METHODS

Patient samples. DNAs from leukemic T-ALL blasts at diagnosis and relapse and matched remission lymphocytes were provided by the Hemato-Oncology Laboratory at University of Padua, Italy; the Eastern Cooperative Oncology Tumor Bank Laboratory in New York, USA; and the Department of Pediatric Oncology/Hematology at the Charité-Universitätsmedizin Berlin in Berlin, Germany. We collected and analyzed samples under the supervision of local Columbia University Medical Center Instutional Review Board. We selected samples for Whole Exome

sequencing based on the availability of sufficient DNA from diagnosis, remission and relapse and we evaluated high tumor content at relapse based on the copy number analysis of T-cell receptor associated deletions.

Whole exome capture and nextgen sequence analysis. We used matched diagnostic remission and relapsed DNA samples from 5 T-ALL patients from University of Padua treated in AIEOP protocols (**Supplementary Table 1**) for exome capture with the SureSelect 50 Mb All Exon kit (Agilent Technologies) following standard protocols. We performed paired-end sequencing (2 × 100 bp) by using HiSeq2000 sequencing instruments at Centrillion Biosciences. Illumina HiSeq analysis produced between 60 and 120 million paired-end reads per sample. We mapped reads to the reference genome hg19 using the Burrows-Wheeler Aligner (BWA) alignment tool version 0.5.9. Mean depth (defined as mean number of reads covering the captured coding sequence of a haploid reference) was 50x with 80% of the genome covered more than 10x and 57% covered more than 30x. We identified sites that differ from reference (called here variants) in each sample independently. We constructed empirical priors for the distribution of variant frequencies for each sample. We obtained high-credibility intervals (posterior probability ≥ $1–10^{-5}$) for the corresponding change in frequency between tumor and normal samples, using the SAVI algorithm (Statistical Algorithm for Variant Identification) developed at Columbia University[32,33]. The number of germline SNPs in the coding region were 18,000 comparable with previous reports[32]. Most of the candidate germline SNPs (16,000, or ~90% of germline variants) were reported in dbSNP database. We identified candidate somatic variants using the following criteria: variant total depth in tumor and normal larger than 10x and smaller than 300x, variant frequency larger than 15% in tumor and less than 3% in normal, and at least 1% change in frequency from the normal with high posterior probability (≥ $1–10^{-5}$). Also to remove systematic errors, we excluded all variants that were found present in any of the normal cases. In addition, to eliminate ambiguous mapping from captured pseudogenes, and regions of low complexity, each variant with a flanking 20-base context sequence around its genomic position was mapped to the hg19 reference using the BLAST algorithm. We kept in the list only those with unique mappability, i.e. we required the 41-base sequence to uniquely map to the reference genome, with only one mismatch.

To discern the regions of Loss of Heterozygousity (LOH), we used the SAVI-calculated high-credibility intervals for the present dbSNP variants, which correspond to the change in their frequency between tumor and normal samples. In an LOH event, depending on whether the reference or the dbSNP allele is lost, at least 1% or at most –1% change in frequency from the

normal is expected. Therefore, by segmenting the regions covering more than 10 dbSNP variants with significantly changed frequencies, we were able to identify the LOH regions.

Mutation validation and analysis of recurrence. We designed primers flanking exons containing candidate somatic variants using Primer3 (http://frodo.wi.mit.edu/primer3/), and used for PCR amplification from Whole-genome-amplified (WGA) tumor, relapse and matched normal (remission) DNAs. We analyzed the resulting amplicons by direct bidirectional dideoxynucleotide sequencing with a validation rate of 97%. After exome sequence analysis of 5 diagnostic relapse and remission T-ALL AIEOP samples from University of Padua (**Supplementary Table 1**); we used 18 additional patient samples from the same institution for the analysis of recurrence of *TP53, BANP, RPL11, NRAS* and *NT5C2* (**Supplementary Table 1**). We subsequently extended this series to additional relapse T-ALL patient samples from University of Padua (n = 13), the Charité-Universitätsmedizin Berlin (n = 67) (**Supplementary Table 1**); and relapsed B-precursor ALL patients from University of Padua (n = 35) for extended mutation analysis of *NT5C2* (**Supplementary Table 1**). We used two cohorts of diagnostic T-ALL patients from ECOG (n = 23) and diagnostic B-precursor ALL patients from University of Padua (n = 27) to verify the absence of *NT5C2* mutations in diagnostic ALL specimens (**Supplementary Table 1**).

Structural depiction and analysis. we identified structural coverage of the NT5C2 protein through use of the PSI-Blast and SKAN algorithms; we subsequently mapped viable structures to all NT5C2 isoforms, and analyzed with the use of Chimera Suite[34,35]. We aligned structurally protein database (PDB) structures 2XCW, 2XCX, 2J2C, 2XCB, 2XCV, 2XJB, 2XJC, 2XJD, 2XJE, 2XJF, 2J2C, and 2JC9 and the resulting and subsequently analyzed the composite structure to assess conformational flexibilities[34]. We structurally modeled NT5C2 mutations utilizing the I-TASSER software suite, and subsequently refined and analyzed via minimization and rotamer library analysis in Chimera[34,36]. We predicted protein stability changes upon mutation through use of the SDM potential energy statistical algorithm and associated software[37]. We created all structural images using UCSF Chimera[34].

Site directed mutagenesis. We generated the *NT5C2* mutations at residues K359Q, R367Q and D407A by site directed mutagenesis on the mammalian expression pLOC-NT5C2 vector (Open Biosystems) using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions.

Cell lines. We cultured CCRF-CEM and CUTLL1 cells in RPMI-1640 media supplemented with 10% fetal bovine serum, 100 U ml^{-1} penicillin G and 100 µg ml^{-1} streptomycin at 37ºC in a humidified atmosphere under 5% CO2. We maintained HEK293T under similar conditions in DMEM media.

Lentiviral production and infection. We transfected the lentiviral constructs pLOC-NT5C2, pLOC-NT5C2-359, pLOC-NT5C2-367, pLOC-NT5C2-407 and the pLOC-RFP control plasmid with gag-pol and V-SVG expressing vectors into HEK293T cells using JetPEI transfection reagent (Polyplus). We collected viral supernatants after 48h and used them for infection of CCRF-CEM and CUTLL1 cells by spinoculation. After infection, we selected cells for 5 days in blasticidin and ficolled them the day before experiments.

Western blot. Western blot analysis was performed using a rabbit polyclonal antibody against NT5C2 (1:1,000; Abcam, ab96084); and a goat polyclonal antibody against GAPDH (1:7000; Santa Cruz Biotechnology, sc-20357) using standard procedures.

Cell viability and chemotherapy drug response. We determined cell viability by measurement of the metabolic reduction of the tetrazolium salt MTT using the Cell Proliferation Kit I (Roche) following the manufacturer instructions. We performed experiments in triplicate. We analyzed viability after 48 or 72 hours of initiation of treatment with 6-mercaptopurine, 6-thioguanine, nelarabine and Ara-G.

Recombinant protein production and purification. We cloned full length cDNA constructs encoding wild type NT5C2, NT5C2 K359Q, NT5C2 R367Q and NT5C2 D407A with an N-terminal His6 tag in the pET28a-LIC expression vector using the In-Fusion HD PCR cloning system (Clontech), as per manufacturer's instructions. We expressed recombinant proteins from Rosetta 2(DE3) *E. coli* cells by induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside for 3 hours at 37°C. We harvested cells and lysed them in lysis buffer (50 mM sodium phosphate pH 7.4, 100 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 1% Triton X-100, 0.5 mg ml^{-1} lysozyme, 20 mM imidazole) supplemented with Complete EDTA-free protease inhibitor (Roche). We purified His-tagged NT5C2 proteins binding them to Nickel-sepharose beads and eluting them with 50 mM sodium phosphate pH 7.4, 100 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 300 mM imidazole. We removed imidazole by buffer exchange using PD-10 desalting columns (GE Healthcare). We assessed protein expression and purity by SDS-PAGE and Coomassie staining.

5'-Nucleotidase assay. We assessed 5'-Nucleotidase activity of purified recombinant wild type and mutant NT5C2 proteins using the 5'-nucleotidase (5'-NT) Enzymatic Test Kit (Diazyme), according to the manufacturer's instructions. The assay measures the enzymatic hydrolysis of inosine 5'-monophosphate to inosine, which is reacted further to hypoxanthine by purine nucleoside phosphorylase, and then to uric acid and hydrogen peroxide by xanthine oxidase. H_2O_2 is quantified using a Trinder reaction. We calculated 5'-Nucleotidase activity levels using a calibrator of known 5'-NT activity as standard. We performed assays in triplicate in an Infinite M200 Tecan plate reader.

Statistical analysis. We evaluated differences in the percentages of NT5C2 wild type and mutant ALL patients in different relapsed categories using the Fisher's exact test. We analyzed equality of categorical and continuous variables by Fisher's exact test and by Mann-Whitney U test, respectively.

Methods References

32. Tiacci, E., et al. BRAF mutations in hairy-cell leukemia. N Engl J Med 364, 2305-2315.
33. Pasqualucci, L., et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 43, 830-837.
34. Pettersen, E.F., et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612 (2004).
35. Altschul, S.F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402 (1997).
36. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5, 725-738.
37. Worth, C.L., Preissner, R. & Blundell, T.L. SDM--a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39, W215-222.