SUPPLEMENTARY INFORMATION

Contents:

SUPPLEMENTAL MATERIALS AND METHODS

Human samples

RNA-seq data relative to 272 patients diagnosed with T-cell lymphoblastic leukemia/ lymphoblastic lymphoma, including the index case (Patient_Rx), were obtained through the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative as well as different biobanks belonging to the Spanish Network of Biobanks in Hospitals (RetBioH; www.redbiobancos.es): Hospital Universitario 12 de Octubre i+12, Hospital Universitario Ramón y Cajal, Hospital General Universitario Gregorio Marañón, Hospital La Paz, Hospital de Murcia and Fundación Jiménez Díaz. RNA-seq data from TARGET, accession number phs000218 and sub-study specific accession number phs000464.v19.p8 (TARGET Acute Lymphoblastic Leukemia (ALL) Expansion Phase 2), are managed by the National Cancer Institute (NCI). Information about TARGET can be found at: http://ocg.cancer.gov/programs/target. Data used for this analysis are accessible in the dbGaP database (database of Genotypes and Phenotypes, https://www.ncbi.nlm.nih.gov/projects/gap/cgibin/study.cgi?study_id=phs000464.v19.p8).

For differential expression analysis of Patient_Rx, it was compared with CD34+ cells isolated from human postnatal thymocytes from thymuses removed during cardiac surgery in pediatric individuals, as described previously[1].

All samples were collected after informed consent, according to the declaration of Helsinki, following legal standards for clinical studies in Spain and regulations of the ethics committees from the respective hospitals. The work performed was approved by the Research Ethics Committee of the Universidad Autónoma de Madrid (references CEI-31-773 and CEI-70-1260).

Case of interest

A 51-year-old patient presented symptoms of edema and pain in the left arm. Positron emission tomography (PET/CT) revealed a hypermetabolic mass in the anterior mediastinum with evident lymph node involvement at the supradiaphragmatic level and probably incipient at the infradiaphragmatic level. The presence of blasts was prominent in the peripheral blood and a bone marrow biopsy was therefore performed. Bone marrow analysis showed hypercellularity and clear predominance of a monomorphic cell population with medium-large size, high nucleus/cytoplasm ratio, loose chromatin and evident nucleolus in the nuclear periphery. The presence of frequent mitoses stood out. Immunophenotypic analysis revealed an abnormal population of cells compatible with T-lymphoblasts at the pro-T stage (CD34+, CD5+, CD7++, CD3+ic (50%), CD3s-, CD10+, TdT+, CD45+d, CD10+++, HLA-DR++, CD2-, CD4-, CD8-, TCDab-, TCRgd-, CD79a+, CD19-, CD22ic-, CD13+,MPO-, CD15-, CD11b-, CD117-, CD16-, CD56-, CD36-, CD14-, CD3ic-). The patient was diagnosed with pro-T lymphoblastic leukaemia and achieved complete remission (CR) after being treated according to LAL/SEHOP-PETHEMA 2011 and also receiving a haematopoietic stem cell transplant (HSCT) from his HLA-identical brother. Reassessment at 1, 3, 6 and 12 months after HSCT showed MRD <0.01% and complete chimerism in bone marrow. However, 18 months after HSCT, the MRD increased to 0.26% and a symptomatic relapse of T-ALL was subsequently confirmed.

Cell culture

M07e (DSMZ#ACC104) and Ba/F3 (DSMZ#ACC300) cell lines were purchased from the Leibniz Institute-DSMZ. The HEK293T (ATTC# CRL-11268) cell line was purchased from ATCC. The U4A cell line was generously provided by Dr. Elise Holvey-Bates (Cleveland Clinic). Cell experimentation was always performed within a period not exceeding 6 months after thawing. Cultures tested negative for mycoplasma and were maintained at 37°C in 5% CO_2 humidified atmosphere. Adherent cell lines were grown in Dulbecco's modified Eagle's medium (DMEM) (GE Healthcare Life Sciences) supplemented with 10% fetal bovine serum (GE Healthcare Life Sciences), 2 mM L-glutamine (Merck Millipore) and 1 mM sodium pyruvate (Merck Millipore). Suspension cell lines were grown in RPMI 1640 medium (Gibco, Life Technologies) supplemented with 10% fetal bovine serum and 2 mM L-glutamine. M07e and Ba/F3 cells, which require growth factors/cytokines for normal viability and proliferation, were routinely cultured with 10 ng/ml granulocyte-monocyte colony-stimulating factor (R&DSystems) or with 5 ng/ml of interleukin-3 (Cell Signaling) respectively.

Pharmacological inhibitors

Ruxolitinib and Tofacitinib were purchased from Selleckchem and used within the dose range recommended by manufacturers: the JAK1 inhibitor ruxolitinib (#INCB018424, 1 µM, 0.5 µM and 0.1 µM); and the JAK3 inhibitor tofacitinib (#CP-69055 1 µM, 0.5 µM and 0.1 µM). In functional assays involving pharmacological inhibitors, the so-called "untreated cells" were cultured with dimethyl sulfoxide (DMSO) as a negative control.

Expression vectors

The lentiviral vector expressing $JAK3^{WT}$ was purchased from Vectorbuilder. $JAK3^{M511I}$, $JAK3^{L857P}$ and $JAK3^{Q988P}$ mutations were introduced using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies). The oligonucleotides used are summarized in Supplemental Table S1. The different plasmids were amplified in DH5α competent bacteria and subsequently purified using Wizard® Plus SV Minipreps DNA purification system (Promega) or GenoPure Plasmid Maxi Kit (Roche Applied Science). Cell transfection was accomplished using Lipofectamine 2000 (Invitrogen) and Opti-MEM medium (Gibco). Cell transduction was accomplished with lentiviral particles obtained after HEK293T-mediated packaging using pMD2.G and psPAX2 (Addgene; RRID:Addgene_12260). Transduced cells were sorted using FACS (FACSCVantage SE, BD Biosciences, RRID:SCR_013311) for similar EGFP levels.

Functional assays

To analyze growth and viability, cells were washed three times with PBS (1X), seeded at 500.000 cells/ml and cultured during 96 hours in medium depleted of cytokines and growth factors. Cells were counted using trypan blue exclusion and the TC10 Automated Cell Counter (Bio-Rad Laboratories).

DNA extraction

Total DNA was isolated using the DNeasy 96 Blood and Tissue kit (Qiagen) according to the manufacturer's instructions. DNA quantification and quality were checked with Nanodrop (Thermo Fisher Scientific Inc.), Qubit (Thermo Fisher Scientific Inc.) and TapeStation (Agilent Technologies).

Whole exome sequencing (WES)

WES was performed by Sistemas Genómicos S.L. using Illumina HiSeq sequencing platform (Illumina, Inc.). Paired end reads 101nt length were generated; then, targeted regions were enriched using targeted sequencing protocol; reads were aligned against the human reference genome version GRCh38/hg38; read alignment was performed using BWA and 'in-house' scripts. Variant calling was performed using a combination of VarScan[2] and GATK[3] algorithms. 'In-house' scripts were developed to combine and filter variants. Identified variants were annotated using the Ensembl database[4].

Polymerase chain reaction (PCR)

Amplifications of genomic regions encompassing the JAK3^{M511I} and JAK3^{Q988P} mutations were performed by PCR using FastStart High Fidelity PCR System (Roche Applied Science) and appropriate primer oligonucleotides described in Supplemental Table S1, following manufacturer´s instructions. For each reaction, 300 ng of template genomic DNA were used. Annealing temperature was set at 55ºC for JAK3^{M511I} and JAK3^{Q988P} amplicons. Elongation time was set at 30 seconds and a final elongation step of 7 minutes was carried out. Amplification products were purified using Wizard® SV Gel and PCR Clean-Up System (Promega) following manufacturer´s instructions.

RNA extraction

Total RNA was isolated using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA quantification and quality were checked by Nanodrop (Thermo Fisher Scientific Inc.), Qubit (Thermo Fisher Scientific Inc) and TapeStation (Agilent Technologies).

RNA sequencing (RNA-seq)

When perfomed by Sistemas Genómicos S.L., a quality study was carried out using the FASTQC programme. For bioinformatics analysis, GRCh38/hg38 genome was used as a reference. The process was carried out using the algorithm proposed by Tophat2 v2.1.0[5]. Subsequently, poor quality mapping reads were removed using Picard Tools (http://picard.sourceforge.net). The assembly and identification were performed by Bayesian inference using the algorithm proposed by cufflinks v2.2.1[6]. Gene quantification was carried out using the htseq_count 0.6.1p1 method[7].

When perfomed by NIMGenetics S.L., the TruSeq Stranded Total RNA Library Prep (Illumina, Inc.) was used, which included rRNA depletion, fragmentation, cDNA synthesis and adaptor ligation. The generated libraries were normalized and combined in equimolecular concentrations for optimal generation of DNA clusters. Paired-end sequencing (2x100bp) of the previously enriched, indexed and multiplexed libraries were performed on the high-throughput NovaSeq 6000 platform (Illumina Inc.), with a minimum of 100M PE reads (50+50) per sample, with a read quality of 85%>Q30. For bioinformatics analysis, GRCh38/hg38 (Ensembl version 103) genome was used as a reference. Briefly, quality check and sequence trimming were performed using FASTQC and fastp[8] respectively. Then, the trimmed RNA-seq reads were aligned against the reference genome and the transcripts were assembled using HISAT2 tool[9]. Corresponding genes were obtained and their expression abundance was determined using StringTie suite

(https://ccb.jhu.edu/software/stringtie/). On gene counting matrices, reads were subjected to unsupervised filtering in order to discard those genes with very few or no reads throughout all the samples of the study (https://bioconductor.org/packages/release/bioc/html/genefilter.html). Genes with a total abundance below 15 reads were excluded from further analysis.

Differential expression analysis

The statistical package DESeq2[10] was used for differential expression analysis using a negative binomial distribution for the determination of statistical significance. For multiple testing, p-value was adjusted using the Benjamini-Hochberg procedure[11].

Amplification and sequencing of the cDNA region encompassing both the JAK3^{M511I} and JAK3^{Q988P} mutations

One microgram of RNA was reverse-transcribed using SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen, Carlsbad, CA, USA), and a cDNA region encompassing both the JAK3^{M511I} and JAK3^{Q988P} mutations was amplified by PCR using Expand High Fidelity PCR System (Roche Applied Science) and appropriate primer oligonucleotides described in Supplemental Table S1, following manufacturer´s instructions. For each reaction, 200 ng of template cDNA were used. Annealing temperature was set at 62ºC. Ten cycles with an elongation time of 60 seconds were followed by 20 cycles with an elongation time of 60 seconds plus 5 additional seconds per cycle. The amplification product (1518 bp) was purified using Wizard® SV Gel and PCR Clean-Up System (Promega) following manufacturer´s instructions and cloned into pGEM-T-Easy vector system (Promega). Sanger sequencing of six individual clones was performed using a forward and a reverse oligonucleotide for the two positions of interest.

Sanger sequencing

Expression vectors, PCR products and clones, together with the corresponding oligonucleotides, were sent to Macrogen for Sanger sequencing. The resulting sequences were subsequently analyzed with Chromas Pro (Technelysium Pty Ltd). The oligonucleotides used are summarized in Supplemental Table S1.

Western-Blot (WB)

Protein extraction was performed with RIPA lysis buffer (50mM Tris-HCl pH 7.4, 150mM NaCl,1% triton X-100, 0.5% Deoxycholate and 0.1% SDS). Protein extracts were supplemented with 2mM phenylmethylsulphonyl fluoride, 2.5µl/ml Protease Inhibitor Cocktail and 10µl/ml Phosphatase Inhibitor Cocktail 2 (Roche Diagnostics GmbH). Ten-microgram aliquots of total protein were electrophoresed in 30% acrylamide/bis-acrylamide solution 29:1 (Bio-Rad Laboratories, RRID:SCR_008426) and then electro-transferred to mini-sized nitrocellulose membranes using the Transfer Blot® Turbo™ Transfer System (Bio-Rad Laboratories). Nitrocellulose membranes were incubated with primary and, subsequently, secondary antibodies. Secondary antibodies were conjugated to horseradish peroxidase (GE Healthcare Life Sciences), and the bands were visualized using a cooled charge-coupled device camera (ImageQuant LAS-4000; GE Healthcare Life Sciences). The antibodies used are summarized in Supplemental Table S2.

Statistics

Statistical analyses were performed with GraphPad Prism 8 and R-version 4.1.0. The Shapiro-Wilk and Kolmogorov-Smirnov tests were used to check the normality of data groups, and the Levene test was used to check the homogeneity of variances. The differences between independent samples were analyzed using the nonparametric Mann-Whitney and Kolmogorov-Smirnov tests for variables not adjusted to normality. Multiple comparisons were performed by one-way or two-way analysis of variance (ANOVA) or Krustal-Wallis analysis followed by Dunnett's test. All p-values below 0.05 were considered statistically significant.

Data Sharing Statement

Data from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative are managed by the National Cancer Institute (NCI) and accessible through the genotypes and phenotypes database (dbGaP, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study). Additional data that support the findings of this study are available in the supplementary material of this article and/or from the corresponding authors upon reasonable request.

SUPPLEMENTAL TABLE AND FIGURE LEGENDS

Supplemental Table S1. List of oligonucleotides, indicating its application, target gene, identification and sequence.

Supplemental Table S2. List of antibodies, indicating the commercial name, dilution, species, company and catalog number.

Supplemental Figure S1. A. Mutations identified in the tumor sample at relapse. Comprehensive summary of those mutations affecting the canonical isoform of proto-oncogenes and tumor suppressor genes that are considered relevant for T-ALL development. B.Schematic representation of the approach followed to study the configuration of $JAK3^{Q988P}$ and $JAK3^{M511I}$ mutations in the index patient. Representative electropherograms are shown. C. Electropherograms confirming the presence of the single-point mutation 1533G>A *JAK3* at diagnosis.

Supplemental Figure S2. A. mRNA expression of genes belonging to JAK and STAT families in the tumor sample at relapse and control thymocytes. Expression levels are referred to the mean of both subgroups. *: $p<0.05$; **: $p<0.01$. B. mRNA levels for multiple transcriptional targets of activated STAT proteins in the tumor sample at relapse and control thymocytes. Expression levels are referred to the mean of both subgroups. *: $p<0.05$; **: $p<0.01$. C. *In-silico* analysis of the $JAK3^{Q988P}$ mutation using three different predictors (Sift, Provean and Polyphen). All of them catalogued $JAK3^{Q988P}$ as a likely-oncogenic mutation. D. Cell growth (left) and viability (right) assays of M07e cells untransduced (0) or transduced with $JAK3^{WT}$ or $JAK3^{Q988P}$. E. Western blot for JAK3, p-STAT5 and STAT5 in M07e cells untransduced (-) or transduced with $JAK3^{WT}$ or $JAK3^{Q988P}$. F. Western blot for JAK3, p-STAT1, STAT1, c-MYC and BCL-XL in Ba/F3 cells transduced with JAK3 mutants untreated or treated with ruxolitinib (0.5µM or 1µM). G. Western blot for JAK3, p-STAT1, STAT1, c-MYC and BCL-XL in Ba/F3 cells transduced with JAK3 mutants untreated or treated with tofacitinib (0.5µM or 1µM). The graphics show the mean ± standard deviation (s.d.) after three independent experiments. *: $p<0.05$; **: $p<0.01$. Images are representative examples of at least three independent experiments.

SUPPLEMENTAL TABLES AND FIGURES

Supplemental Table S1.

Gene	Application	Oligonucleotide	Sequence (5´ > 3´)
JAK3	Sequencing	JAK3 M511I_Fw	CTCCCTCATCCTCTCCCCATAG
		JAK3 M511I_Rv	AGATGGGAAAACCGAGGCAA
		JAK3 Q988P_Fw	GGGTTATAGTTGGGGTCTGGGTTG
		pGEM-T-easy_T7	TAATACGACTCACTATAGGG
		pGEM-T-easy_SP6	ATTTAGGTGACACTATAGAA
		JAK3 cDNA_Fw	CCTCTCAGACCCCACACCT
		JAK3 cDNA_Rv	CTTCCGGGCAGAGACATTG
	Mutagenesis	JAK3 Q988P_Rv	AGTGCTGTTGAGGGGGCG
		JAK3 1533G>C_Fw	GGATCTTGTGAAATGTGATCTGACTCAGCTGGTATTGG
		JAK3 1533G>C_Rv	CCAATACCAGCTGAGTCAGATCACATTTCACAAGATCC
		JAK3 2570T>C_Fw	CCCGCTGTGCTGCGGCTGTTTCACGGC
		JAK3 2570T>C_Rv	GCCGTGAAACAGCCGCAGCACAGCGGG
		JAK3 2963A>C_Fw	GAAAATGGGGCTCGGGCCTGGCTCGCG
		JAK3 2963A>C_Rv	CGCGAGCCAGGCCCGAGCCCCATTTTC
	Amplification	JAK3 CIS/TRANS_Fw	CAGCCCCAATCCCAATACCAGC
		JAK3 CIS/TRANS_Rv	GCGAGAGAAGATGTTGTCCGAGAGG

Supplemental Table S2.

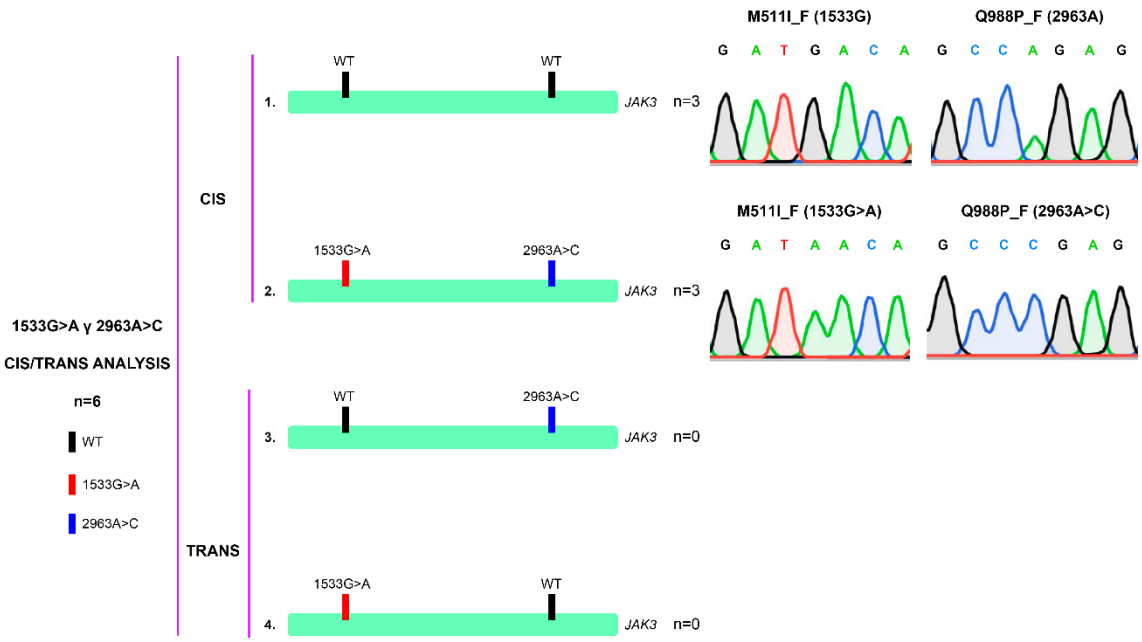
Antibody	Dilution	Species	Company, catalog No
JAK3 (D7B12)	1/1000	Rabbit	Cell Signaling, 8863
STAT1	1/1000	Rabbit	Cell Signaling, 9172
P-STAT1 (Tyr701) (D4A7)	1/1000	Rabbit	Cell Signaling, 7649
STAT5 (D2O6Y)	1/1000	Rabbit	Cell Signaling, 94205
c-MYC (D84C12)	1/1000	Rabbit	Cell Signaling, 5605
Bcl-xL	1/1000	Rabbit	Cell Signaling, 2762
P-STAT5 (Tyr694)	1/1000	Mouse	BD Biosciences, 611965
β-Actin (AC-15)	1/20000	Mouse	Sigma Aldrich, A5441
HRP-anti-mouse igG antibody	1/1000	-	Cell Signaling, 7076
HRP-anti-rabbit igG antibody	1/1000	-	Cell Signaling, 7074

Supplemental Figure S1.

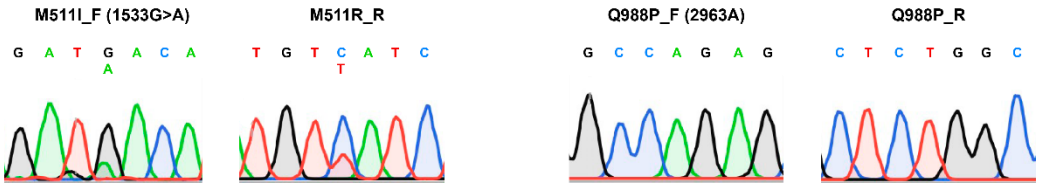
A

Gene	Chromosome	Exon	Var_effect	Ref_Allele	Var_Allele	Codons	Var_Depth
NOTCH1	chr9	27/34	missense	T	A	gTg/gAg	0.466
JAK3	chr19	11/24	missense	G	A	atG/atA	0.475
JAK3	chr19	21/24	missense	A	C	cAg/cCg	0.5
WT1	chr11	7/10	frameshift	G	TG	cga/cTga	0.778
KMT2C	chr7	36/59	frameshift	C	AC	aac/aaAc	0.449
KMT2C	chr7	14/59	missense	G	A	Ggt/Agt	0.485
KMT2C	chr7	14/59	missense	C	T	tCa/tTa	0.469
KMT2C	chr7	7/59	missense	C	T	Ctt/Ttt	0.139

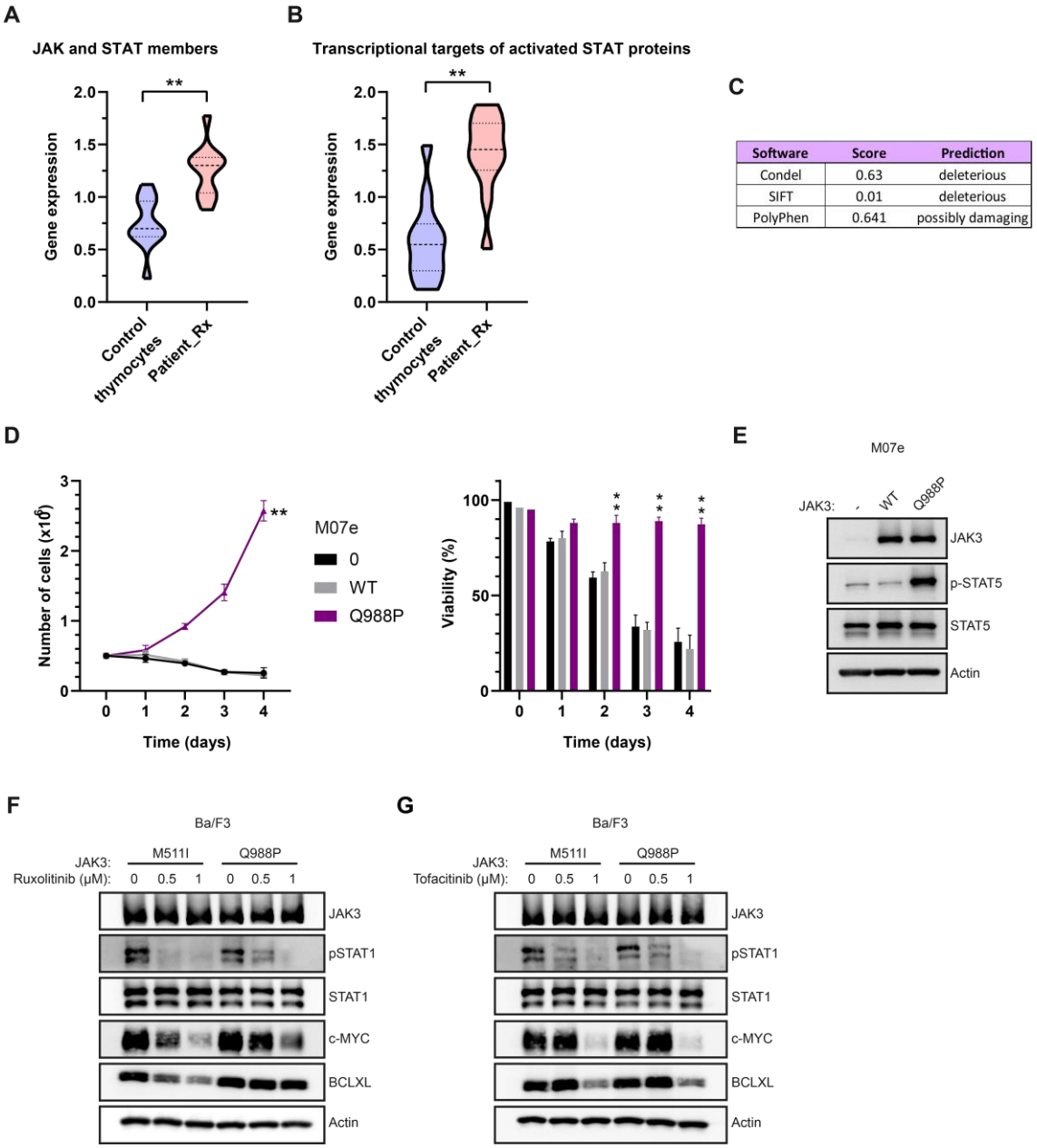
B



C



Supplemental Figure S2



SUPPLEMENTAL REFERENCES

1. Lopez-Nieva P, Fernandez-Navarro P, Cobos-Fernandez MA, et al. Patterns of Differentially Expressed circRNAs in Human Thymocytes. *Noncoding RNA*. Mar 30 2022;8(2)doi:10.3390/ncrna8020026
2. Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. Sep 1 2009;25(17):2283-5. doi:10.1093/bioinformatics/btp373
3. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 10 1-11 10 33. doi:10.1002/0471250953.bi1110s43
4. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Res*. Jan 7 2022;50(D1):D988-D995. doi:10.1093/nar/gkab1049
5. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. May 1 2009;25(9):1105-11. doi:10.1093/bioinformatics/btp120
6. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. May 2010;28(5):511-5. doi:10.1038/nbt.1621
7. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. Jan 15 2015;31(2):166-9. doi:10.1093/bioinformatics/btu638
8. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. Sep 1 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560
9. Zhang Y, Park C, Bennett C, Thornton M, Kim D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res*. Jun 8 2021;doi:10.1101/gr.275193.120
10. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
11. Benjamini Y, Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57(1):12. doi:https://www.jstor.org/stable/2346101