cells (500,000 per sample in 100 μL Amaxa human B-cell Nucleofector Solution) were transfected by nucleofection with the expression vectors (8 μg plasmid DNA/sample) for the wild-type IK or SYK-resistant mutant IK M1-1 using the Amaxa Nucleofector II Device and Amaxa Human B-cell Nucleofector Kit (Catalog No. VPA-1001; Lonza) according to the Nucleofector Program U-015. After transfection, cells were transferred to a six-well tissue culture plate and cultured in RPMI supplemented with 10% FBS (1.5 mL final volume per well) for 96 h at 37 °C/5% CO_2. After 96 h of culture, cells were examined for expression of the *CD20* and *CD22* genes using RT-PCR. Total cellular RNA was extracted from cells using the QIAamp RNA Blood Mini Kit (Catalog No. 52304) (Qiagen) following the manufacturer's instructions. Oligonucleotide primers were ordered from Integrated DNA Technologies to amplify a 240-bp region of the *CD20* cDNA (NCBI Gene ID 931: 5′-GCTGCCATTTCTGGAATGAT-3′ and 5′-TTCCTGGAAGAAGGCAAAGA-3′), and the P10 primer set (Forward: ATCCAGCTCCCTCCAGAAAT, Reverse: CTTCCCATGGTGACTCCACT) to amplify a 213-bp region (c.433–c.645) of exon 4 of the *CD22* cDNA (NCBI Gene ID 933). After 96 h of culture posttransfection, cells were also examined for surface expression of CD19 and CD20 proteins by standard two-color immunofluorescence and multiparameter flow cytometry using PE-labeled anti-human CD19 monoclonal antibody (BD Biosciences; Cat. #: 555413) and FITC-labeled anti-human CD20 monoclonal antibody (BD Biosciences; Cat. #: 556632) (1, 8). The labeled cells were analyzed using a The LSR II flow cytometer from Becton Dickinson. To determine the transfection efficiency, cells in duplicate samples were electroporated in parallel with the EGFP reporter vector pmaxGFP provided in the Amaza Nucleofector Kit, cultured for 48 h and EGFP-expressing cells were identified using a Nikon Eclipse TS100 inverted microscope. A B-2A long-pass emission fluorescence filter with an excitation of 450–490 nm was used to detect EGFP expressing cells with excitation maximum at 488 nm and emission maximum at 510 nm.

1. Uckun FM, Goodman P, Ma H, Dibirdik I, Qazi S (2010) CD22 EXON 12 deletion as a pathogenic mechanism of human B-precursor leukemia. *Proc Natl Acad Sci USA* 107 (39):16852–16857.
2. Uckun FM, Qazi S, Ma H, Tuel-Ahlgren L, Ozer Z (2010) STAT3 is a substrate of SYK tyrosine kinase in B-lineage leukemia/lymphoma cells exposed to oxidative stress. *Proc Natl Acad Sci USA* 107(7):2902–2907.
3. Uckun FM, et al. (1996) BTK as a mediator of radiation-induced apoptosis in DT-40 lymphoma B cells. *Science* 273(5278):1096–1100.
4. Sun L, et al. (1999) Expression of dominant-negative and mutant isoforms of the antileukemic transcription factor Ikaros in infant acute lymphoblastic leukemia. *Proc Natl Acad Sci USA* 96(2):680–685.
5. Uckun FM, Ozer Z, Qazi S, Tuel-Ahlgren L, Mao C (2010) Polo-like-kinase 1 (PLK1) as a molecular target to overcome SYK-mediated resistance of B-lineage acute lymphoblastic leukaemia cells to oxidative stress. *Br J Haematol* 148(5):714–725.
6. Uckun FM, Ozer Z, Vassilev A (2007) Bruton's tyrosine kinase prevents activation of the anti-apoptotic transcription factor STAT3 and promotes apoptosis in neoplastic B-cells and B-cell precursors exposed to oxidative stress. *Br J Haematol* 136:574–589.
7. Mahajan S, et al. (2001) Transcription factor STAT5A is a substrate of Bruton's tyrosine kinase in B cells. *J Biol Chem* 276(33):31216–31228.
8. Uckun FM, et al. (1993) Signal transduction through the CD19 receptor during discrete developmental stages of human B-cell ontogeny. *J Biol Chem* 268(28):21172–21184.
9. Uckun FM, Ek RO, Jan ST, Chen CL, Qazi S (2010) Targeting SYK kinase-dependent anti-apoptotic resistance pathway in B-lineage acute lymphoblastic leukaemia (ALL) cells with a potent SYK inhibitory pentapeptide mimic. *Br J Haematol* 149(4):508–517.
10. Xia Z, Guo M, Ma H (2011) Functional analysis of novel phosphorylation sites of CREB-binding protein using mass spectrometry and mammalian two-hybrid assays. *Proteomics* 11(17):3444–3451.
11. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
12. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
13. Tovchigrechko A, Vakser IA (2005) Development and testing of an automated approach to protein docking. *Proteins* 60(2):296–301.
14. Huang Z, et al. (2011) ASD: A comprehensive database of allosteric proteins and modulators. *Nucleic Acids Res* 39(Database issue):D663–D669.
15. Molnár A, Georgopoulos K (1994) The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. *Mol Cell Biol* 14(12):8292–8303.
16. Cobb BS, et al. (2000) Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev* 14(17):2146–2160.
17. Gurel Z, et al. (2008) Recruitment of Ikaros to pericentromeric heterochromatin is regulated by phosphorylation. *J Biol Chem* 283(13):8291–8300.
18. Jin H, Ralston SH (2012) Analysis of transcriptional regulation in bone cells. *Methods Mol Biol* 816:233–247.
19. Zhang J, et al. (2012) Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat Immunol* 13(1): 86–94.
20. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
21. Yoshida T, Ng SY, Georgopoulos K (2010) Awakening lineage potential by Ikaros-mediated transcriptional priming. *Curr Opin Immunol* 22(2):154–160.
22. Popescu M, et al. (2009) Ikaros stability and pericentromeric localization are regulated by protein phosphatase 1. *J Biol Chem* 284(20):13869–13880.
23. No D, Yao T-P, Evans RM (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* 93(8):3346–3351.
24. Dias S, Månsson R, Gurbuxani S, Sigvardsson M, Kee BL (2008) E2A proteins promote development of lymphoid-primed multipotent progenitors. *Immunity* 29(2):217–227.
25. Yoshida T, Ng SY, Zuniga-Pflucker JC, Georgopoulos K (2006) Early hematopoietic lineage restrictions directed by Ikaros. *Nat Immunol* 7(4):382–391.
26. DeFranco AL (1996) The two-headed antigen. B-cell co-receptors. *Curr Biol* 6(5): 548–550.
27. Uckun FM, et al. (2011) Targeting human B precursor acute lymphoblastic leukemia cells with recombinant human CD19 ligand. Available at https://ash.confex.com/ash/2010/webprogram/Paper32270.html. Accessed December 4, 2010.
28. Depoil D, et al. (2008) CD19 is essential for B cell activation by promoting B cell receptor-antigen microcluster formation in response to membrane-bound ligand. *Nat Immunol* 9(1):63–72.
29. Uckun FM, Sun L, Qazi S, Ma H, Ozer Z (2011) Recombinant human CD19-ligand protein as a potent anti-leukaemic agent. *Br J Haematol* 153(1):15–23.

a



b

Affy ID *Up in High SYK and Ikaros*	Genesymbol	StdDev [High *SYK* (N = 285) – Low *SYK* (N = 270)]	T-test P-value	StdDev [High Ikaros (N = 302) – Low Ikaros (N = 318)]	T-test P-value
209200_at	*MEF2C*	1.35	2.6×10^{-58}	0.97	4.4×10^{-34}
209199_s_at	*MEF2C*	1.25	1.9×10^{-45}	0.82	3.0×10^{-25}
207968_s_at	*MEF2C*	1.02	2.8×10^{-32}	0.70	8.2×10^{-18}
206591_at	*RAG1*	0.82	6.4×10^{-25}	0.66	6.6×10^{-16}
204661_at	*CD52*	0.67	5.5×10^{-15}	0.35	1.7×10^{-5}
34210_at	*CD52*	0.66	1.5×10^{-14}	0.33	5.1×10^{-5}
206631_at	*PTGER2*	0.60	4.3×10^{-12}	0.57	3.8×10^{-12}
206674_at	*FLT3*	0.58	6.3×10^{-11}	0.50	5.8×10^{-11}
205798_at	*IL7R*	0.49	1.2×10^{-8}	0.49	8.6×10^{-9}
212592_at	*IGJ*	0.48	7.4×10^{-10}	0.41	9.8×10^{-7}
210487_at	*DNTT*	0.34	9.7×10^{-5}	0.50	2.3×10^{-9}
207339_s_at	*LTB*	0.29	8.6×10^{-4}	0.41	4.0×10^{-7}
203104_at	*CSF1R*	0.26	1.6×10^{-3}	0.50	2.5×10^{-9}
208025_s_at	*HMGA2*	0.16	5.0×10^{-2}	0.29	7.8×10^{-5}

Fig. S1. Effects of *SYK* expression levels on expression levels of lymphoid-priming genes in primary lymphocyte precursors from ALL patients. (*A*) Coregulation of *IKZF1*, *SYK* and lymphoid-priming genes in ALL patients. Gene Pattern (www.broadinstitute.org/cancer/software/genepattern) was used to extract expression values for the 20 lymphoid-priming genes in the combined dataset from five studies with a total of 1,104 primary leukemia samples of human lymphocyte precursors for further analysis. For each study, the SD values were calculated from the study mean for all of the patients. The datasets were combined to test for consistent differences in the *z*-scores for high SYK (>0.5 SD units; $n = 285$ samples) and low SYK ($\leq$ 0.5 SD units; $n = 270$ samples) groups. Fourteen transcripts representing 11 lymphoid-priming genes were up-regulated in specimens with both high *IKZF1* and *SYK* expression. A one-way agglomerative hierarchical clustering technique was used to organize expression patterns using the average distance linkage method, such that genes (rows) having similar expression across patients were grouped together (average distance metric). The heat map depicts expression values represented by SD units above (red) and below the mean (green). Dendrograms were drawn to illustrate similar gene-expression profiles from joining pairs of closely related gene-expression profiles, whereby genes joined by short branch lengths showed most similarity in expression profile across patients. Hierarchical cluster analysis revealed a subset of four genes [*MEF2C* (three transcripts), *FLT3*, *DNTT*, and *RAG*] that were highly coregulated with three transcripts of *IKZF1* and two transcripts of *SYK*. (*B*) Expression values expressed as SD units calculated from 1104 samples were compiled for the five studies and rank-ordered according to the mean expression of three highly correlated SYK transcripts [207540_s_at, 209269_s_at, 226068_at and 244023_at (207540_s_at and 209269_s_at were common transcripts in all five Affymetrix platforms). These samples were also rank ordered according to IKZF1 expression level (205038_at, 205039_s_at, 216901_s_at, 227344_at and 227346_at; three of these were common in all Affymetrix platforms (205038_at, 205039_s_at, 216901_s_at)]. *t*-Tests were performed for the combined SD units from the five datasets (two-sample, unequal variance correction, P values < 0.05 deemed significant) to reveal 14 transcripts significantly up-regulated in specimens with both high *SYK* and high *IKZF1* expression. The five most significantly up-regulated lymphoid priming genes in specimens with high *SYK* expression were *MEF2C* (three transcripts ranging from 1.35 to 1.02 SD units, $P = 2.6 \times 10^{-58}$ to 2.8×10^{-32}), *RAG1* (0.82 SD Units, $P = 6.4 \times 10^{-25}$), *CD52* (two transcripts ranging from 0.66 to 0.67 SD units, $P = 5.5 \times 10^{-15}$ to 1.5×10^{-14}), *PTGER2* (0.60, $P = 4.3 \times 10^{-12}$), and the lymphoid lineage-promoting gene *FLT3* (0.58, $P = 6.3 \times 10^{-11}$).

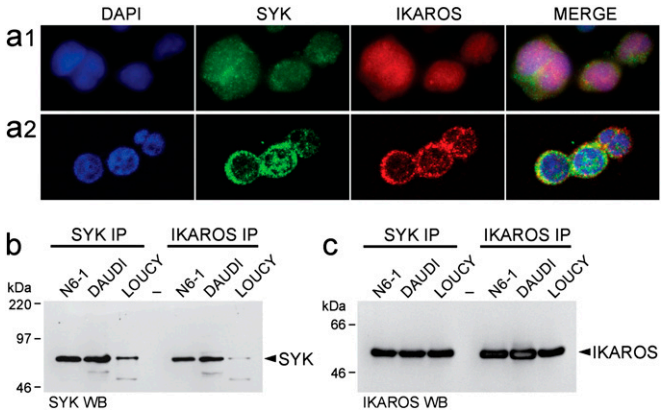


Fig. S2. Colocalization and physical interactions of native IK and SYK proteins in human cells. (*A*) Nuclear colocalization of native IK and SYK. NALM-6 and 293T cells (depicted in *1* and *2*, respectively) were fixed and stained with polyclonal rabbit anti-IK1 (primary Ab)/Alexa Fluor 568 F(ab')$_2$ fragment of goat anti-rabbit IgG (secondary Ab) (red) and mouse anti-SYK MoAb (primary Ab)/Alexa Fluor 488 goat anti-mouse IgG (secondary Ab) (green) antibodies. Nuclei were stained with blue fluorescent dye DAPI. MERGE panels depict the merged three-color confocal image showing colocalization of IK1 and SYK in DAPI-stained nucleus as magenta immunofluorescent foci, as well as colocalization of IK1 and SYK in the cytoplasm as "enhanced yellow" immunofluorescent foci. (Magnification: 400×.) (*B* and *C*) Coimmunoprecipitation of native IK and SYK. (*B*) Depiction of the results of the SYK Western blot analysis of SYK and IK immune complexes immunoprecipitated from the N6-1, DAUDI, and LOUCY cell lines. (*C*) Depiction of the results of the IK Western blot analysis of SYK and IK immune complexes from the same cells.

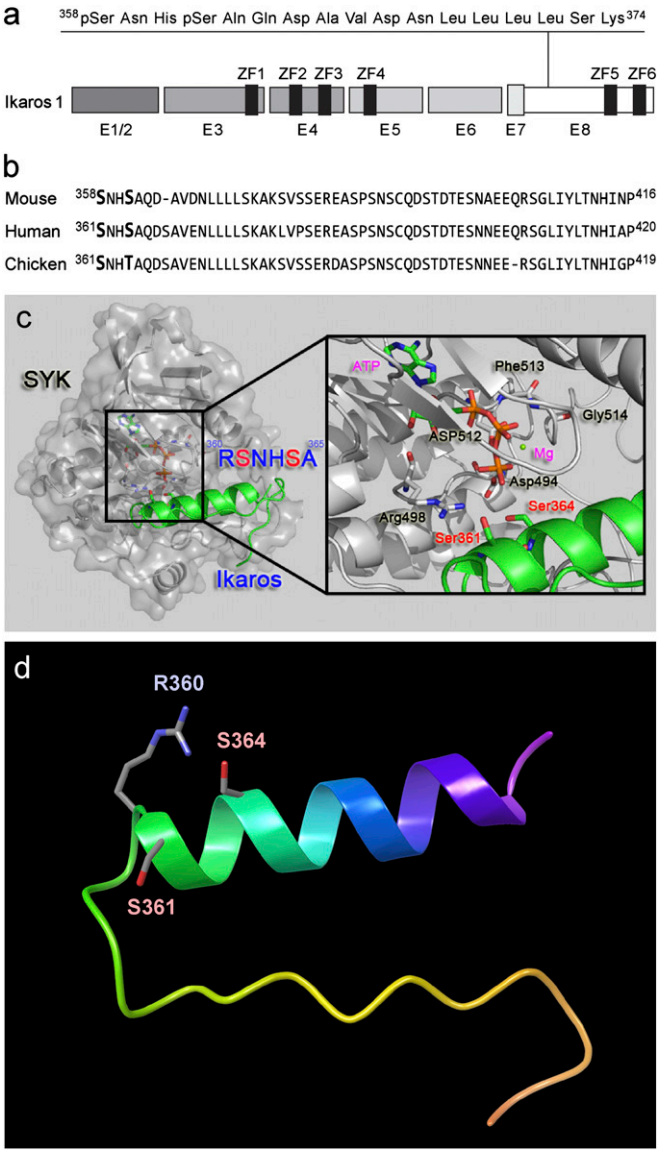


Fig. S3. SYK phosphorylation sites of IK. (*A*) A MS analysis was performed on trypsin-digested recombinant IK after an in vitro kinase reaction with purified recombinant SYK. We identified Ser[358] (S358) and Ser[361] (S361) as two unique SYK phosphorylation sites within the mouse Ikaros peptide [358]SNHSAQ-DAVDNLLLLSK[374] corresponding to a consensus sequence encoded by exon 8 and found in IK from mouse (NM_001025597), human (NM_006060.4), and chicken (NM_205088). [(*B*) Allignment of mouse, human, and chicken IK protein segments containing the identified SYK phosphorylation sites. S358 corresponds to S361 and S361 corresponds to S364 of human IK1. (*C*) Schematic diagram of SYK-Ikaros (341–375) complex. SYK is shown as molecular surface colored in gray, and IK is shown as secondary structure colored in green. Motif DFG (Asp512-Phe513-Gly514) and other key residues in catalytic site of SYK, ATP/Mg^{2+} and substrate residues (Ser361 and Ser364) in the peptide of IK are shown as a stick model. The model was built by Modeler and shown with Pymol. S361 and S364 are located at the N terminus of an α-helical secondary structure. This model shows that a helix of target IK peptide would readily bind to SYK catalytic site with a compact conformation because of its narrow and deep shape, leading to a microenvironment of transphosphorylation composed of ATP, Mg ion, S361, and S364 residues of IK, the activation DFG loop (Asp512-Phe513-Gly514) in SYK, and other potential key residues (e.g., Asp494 and Arg498 of SYK). (*D*) Locally stabilizing effects of SYK-induced phosphorylation of IK. Protein model of human IK amino acids 341–375. Side-chains of R360, S361, and S364 are represented by stick models. Based on the model, S361 and S364 are located at the N terminus of an α-helical secondary structure. S364 forms a close contact with the side-chain of R360. Our model indicates that phosphorylation of S364 would enhance the side-chain interaction by a positive-negative salt bridge formation and thereby stabilize the folded protein conformation of IK in this segment. S361 is positioned at the N-terminal cap of the α-helix. According to this model, the positively charged environment constituted by the macrodipole of the α-helix favors the interaction with a SYK-phosphorylated S361, which further stabilizes the local protein conformation.

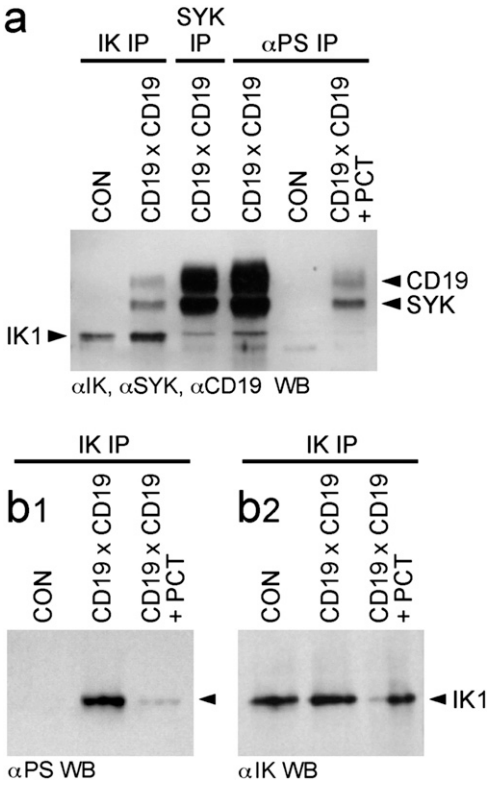


Fig. S4. SYK inhibitor PCT inhibits serine phosphorylation of IK in FL112 human Pro-B cells after CD19-receptor mediated SYK activation. (*A*) Western blot analysis of anti-IK, anti-SYK, and antiphosphoserine immunoprecipitates with a mixture of rabbit polyclonal antibodies reactive with IK, SYK, and CD19. CD19-receptor engagement with an anti-CD19 monoclonal antibody homoconjugate (CD19×CD19) (1 μg/mL) for 15 min promoted a physical association between CD19, SYK, and IK, as documented by detection of CD19 and SYK proteins in IK immune complexes. None of the three proteins were readily detectable in antiphosphoserine immunoprecipitates from lysates of unstimulated control FL112 cells. CD19 engagement caused serine phosphorylation of all three proteins, as documented by detection of all three proteins in antiphosphoserine immunoprecipitates from lysates of FL112 cells stimulated with CD19×CD19. Increased serine phosphorylation of IK in CD19×CD19 stimulated FL112 cells was abrogated by the SYK inhibitor PCT (50 μM). (*B*) Antiphosphoserine immunoblotting of IK immune complexes. CD19 engagement triggered serine phosphorylation of IK in a SYK-dependent manner as it was blocked by the SYK inhibitor PCT (50 μM).

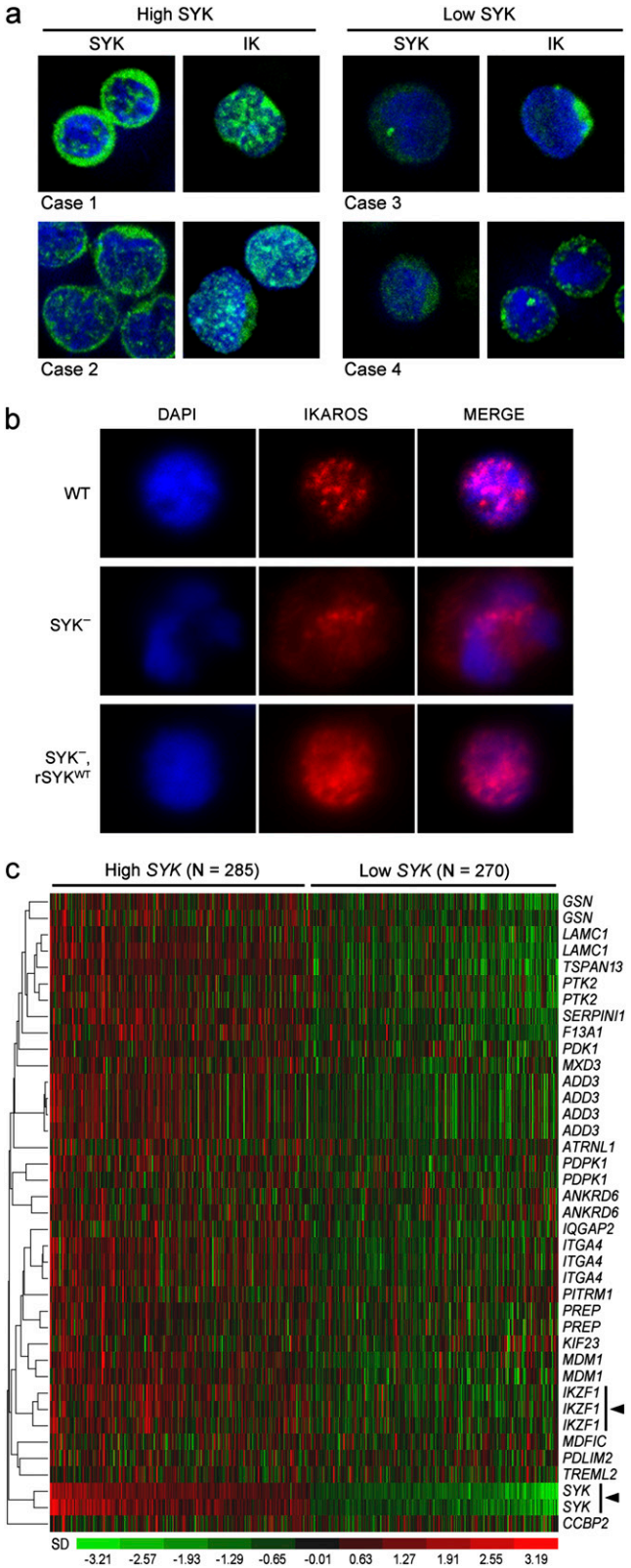


Fig. S5. SYK expression levels control nuclear localization and transcription factor function of IK in lymphoid cells. (*A*) Depicted are the confocal two-color (green/blue: SYK/TOTO-3 and IK/TOTO-3) merge images of leukemic cells from B-precursor ALL patients. Normal punctate nuclear staining for IK was observed in the confocal fluorescence images of primary leukemic B-cell precursors from case 1 and case 2 with wild-type SYK consistent with the localization of IK to PC-HC. In contrast, SYK-deficient leukemic cells from case 3 and case 4 showed an aberrant, predominantly cytoplasmic localization of IK. No homozygous IK mutations or deletions were detected by genomic PCR in case 3 or case 4 with abnormal IK localization (Fig. S6). Previous studies on leukemic cells from case 3 identified a 113-bp deletion (G^{1870}–A^{1982}) in the aberrant coding sequence of *syk* mRNA that corresponds to deletion of exon 12 and results in a truncating frame-shift mutation starting after Arg^{574} of the catalytic domain with the addition of 19 novel amino acids until the stop codon (1). In the *syk* coding-sequence PCR amplified from

Legend continued on following page

leukemic cells of case 4, a 141-bp deletion (G^{1729}–A^{1869}) corresponding to deletion of exon 13 was found, which results in an in frame deletion of 47 amino acids (Ala^{528}–Arg^{574}) in the SYK catalytic domain (1). (Magnification: 500×.) (*B*) Depicted are the representative confocal two-color (red/blue) Ikaros/DAPI merge images of wild-type DT40 cells, SYK-deficient DT40 cells, and SYK-deficient DT40 cells reconstituted with wild-type *syk*. Wild-type DT40 cells and SYK-deficient DT40 cells reconstituted with wild-type *syk* (but not the SYK-deficient DT40 cells) showed normal IK localization. (Magnification: 630×.) (*C*) A one-way agglomerative hierarchical clustering technique was used to organize gene expression patterns of lymphocyte precursors from ALL patients using the average distance linkage method such that genes (rows) having similar expression across patients were grouped together (average distance metric) (*SI Text*). The heat map depicts expression values represented by SD units above (red) and below the mean (green). Dendrograms were drawn to illustrate similar gene-expression profiles by joining pairs of closely related gene-expression profiles, whereby genes joined by short branch lengths showed most similarity in expression profile across patients.

1. Goodman PA, Wood CM, Vassilev A, Mao C, Uckun FM (2001) Spleen tyrosine kinase (Syk) deficiency in childhood pro-B cell acute lymphoblastic leukemia. *Oncogene* 20(30):3969–3978.

a

	PCR primer (5′ → 3′)	Exon Number	Product (bp)
F	GGTTATGACTCCTTGGTGAAACGCGGCGCATCCCAGCCTGGGCGGGACGCTCGGCCG	1	636
R	TGAGCGTCAAACCTGCGGTC		
F	GCAGTTGCTGCTCAGATATG	2	662
R	ATTCCTCCCCCATGAGCATA		
F	5ATAAATCCTAGTGTACCCGT	3	756
R	AAAGCTTGGACATTTTGCGG		
F	TCAAGGCTGAATGCACGGCG	4	577
R	TACCACTTGGAACCAATCGC		
F	CCCAGCCAGTGAAGCGTTAA	5	646
R	CATACCAGCCTGACAAACGG		
F	CAGGAATTTCACCAAGTCCG	6	533
R	CTAGGAGTTTCTCCCGATTG		
F	GTCATTGGCATTGCTCCGGC	7	785
R	CACTTCTCAGGGGTCAGTCA		
F	TGCAGGTGTGATGTGTCCGC	8-1	1276
R	CTTAGCACAGGGTAGAGCAC		
F	CCTGTCGCTTCCTAGAATCC	8-2	1289
R	TCATTGCCTAGGTGGTTGAG		
F	AGAGTAATGCCACCAGATCC	8-3	1883
R	GGCAAGAAGAGGACACTACA		
F	GGTACCGCTGTGTACTACTG	8-4	1737
R	GCTGGGATGCGCCGCGTTTCACCAAGGAGTCATAACCTCCAAGA		

b

Sequencing Primer	Sequence (5′ → 3′)	Exon Number
IK-seq1rn	TTCACCAGTTACCCCAAGTT	1
IK-seqF2N	GCAGTTGCTGCTCAGATATG	2
IK-seq3	CCAGGCACCTTGACCATGAC	3
IK-seqR4	TACCACTTGGAACCAATCGC	4
IK-seqF5	CCCAGCCAGTGAAGCGTTAA	5
IK-seqF6	CAGGAATTTCACCAAGTCCG	6
IK-seqR6	CTAGGAGTTTCTCCCGATTG	6
IK-seqF7N	GTCATTGGCATTGCTCCGGC	7
IK-seqF8-1	TCAGCTGTTGCTGCCAGACC	8-1
IK-seqF8-2	TGTACAAGTGCGAACACTGC	8-1
IK-seqF9	CCTGTCGCTTCCTAGAATCC	8-2
IK-seqF9-2	CTTCCTTCCCTGGCTGGTTC	8-2
IK-seqF10N	AGAGTAATGCCACCAGATCC	8-3
IK-seqF10r	GGCGGAACAAGACACAGCGA	8-3
IK-seqR10	GGCAAGAAGAGGACACTACA	8-3
IK-seqF12	GGTACCGCTGTGTACTACTG	8-4
IK-seqF13	TTGGTGCGATGGCACTCACT	8-4
IK-seqF14	GGACAGCTGCAATACCACTT	8-4

c

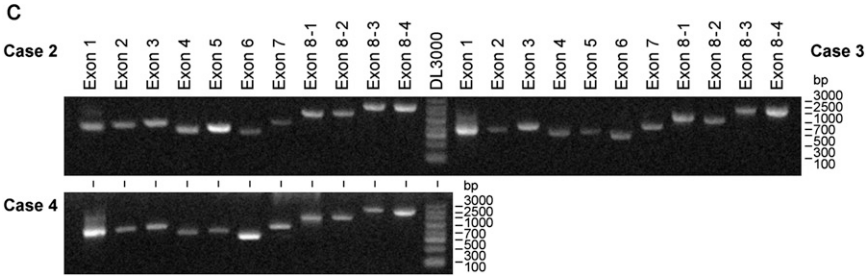


Fig. S6. Genomic PCR analysis of the human *IKFZ1*/*Ikaros* gene exons 1–8. (*A*) PCR primers. (*B*) Sequencing primers. (*C*) PCR product.

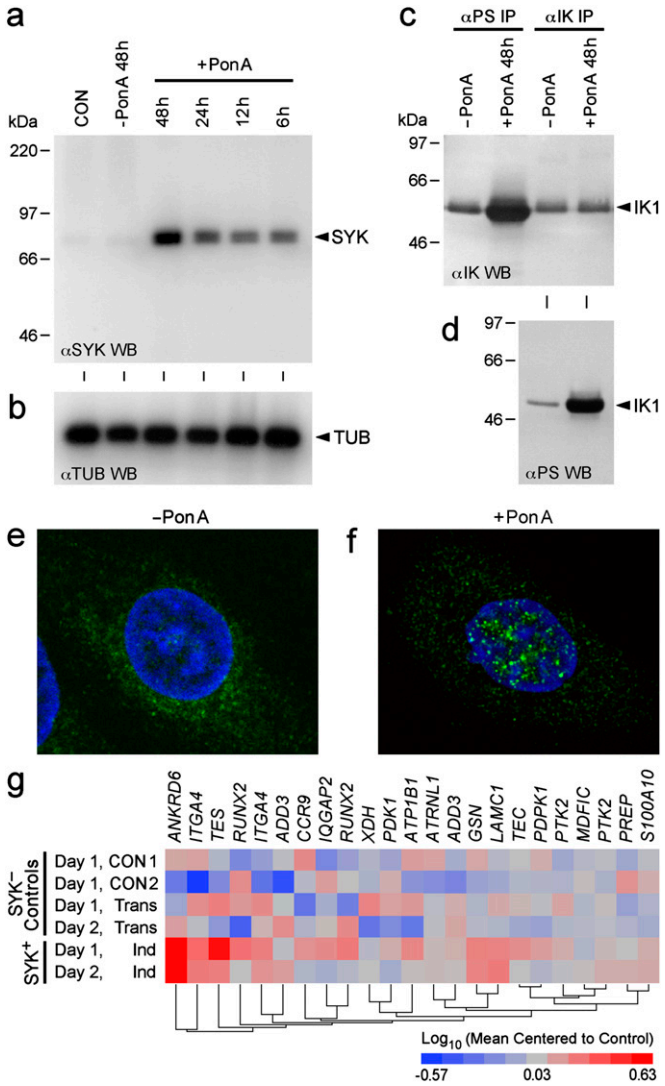
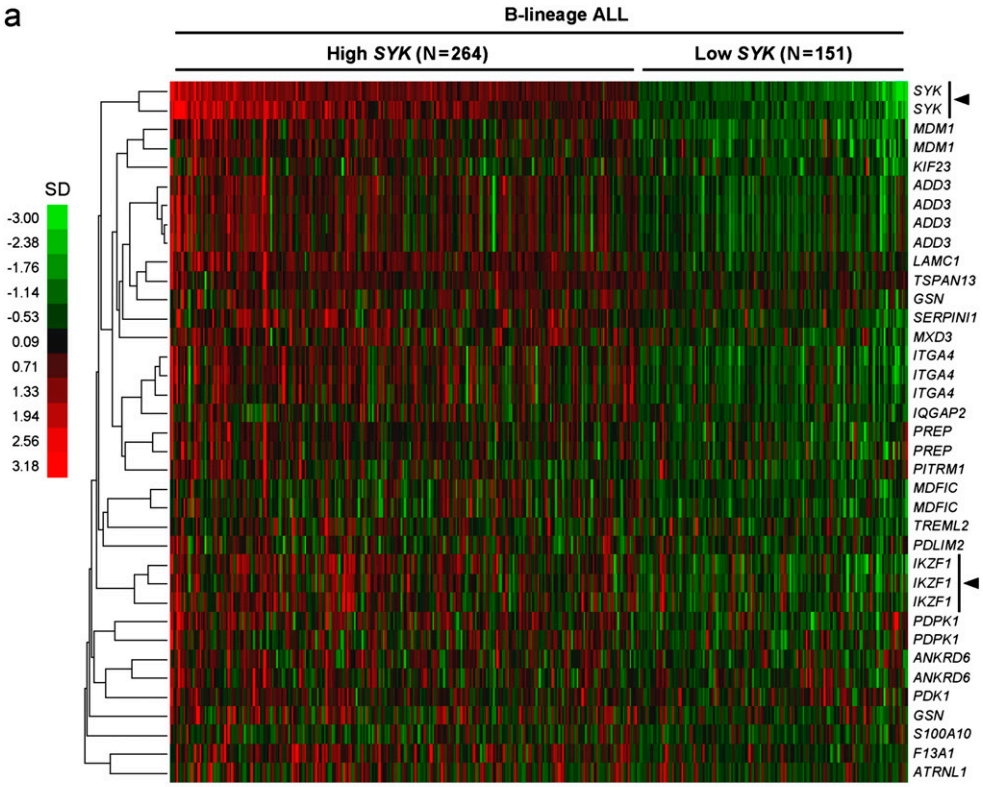


Fig. S7. Activation of IK in an ecdysone-inducible mammalian expression system for SYK. (*A*) Anti-SYK Western blot analysis of whole-cell lysates of untransfected vs. transfected U373 cells before and 6–48 h after exposure to the ecdysone-analog Pon-A (10 μM). (*B*) Anti-tubulin (TUB) Western blot analysis of the same lysates used in *A*. (*C*) Anti-IK Western blot analysis of antiphosphoserine and anti-IK immunoprecipitates from *SYK*-transfected U373 cells before and 48 h after Pon-A treatment for induction of *SYK* expression. (*D*) Antiphospho serine Western blot analysis of IK immunoprecipitated from U373 cells before and 48 h after Pon-A–mediated induction of *SYK* expression. (*E* and *F*) These two-color (green/blue) merge IK/TOTO3 confocal images illustrate the cytoplasmic localization of native IK in a representative *SYK*-transfected U373 cell before exposure to Pon-A (depicted in *E*) vs. the corresponding confocal image showing substantial nuclear (in addition to cytoplasmic) localization of native IK in a representative *SYK*-transfected U373 cell 24 h after exposure to Pon-A (10 μM) (depicted in *F*). U373 cells imaged before Pon-A exposure showed a predominantly cytoplasmic localization of native IK, whereas U373 cells imaged 48 h after Pon-A exposure showed a predominantly nuclear localization of native IK. (Magnification: 500×.) (*G*) Pon-A (10 μM)-induced expression of SYK leads to up-regulation of previously published and validated IK target genes. The cluster figure depicts $\log_{10}$ transformed expression values mean centered to the four control samples (red shows an increase in *SYK*-induced cells). Control (CON)1/CON2; no transfection, no induction; Trans, transfected but not induced; Ind, transfected and induced. Nineteen genes (23 transcripts) of 30 validated IK target genes (36 transcripts) represented on the Human Affymetrix U95Av2 platform were up-regulated with *SYK* induction in U373 cells, of which a cluster of three genes showed pronounced overexpression (*ANKRD6*: 6.5-fold increase, $P = 7.9 \times 10^{-6}$; *TES*: 2.8-fold increase, $P = 0.01$; and *ITGA4*: 2.25-fold increase, $P = 0.005$).

a



b

Affy ID	GeneSymbol	StdDev [High *SYK* (N=264) - Low *SYK* (N=151)]	T-test P-value
213761_at	*MDM1*	1.28	1.3×10^{-31}
201753_s_at	*ADD3*	0.93	8.3×10^{-27}
201034_at	*ADD3*	0.93	2.5×10^{-25}
213416_at	*ITGA4*	1.04	3.3×10^{-25}
205884_at	*ITGA4*	1.01	1.5×10^{-23}
201752_s_at	*ADD3*	0.92	2.3×10^{-23}
205882_x_at	*ADD3*	0.89	1.2×10^{-21}
200771_at	*LAMC1*	0.73	5.1×10^{-19}
205885_s_at	*ITGA4*	0.78	1.0×10^{-17}
220397_at	*MDM1*	0.96	1.6×10^{-17}
205352_at	*SERPINI1*	0.78	1.4×10^{-16}
221250_s_at	*MXD3*	0.71	1.8×10^{-12}
204709_s_at	*KIF23*	0.64	1.0×10^{-10}
204117_at	*PREP*	0.64	2.7×10^{-09}
204524_at	*PDPK1*	0.63	5.7×10^{-09}
217979_at	*TSPAN13*	0.38	1.5×10^{-08}
211675_s_at	*MDFIC*	0.48	1.6×10^{-08}
203474_at	*IQGAP2*	0.52	2.2×10^{-08}
203305_at	*F13A1*	0.57	5.5×10^{-08}
217599_s_at	*MDFIC*	0.52	1.1×10^{-07}
204672_s_at	*ANKRD6*	0.42	2.6×10^{-05}
200696_s_at	*GSN*	0.36	5.0×10^{-05}
37950_at	*PREP*	0.44	6.2×10^{-05}
205273_s_at	*PITRM1*	0.39	1.4×10^{-04}
214040_s_at	*GSN*	0.38	1.9×10^{-04}
221244_s_at	*PDPK1*	0.37	2.9×10^{-04}
213745_at	*ATRNL1*	0.31	2.0×10^{-03}
219165_at	*PDLIM2*	0.28	4.6×10^{-03}
219748_at	*TREML2*	0.28	6.3×10^{-03}
200872_at	*S100A10*	0.23	1.1×10^{-02}
204671_s_at	*ANKRD6*	0.20	4.2×10^{-02}
206686_at	*PDK1*	0.19	5.0×10^{-02}
205038_at	*IKZF1*	1.01	7.6×10^{-17}
205039_s_at	*IKZF1*	0.75	1.2×10^{-11}
216901_s_at	*IKZF1*	0.60	2.8×10^{-08}
207540_s_at	*SYK*	2.04	9.9×10^{-79}
209269_s_at	*SYK*	1.78	2.0×10^{-77}

Fig. S8. Effects of *SYK* expression levels on expression levels of IK target genes in primary lymphocyte precursors from B-lineage ALL patients. (*A* and *B*) Gene Pattern (www.broadinstitute.org/cancer/software/genepattern) was used to extract expression values for the IK target genes in the combined dataset from five studies with a total of 884 B-lineage ALL cases. The datasets were combined to test for consistent differences in the *z*-scores for high *SYK* (>0.5 SD units; $n = 264$ samples) versus low *SYK* (≤ 0.5 SD units; $n = 151$ samples) expression groups. Thirty-two transcripts representing 21 IK target genes were up-regulated in specimens with both high *IKZF1* and *SYK* expression. A one-way agglomerative hierarchical clustering technique was used to organize expression patterns using the average distance linkage method, such that genes (rows) having similar expression across patients were grouped together (average distance metric). The heat map depicts expression values represented by SD units above (red) and below the mean (green). Dendrograms were drawn to illustrate similar gene-

Legend continued on following page

expression profiles from joining pairs of closely related gene-expression profiles, whereby genes joined by short branch lengths showed most similarity in expression profile across patients. The five most significantly up-regulated genes in B-lineage ALL specimens with high *SYK* expression were *MDM1* (two transcripts ranging from 0.96 to 1.28 SD units, $P = 1.6 \times 10^{-17}$ to 1.28×10^{-31}), four transcripts for *ADD3* (0.89–0.93 SD units, $P = 1.2 \times 10^{-21}$ to 8.3×10^{-27}), three transcripts for *ITGA4* (0.78–1.04 SD units, $P = 1.0 \times 10^{-17}$ to 3.3×10^{-25}), *LAMC1* (0.73 SD units, $P = 5.1 \times 10^{-19}$), and *SERPINI1* (0.78 SD units, $P = 1.4 \times 10^{-16}$).

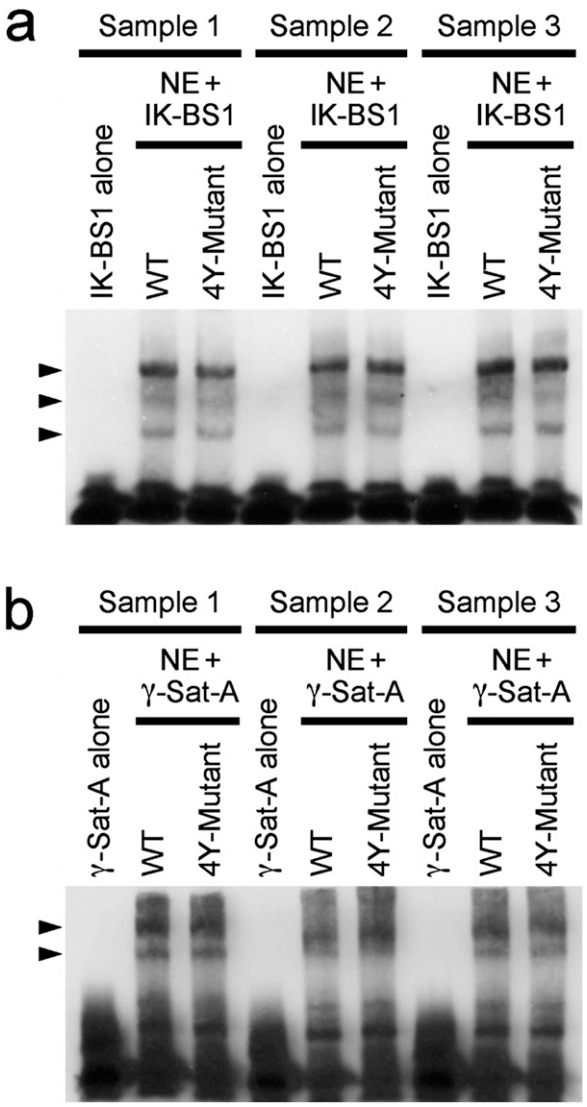


Fig. S9. DNA binding activity of the Y-mutant of IK protein. A mutant IK protein with phenylalanine substitutions of four tyrosine residues (Y292F+Y409F+Y493F+Y499F) was evaluated in triplicate for its DNA binding activity using EMSAs. EMSAs were performed on nuclear extracts (NE) from 293T cells transfected with expression vectors for wild-type or mutant IK protein using biotin-labeled DNA probes IK-BS1 (*A*) and γ-satellite A (*B*). IK activity was measured by the electrophoretic mobility shifts of the biotin-labeled γ-satellite A or IK-BS1 probe, representing IK-containing nuclear complexes (indicated with arrow heads).

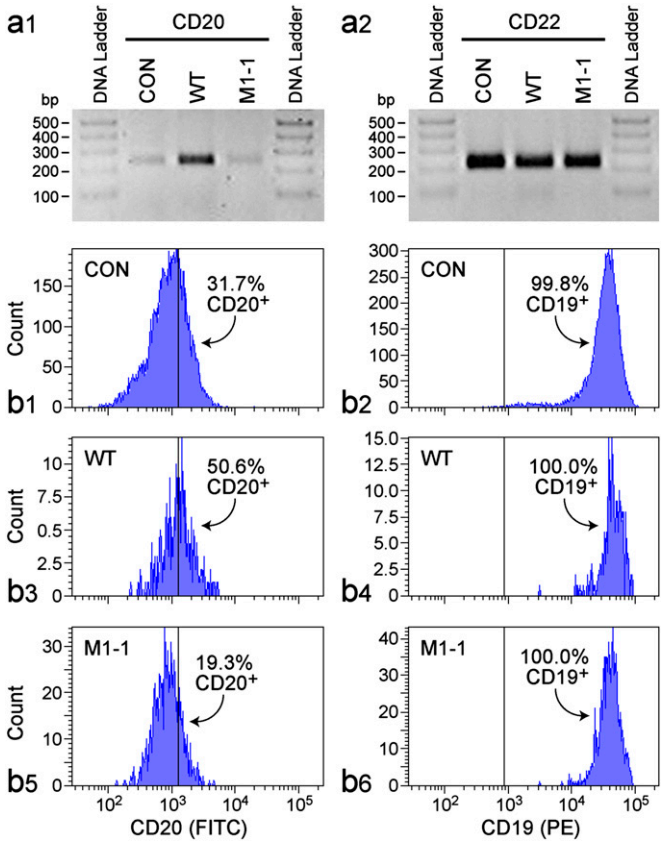


Fig. S10. Effects of wild-type and SYK-resistant mutant IK proteins on maturational stage of human B-cell precursors. In differentiation experiments, we examined the effects of wild-type and SYK-resistant mutant IK proteins on the maturation stage of the human pre-pre-B ALL cell line ALL-1 by using the gene and protein expression levels of the mature B-cell antigen CD20 as a differentiation marker. ALL-1 cells lack IK1/IK2 and express only truncated non-DNA binding IK isoforms (1). Cells were analyzed at 96 h after sham-transfection or transfection with plasmids for wild-type vs. SYK-resistant mutant IK. To determine the transfection efficiency, cells were electroporated in parallel with the EGFP reporter vector pmax×GFP, cultured for 48 h, and EGFP-expressing cells were identified using a Nikon Eclipse TS100 inverted microscope. The average transfection efficiency in triplicate samples was 47% (individual values: 15/30 = 50%; 20/42 = 48%; 11/25 = 44%). (*A*) A NanoDrop 1,000 Spectrophotometer (Thermo Scientific) was used to measure the total RNA concentration before RT-PCR. The amounts of total RNA used in the PCR assays were 118 ng for sham-transfected sample, 110 ng for the sample transfected with wild-type IK, and 110 ng for the sample transfected with mutant IK. RT-PCR analysis of showed markedly increased expression levels for *CD20* gene transfected with wild-type (but not mutant) IK. Transfections did not alter the transcript expression levels of the pan-B marker CD22. (*B*) The percentage of CD20 positivity was determined flow cytometrically (2, 3). Depicted are single-parameter FACS histograms obtained at 96 h showing CD20 (FITC) and CD19 (PE) expression levels of sham-transfected control ALL-1 cells and ALL-1 cells transfected with plasmids for wild-type vs. SYK-resistant murine IK proteins. Background fluorescence levels indicated with vertical lines were obtained with PE/FITC labeled control monoclonal antibodies. Only viable cells appearing in the lymphocyte/blast window during the light scatter analysis were included in the depicted fluorescence evaluations. Cells transfected with wild-type IK showed higher levels of CD20 antigen on their surface than sham-transfected cells or cells transfected with mutant IK.

1. Sun L, et al. (1999) Expression of aberrantly spliced oncogenic Ikaros isoforms in childhood acute lymphoblastic leukemia. *J Clin Oncol* 17(12):3753–3766.
2. Uckun FM, Goodman P, Ma H, Dibirdik I, Qazi S (2010) CD22 EXON 12 deletion as a pathogenic mechanism of human B-precursor leukemia. *Proc Natl Acad Sci USA* 107(39): 16852–16857.
3. Uckun FM, et al. (1993) Signal transduction through the CD19 receptor during discrete developmental stages of human B-cell ontogeny. *J Biol Chem* 268(28):21172–21184.

Table S1. Expression levels of IK target genes in knockout mice with *IK*-null mutation

Affy ID	Gene symbol	*Z*-score decrease in IK-null mutation	*t*-Test *P* value	Abundance rank
1423890_x_at	*Atp1b1*	−1.82	0.00002	3,004
1429950_at	*Unc5cl*	−1.82	0.00096	4,279
1437217_at	*Ankrd6*	−1.81	0.00098	5,474
1416762_at	*S100a10*	−1.81	0.00020	1,571
1415812_at	*Gsn*	−1.81	0.00018	3,592
1429274_at	*Lypd6b*	−1.81	0.00098	6,920
1453025_at	*Macrod2*	−1.80	0.00090	22,007
1422796_at	*Prep*	−1.80	0.00062	2,081
1456642_x_at	*S100a10*	−1.80	0.00048	2,043
1439348_at	*S100a10*	−1.79	0.00538	11,449
1423297_at	*Add3*	−1.79	0.00254	10,071
1418166_at	*Il12rb1*	−1.78	0.00172	12,275
1460378_a_at	*Tes*	−1.78	0.00142	9,956
1417977_at	*Eif4e3*	−1.78	0.00381	2,213
1432431_s_at	*Macrod2*	−1.78	0.01337	18,504
1448929_at	*F13a1*	−1.78	0.00547	1,637
1433885_at	*Iqgap2*	−1.78	0.00310	10,285
1424246_a_at	*Tes*	−1.77	0.01016	8,922
1428737_s_at	*Gramd3*	−1.77	0.00560	16,200
1451152_a_at	*Atp1b1*	−1.76	0.01459	3,533
1418453_a_at	*Atp1b1*	−1.76	0.00429	5,023
1417978_at	*Eif4e3*	−1.76	0.00829	1,944
1444667_at	*Brdt*	−1.75	0.01305	12,123
1429399_at	*Rnf125*	−1.75	0.02099	5,084
1421194_at	*Cerkl /// Itga4*	−1.75	0.00290	18,730
1439942_at	*Prep*	−1.74	0.00354	10,312
1428736_at	*Gramd3*	−1.74	0.01705	6,808
1450178_at	*Brdt*	−1.74	0.00855	13,784
1448390_a_at	*Dhrs3*	−1.74	0.00310	2,611
1440298_at	*Treml2*	−1.73	0.01475	7,560
1424704_at	*Runx2*	−1.73	0.00507	29,264
1423748_at	*Pdk1*	−1.73	0.00735	4,281
1423885_at	*Lamc1*	−1.73	0.02527	15,634
1437540_at	*Mcoln3*	−1.72	0.00693	9,941
1423298_at	*Add3*	−1.72	0.00558	11,992
1422970_at	*Mxd3*	−1.72	0.00513	10,248
1449831_at	*Tctex1d1*	−1.71	0.00666	4,783
1436045_at	*Tsga10*	−1.70	0.02148	12,060
1460204_at	*Tec*	−1.70	0.02428	3,001
1433795_at	*Tgfbr3*	−1.70	0.03494	4,356
1439036_a_at	*Atp1b1*	−1.69	0.02307	2,547
1416702_at	*Serpini1*	−1.69	0.01932	19,998
1417110_at	*Man1a*	−1.68	0.01045	7,555
1423946_at	*Pdlim2*	−1.68	0.03003	7,703
1436037_at	*Cerkl /// Itga4*	−1.67	0.01846	3,926
1423059_at	*Ptk2*	−1.67	0.02046	25,620
1451053_a_at	*Mdm1*	−1.67	0.01053	7,381
1427040_at	*Mdfic*	−1.67	0.01921	14,626
1421919_a_at	*Ccr9*	−1.66	0.01185	739
1450227_at	*Ankrd6*	−1.66	0.02413	11,481
1460239_at	*Tspan13*	−1.65	0.04370	1,803
1452948_at	*Tnfaip8l2*	−1.64	0.01945	11,863
1425814_a_at	*Calcrl*	−1.63	0.01718	6,786
1451006_at	*Xdh*	−1.62	0.01779	15,505
1419922_s_at	*Atrnl1*	−1.62	0.03996	11,138
1434306_at	*Rab3ip*	−1.62	0.02401	10,607
1433596_at	*Dnajc6*	−1.61	0.04376	5,053
1428595_at	*Slc6a19*	−1.60	0.02385	33,893
1435836_at	*Pdk1*	−1.56	0.03515	3,794
1453748_a_at	*Kif23*	−1.56	0.04814	20,110
1455442_at	*Slc6a19*	−1.52	0.04443	12,532

The publicly available archived GSE32311 database was used to compare gene expression changes in $CD4^+CD8^+$ double-positive wild-type (n = 3; GSM800500, GSM800501, GSM800502) vs. *Ikaros*-null thymocytes (n = 3, GSM800503, GSM800504, GSM800505) from the same genetic background of (C57BL/6 ×129S4/SvJae). Probe level "Robust Multi-array Average" (RMA) signal intensity values were obtained from the mouse 430_2.0 Genome Array. We identified 1,158 transcripts representing 924 genes that were down-regulated in *IK*-null mice with a subset of 201 transcripts representing 137 genes exhibiting a greater than twofold decrease. Cross-referencing of this IK-regulated gene set with CHiPseq data (GSM803110) identified 45 IK target genes that harbored IK binding sites. *z*-Scores were calculated by dividing the difference in the mean value of expression (null mutation – wild-type) by the SD of the pooled samples (n = 6). Abundance for IK direct targets was calculated and ranked according to mean level of expression in three control mice for 45,101 transcripts.

Table S2. IK target gene expression levels in human lymphocyte precursors according to *Ikaros* and *SYK* transcript levels

Affy ID	Gene symbol	SD [High SYK (n = 285) –Low SYK (n = 270)]	t-Test P value	SD [High IK (n = 302) – Low IK (n = 318)]	t-Test P value
200771_at	*LAMC1*	1.14	1.8×10^{-43}	0.82	3.0×10^{-25}
217979_at	*TSPAN13*	1.06	4.3×10^{-33}	0.56	3.1×10^{-13}
205352_at	*SERPINI1*	0.94	5.0×10^{-30}	0.71	2.9×10^{-18}
213761_at	*MDM1*	0.90	7.6×10^{-25}	0.84	1.7×10^{-26}
205884_at	*ITGA4*	0.86	2.2×10^{-24}	0.97	4.2×10^{-32}
203305_at	*F13A1*	0.84	6.6×10^{-21}	0.46	7.8×10^{-09}
213416_at	*ITGA4*	0.78	9.1×10^{-21}	0.84	1.8×10^{-26}
200696_s_at	*GSN*	0.81	2.3×10^{-20}	0.50	1.3×10^{-09}
201752_s_at	*ADD3*	0.82	3.0×10^{-20}	0.58	4.9×10^{-12}
201753_s_at	*ADD3*	0.81	8.5×10^{-20}	0.56	1.4×10^{-11}
201034_at	*ADD3*	0.81	1.4×10^{-19}	0.65	1.3×10^{-14}
220397_at	*MDM1*	0.77	3.6×10^{-19}	0.82	8.7×10^{-25}
205882_x_at	*ADD3*	0.79	3.8×10^{-19}	0.57	1.2×10^{-11}
205885_s_at	*ITGA4*	0.66	9.5×10^{-16}	0.86	6.9×10^{-26}
221250_s_at	*MXD3*	0.68	1.1×10^{-14}	0.23	5.7×10^{-03}
200770_s_at	*LAMC1*	0.66	6.6×10^{-14}	0.38	3.8×10^{-06}
204524_at	*PDPK1*	0.65	7.2×10^{-14}	0.43	5.4×10^{-07}
203474_at	*IQGAP2*	0.64	1.2×10^{-13}	0.75	8.5×10^{-21}
208820_at	*PTK2*	0.63	7.1×10^{-13}	0.35	5.0×10^{-06}
204117_at	*PREP*	0.58	1.5×10^{-11}	0.72	3.0×10^{-17}
207821_s_at	*PTK2*	0.57	1.3×10^{-10}	0.44	5.3×10^{-08}
214040_s_at	*GSN*	0.54	8.2×10^{-10}	0.41	9.1×10^{-07}
206686_at	*PDK1*	0.49	6.8×10^{-09}	0.17	2.4×10^{-02}
37950_at	*PREP*	0.41	6.4×10^{-06}	0.67	2.8×10^{-16}
221244_s_at	*PDPK1*	0.35	3.0×10^{-05}	0.42	5.8×10^{-07}
204672_s_at	*ANKRD6*	0.32	2.7×10^{-04}	0.35	2.4×10^{-05}
213745_at	*ATRNL1*	0.29	7.2×10^{-04}	0.31	3.0×10^{-04}
204671_s_at	*ANKRD6*	0.25	3.0×10^{-03}	0.18	2.4×10^{-02}
204709_s_at	*KIF23*	0.27	3.5×10^{-03}	0.36	7.4×10^{-06}
206887_at	*CCBP2*	0.26	4.1×10^{-03}	0.27	1.2×10^{-03}
219748_at	*TREML2*	0.23	6.0×10^{-03}	0.47	9.6×10^{-09}
205273_s_at	*PITRM1*	0.19	3.3×10^{-02}	0.52	3.8×10^{-10}
219165_at	*PDLIM2*	0.18	3.3×10^{-02}	0.31	9.6×10^{-05}
217599_s_at	*MDFIC*	0.18	4.7×10^{-02}	0.55	1.1×10^{-11}
205038_at	*IKZF1*	0.94	3.9×10^{-25}	2.20	3.3×10^{-186}
205039_s_at	*IKZF1*	0.69	3.6×10^{-15}	2.00	9.8×10^{-141}
216901_s_at	*IKZF1*	0.65	1.3×10^{-13}	1.70	3.8×10^{-95}
207540_s_at	*SYK*	2.19	3.0×10^{-129}	0.76	1.6×10^{-20}
209269_s_at	*SYK*	2.04	1.6×10^{-116}	0.89	6.8×10^{-26}

The Gene Pattern Web-based software was used to extract expression values from the NCBI GEO database to compile gene-expression profiles of human lymphocyte precursors in 1,104 primary leukemia specimens from ALL patients (GSE3912, n = 113; GSE18497, n = 82; GSE4698, n = 60; GSE7440, n = 99; GSE13159, n = 750). We focused our analysis on 45 previously published and validated IK target genes from Table S1. For each study, the SD values were calculated from the study mean for all of the patients. The datasets were combined to test for consistent differences in the z-scores for high *SYK* (>0.5 SD units; n = 285 samples) and low *SYK* (<0.5 SD units; n = 270 samples) groups. Similar calculation for *IKZF1* rank ordered expression levels three highly correlated transcripts for IK/*IKZF1* (205038_at, 205039_s_at, 216901_s_at, 227344_at and 227346_at) resulted in characterization of 302 ALL samples with high *IKZF1* expression and 318 samples with low *IKZF1* expression. t-Tests were performed using standardized expression values combined from five datasets (two-sample, unequal variance correction, P values < 0.05 deemed significant), revealing an intersect of 34 transcripts representing 22 genes that were significantly up-regulated in both high *SYK* and high *IK* expression groups. Hierarchical cluster analysis identified *LAMC1* (1.14 SD units, $P = 1.8 \times 10^{-43}$), *TSPAN13* (1.06 SD units, $P = 4.3 \times 10^{-33}$), *SERPINI1* (0.94 SD units, $P = 5.0 \times 10^{-30}$), *MDM1* (0.90 SD units, $P = 7.6 \times 10^{-25}$), and *ITGA4* (both transcripts: 0.86 SD units, $P = 2.2 \times 10^{-24}$; 0.78 SD units, $P = 9.1 \times 10^{-21}$), as the most significantly up-regulated five genes with the most significant effect sizes in 285 patient samples with high *SYK* expression.