SUPPLEMENTARY APPENDIX

Oncorequisite role of an aldehyde dehydrogenase in the pathogenesis of T-cell acute lymphoblastic leukemia

Chujing Zhang,[1] Stella Amanda,[1] Cheng Wang,[2] Tze King Tan,[1] Muhammad Zulfaqar Ali,[1] Wei Zhong Leong,[1] Ley Moy Ng,[1] Shojiro Kitajima,[3] Zhenhua Li,[4] Allen Eng Juh Yeoh,[1,4] Shi Hao Tan[1] and Takaomi Sanda[1,5]

[1]Cancer Science Institute of Singapore, National University of Singapore, Singapore; [2]Department of Anatomy, National University of Singapore, Singapore; [3]Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan; [4]VIVA-NUS CenTRAL, Department of Pediatrics, National University of Singapore, Singapore and [5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

 doi:10.3324/haematol.2019.245639

Received: December 18, 2019.

Accepted: May 14, 2020.

Pre-published: May 15, 2020.

Correspondence: *TAKAOMI SANDA* - takaomi_sanda@nus.edu.sg

Supplementary Information

Supplementary Materials and Methods

Cell culture and reagents

All human leukemia cell lines were cultured in RPMI-1640 medium (BioWest) supplemented with 10% FBS (BioWest). 293T cells were maintained in DMEM medium (BioWest) supplemented with 10% FBS and penicillin/streptomycin (Life Technologies). All-*trans* retinaldehyde was purchased from Sigma-Aldrich. Glucose free and Glutamine free RPMI-1640 media were purchased from ThermoFisher.

Patient derived xenograft (PDX) sample

T-ALL patient-derived PDX sample (DFCI15) was kindly provided by Alejandro Gutierrez (Boston Children's Hospital, Boston). Mouse studies were conducted according to the recommendations from the Institutional Animal Care and Use Committee (IACUC) and all protocols were approved by the Committee at the National University of Singapore (NUS). PDX samples were first transplanted in 8-week-old NSG mice through tail-vein injection. Human leukemic cell expansion was monitored through human CD45 staining on blood samples. Single cells were isolated from the bone marrow, which at time of sacrifice contained >85% human CD45+ cells. PDX samples were cultured in reconstituted alpha-minimum essential media supplemented with 10% fetal bovine serum, 10% human AB serum (Invitrogen), 1% L-glutamine, 1% penicillin/streptomycin in the presence of recombinant cytokines stem cell factor (50 ng/mL), Flt3-L (20 ng/mL) and IL-7 (10 ng/mL) at 37° under 5% CO2.

Inhibitor treatment

T-ALL cell lines Jurkat, RPMI-8402, KOPT-K1 and DND-41 were seeded into 96-well plates at 3,000 cells per well, and treated with different concentration of WIN 18,446 (Cayman). Inhibitor-containing medium was changed every three days. Cell viability was measured using CellTiter Glo(Promega) at day 7. Luminescence background was extracted from the readings. The half maximal inhibitory concentration (IC_{50}) was calculated using GraphPad Prism 8.3.

Knockdown experiments

The short-hairpin RNA (shRNA) sequences were designed according to The RNAi Consortium (TRA)'s recommendation (http://www.broadinstitute.org/rnai/trc) and cloned into the pLKO.1-puro lentivirus vector. Lentiviruses were produced by co-transfecting individual shRNA construct with the packaging plasmids pMDLg/pRRE and pRSV-Rev and the envelope plasmid pMD2.G into 293T cells using FuGENE6 (Roche). Supernatants containing lentivirus particles were collected and filtered through a 0.45-μm filter (Thermo). T-ALL cells were infected with lentivirus in the presence of 8 μg/mL polybrene (Millipore) by centrifugation at 1300 rcf for 1.5 hrs at 32 degrees. Successfully-infected cells were selected with 0.7 μg/mL puromycin (Sigma). The shRNA sequences are described below:

Target Genes	**shRNA IDs**	**shRNA sequences**
GFP (control)	shGFP	ACA ACA GCC ACA ACG TCT ATA
ALDH1A2	shALDH1A2 #3	GCA ACC ATG GAA TCC CTA AAT
ALDH1A2	shALDH1A2 #5	TCC TGT AGA TGG AGA CTA TTT
TAL1	shTAL1	GCT CAG CAA GAA TGA GAT CCT
GATA3	shGATA3	GCC TAC ATG CTT TGT GAA CAA

HEB	shHEB	CCA TCC CAT AAT GCA CCA ATT
RUNX1	shRUNX1	CAG AGT CAG ATG CAG GAT ACA
MYB	shMYB	CCA GAT TGT AAA TGC TCA TTT
LMO1	shLMO1	CGC GAC TAC CTG AGG CTC TTT

Overexpression experiments

The MSCV-IRES-GFP retrovirus vector encoding the *ALDH1A2* cDNA cloned from Jurkat cells was co-transfected with the packaging plasmid pMD-MLV and the envelope plasmid pCMV-VSV-G into 293T cells using FuGENE6 (Roche). Supernatants containing retrovirus particles were collected, filtered through a 0.45 μm filter (Thermo) and concentrated 10-times using a centrifugal filter (Amicon Ultra; Merck). Jurkat cells were then infected in the presence of polybrene (8 μg/ml; Millipore) by centrifugation at 1,300 rcf for 1.5 hr at 32 degrees. Successfully-infected cells that express GFP were sorted using FACS Aria Flow Cytometer (BD Biosciences). Single clones were screened by Western blot analysis.

Inhibition of regulatory element

The single-guide RNAs (sgRNAs) targeting the TAL1-bound region within *ALDH1A2* gene were designed according to the Zhang lab's recommendation (http://crispr.mit.edu/) and cloned into the FgH1tUTG inducible lentivirus expression vector which encodes EGFP as a selection marker (Addgene plasmid #70183). The original EGFP sequence of dCas9-KRAB (Addgene plasmid #71237) was replaced with mCherry sequence. Jurkat cells expressing the dCas9-KRAB-mCherry was established first, followed by an infection of FgH1tUTG lentivirus. EGFP-positive and mCherry-positive cells were selected using FACS Aria Flow Cytometer (BD Biosciences). The sgRNA sequences are described below:

Target Genes	sgRNA IDs	sgRNA sequences
ALDH1A2	gRNA #1	CGT AAC ACT TGG CCG TTG ACT GG
ALDH1A2	gRNA #2	CCT GTG GCT GGA TTA TAG ACA GG

Quantitative reverse transcription PCR (qRT-PCR)

The total RNA from cells was extracted using the NucleoSpin RNA kit (Macherey-Nagel). In total, 1 µg of the purified total RNA was then reverse-transcribed using QuantiTect (QIAGEN). Quantitative reverse transcription PCR (qRT-PCR) was performed on a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific) using Power SYBR Green PCR Master Mix (Roche) and specific primers for individual genes. The PCR primer sequences of the genes of interest are as follows:

Target Genes	Direction	PCR primer sequences (5' to 3')
ALDH1A2 Short isoform	Forward	TGC CTC TTC CTC TCT AAC AGG C
	Reverse	TTG CAG GAA TGG TTT GCC AC
ALDH1A2 Full length	Forward	CAAGATAGAGATGCCCGGCG
	Reverse	ACAGGGAACACTCTCCCACT
ACTINB	Forward	CAG TCG GTT GGA GCG AGC A
	Reverse	GTG GGG TGG CTT TTA GGA TGG
TAL1	Forward	TTC CCT ATG TTCACC ACC AA
	Reverse	AAG ATA CGC CGC ACAACT TT
GATA3	Forward	TTC AGT TGG CCTAAG GTG GT

	Reverse	CGC CGG ACT CTT AGAAGC TA
RUNX1	Forward	GTG TCT TCA GCCAGA TG
	Reverse	CGA CTG TGT ACC GTGGAC TG
MYB	Forward	TGT TGC ATG GATCCT GTG TT
	Reverse	AGT TCA GTG CTG GCCATC TT
ALDOC	Forward	CGCACAGGGAGCTGTCAC
	Reverse	GACTCATCCGCAGCCAGAAT
ENO1	Forward	GCCCTGGTTAGCAAGAAACTGA
	Reverse	GGCGTTCGCACCAAACTTAG
HK1	Forward	GCTGAGTGCCTGGGAGATTT
	Reverse	AGGTGATCAGGATGGCCTCA
HK2	Forward	CCTCGGTTTCCCAACTCTGC
	Reverse	TGGTCAACCTTCTGCACTTGG
LDHA	Forward	GCTGTTCCACTTAAGGCCCC
	Reverse	AAGGAATCGGGAATGCACGTC
PKM	Forward	CGAGCCTCAAGTCACTCCAC
	Reverse	GTGAGCAGACCTGCCAGAC
SLC2A1	Forward	TCACTGTCGTGTCGCTGTTT
	Reverse	GATGGCCACGATGCTCAGAT
SLC2A3	Forward	TTCCTGAGGACGTGGAGAAAAC
	Reverse	TATCAGAGCTGGGGTGACCTTC
TPI1	Forward	CGTTGGGGGAAACTGGAAGAT
	Reverse	CACAAACCACCTCGGTGTCG

GPI	Forward	CGACTAGTGCACAGGGAGTG
	Reverse	CCATGGCGGGACTCTTGC
PFKFB3	Forward	CCCTTCAGGAAAGCCTGTGG
	Reverse	CTCCCCGACGTTGAACACTT
PFKL	Forward	AGAGGCGTGGCACCTCA
	Reverse	GCAAAGCCAGTTTCCTGCTG
PFKFB4	Forward	GAGCTTGGGCAGGAACTTCA
	Reverse	TGGTCATGCACAATGTCCCG
ASNS	Forward	GCAGCTGAAAGAAGCCCAAGT
	Reverse	TGTCTTCCATGCCAATTGCA
ASS1	Forward	TCTACAACCGGTTCAAGGGC
	Reverse	TCCAGGATTCCAGCCTCGTA

Protein extraction and Western blot analysis

Cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (20 mM Tris-HCl at pH 7.5, 150 mM NaCl, 1 mM Na2EDTA, 1mM EGTA, 1% NP-40, 1% sodium deoxycholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na3-VO4, 1 μg/mL leupeptin) (Cell Signaling Technology) with a protease inhibitor cocktail (Roche). Equivalent amounts of protein were diluted in Laemmli sample buffer (Bio-Rad) containing β-mercaptoethanol (Bio-Rad) were resolved on an SDS-PAGE gel. The proteins were subsequently transferred onto polyvinylidenedifluoride (PVDF) membranes (Bio-Rad). The membranes were blocked with 5% non-fat milk and subjected to immunoblot analysis with specific antibodies against ALDH1A2 (Abcam), TAL1 (Millipore), PARP, Caspase 3, α-tubulin, and β-actin (Cell Signaling Technology). The membranes were washed with Tris-buffered saline (TBS)-

Tween (0.1%), incubated with secondary antibodies, and subsequently visualized using the enhanced chemiluminescence method (Thermo).

RNA-sequencing (RNA-seq) analysis

RNA was extracted using an miRNeasy kit (Qiagen) followed by DNase treatment (Turbo DNA-free kit, Ambion). Construction of the strand-specific library and sequencing of the single-end 100-bp-long reads by a BGISEQ sequencer were conducted at BGI Biotech Solutions Co., Ltd. (Hong Kong). RNA-seq data were aligned to the hg19 human genome sequence using STAR 2.5.2a with the parameter --outFilterMultimapNmax set to 1. FeatureCount version 1.6.1, which assigns mapped reads in .bam file format, was used to generate count tables based on Ensembl gene annotation. The Bioconductor package DESeq2 version 1.12.4 was used in the intersection-strict mode to analyze differential gene expression. Deeptools version 2.0 was used to generate the bigwig files with normalized cpm for visualization using UCSC genome browser. Ggplot2 was used to generate the volcano plot.

Gene Ontology Analysis

Analysis was performed using Enrichr tool (https://amp.pharm.mssm.edu/Enrichr/). Enriched pathways showed in this paper are ranked by a combined score calculated using p-value and z-score.

Dataset availability

The ChIP-seq datasets for the transcription factors in Jurkat cells were downloaded from the Gene Expression Omnibus (GEO) (GSE29181, GSE68976, GSE59657 and GSE17954)

(1, 2). The ChIP-seq datasets for H3K27ac in various cell samples were downloaded from the GEO (GSM1296384, GSM1442003, GSM2037781, GSM2037790, GSM1013125, GSM772987, SRX290665 and SRX290666) (1, 3-5). The RNA-seq dataset of knocking down of each member of the TAL1 complex (TAL1, E2A, HEB, LMO1, GATA3, RUNX1, and MYB) were previously reported by us(6) and has been deposited in the GEO database (GSE97514). The RNA-seq datasets for baseline expression in various T-ALL cell lines were previously reported by us(7) and has been deposited in the GEO (GSE103046). The RNA-seq after depletion of ALDH1A2 in Jurkat cells has been deposited in the GEO (GSE136416). The RNA-seq datasets for the primary cell samples from the three cohorts were downloaded from the EGA (EGAD00001002151) (8) and the dbGaP (phs001513.v1.p1 and phs000218)(9). Microarray gene expression in murine hematopoiesis cells was reported by Seita J et al.(10) and visualized by the Gene Expression Commons (https://gexc.riken.jp/models/1507/genes). The single-cell sequencing dataset of prenatal murine thymus was reported by Kernfeld et al.(11) and analyzed using Seurat version 2.3.4 (https://satijalab.org/seurat/). The RNA-seq datasets for primary samples from three cohorts were downloaded from the EGA database (EGAD00001002151),(8) dbGaP database (phs001513.v1.p1(9) and phs000218). RNA-seq gene expression in different cancer cell lines was obtained from the Broad Institute Cancer Cell Line Encyclopedia (CCLE)(https://portals.broadinstitute.org/ccle). Other RNA-seq and ChIP-seq datasets were downloaded from public database with the following accession numbers:

Reference	**GEO ID**	**Description of databases**

Casero et al	GSE69239	RNA-seq in different stages of human hematopoiesis
Maehr et al	GSE107910	Single-cell sequencing database of prenatal murine thymus
Wang X et al	GSE120337	RNA-seq in K562

Cluster analysis of primary T-ALL patients sample

Primary T-ALL samples were clustered based on mRNA expression level of transcription factor genes (*TAL1, TAL2, TLX1, TLX3, HOXAs, LMO1, LMO2, LYL1* and *NKX2-1*). An unsupervised consensus clustering analysis based on gene expression level was also performed for the cohort from TARGET(12) using R package of Consensus Cluster Plus version 1.46.0. The top 5,000 genes with the largest median absolute deviation (MAD) among all samples were selected for K-means cluster analysis. Stable consensus clusters were obtained after 1,000 iterations.

Recombinant protein synthesis

Recombinant proteins for the long and short isoforms of ALDH1A2 were synthesized at GenScript (USA). Briefly, target DNA sequences were cloned into vector Pet-30a(+) with His tag. Recombinant plasmids were then transformed to *E.coli* strain BL21-star. Single colony culture followed by scale up culture were inoculated in TB medium and incubated at 37C followed by IPTG induction. Cell pellets were suspended with lysis buffer followed by sonication. Target protein was obtained by one-step purification using Ni column followed by sterilization. Recombinant proteins were stored in the buffer containing 1mM TCEP.(13) The protein purity and molecular weight were determined by standard SDS-PAGE along

with Western blot confirmation. SEC-HPLC analysis was then performed to verify the correct folding of recombinant proteins. Synthesized sequence of full-length and short isoform of ALDH1A2 are as follows:

Full length:

MHHHHHHMTSSKIEMPGEVKADPAALMASLHLLPSPTPNLEIKYTKIFINNEWQNSESGR
VFPVYNPATGEQVCEVQEADKADIDKAVQAARLAFSLGSVWRRMDASERGRLLDKLADL
VERDRAVLATMESLNGGKPFLQAFYVDLQGVIKTFRYYAGWADKIHGMTIPVDGDYFTFT
RHEPIGVCGQIIPWNFPLLMFAWKIAPALCCGNTVVIKPAEQTPLSALYMGALIKEAGFPP
GVINILPGYGPTAGAAIASHIGIDKIAFTGSTEVGKLIQEAAGRSNLKRVTLELGGKSPNIIFA
DADLDYAVEQAHQGVFFNQGQCCTAGSRIFVEESIYEEFVRRSVERAKRRVVGSPFDPT
TEQGPQIDKKQYNKILELIQSGVAEGAKLECGGKGLGRKGFFIEPTVFSNVTDDMRIAKEE
IFGPVQEILRFKTMDEVIERANNSDFGLVAAVFTNDINKALTVSSAMQAGTVWINCYNALN
AQSPFGGFKMSGNGREMGEFGLREYSEVKTVTVKIPQKNS.

Short isoform:

MHHHHHHMDASERGRLLDKLADLVERDRAVLATMESLNGGKPFLQAFYVDLQGVIKTFR
YYAGWADKIHGMTIPVDGDYFTFTRHEPIGVCGQIIPWNFPLLMFAWKIAPALCCGNTVVI
KPAEQTPLSALYMGALIKEAGFPPGVINILPGYGPTAGAAIASHIGIDKIAFTGSTEVGKLIQ
EAAGRSNLKRVTLELGGKSPNIIFADADLDYAVEQAHQGVFFNQGQCCTAGSRIFVEESI
YEEFVRRSVERAKRRVVGSPFDPTTEQGPQIDKKQYNKILELIQSGVAEGAKLECGGKGL
GRKGFFIEPTVFSNVTDDMRIAKEEIFGPVQEILRFKTMDEVIERANNSDFGLVAAVFTNDI
NKALTVSSAMQAGTVWINCYNALNAQSPFGGFKMSGNGREMGEFGLREYSEVKTVTVKI
PQKNS.

Enzymatic assay

A total of 2 ug recombinant enzymes of each isoform was incubated in freshly made assay buffer containing 20 mM HEPES, 150 mM KCl, 1mM EDTA at 37C with 10 mM NAD+ and 250 uM all-trans retinaldehydes. Samples were added into Costar 96 well clear bottom black side plate. The generation of NADH was measured using Picoprobe (Abcam). Wells containing all reagents without recombinant enzymes were used for background reading. Fluorescence signal (Ex/Em:535/587) were detected continuously for 20 mins using Tecan Infinite M200 PRO Microplate Reader. Reaction velocity of each isoform was calculated and normalized.

Luciferase reporter assay

The reported tandem repeats of retinoic acid response element (RARE)(3x-rev-rev-fwd) was cloned into a *luciferase* reporter construct (pGL4.26 luc2/minP/Hygro)(Promega)(14). The construct was transfected into Jurkat cells using a NEON® Transfection System (Life Technologies) and selected using 400 μg/ml of Hygromycin B (Sigma-Aldrich). Stable clones expressing the luciferase was expanded and then infected with lentiviruses encoding shRNA targeting *ALDH1A2* or control (shGFP). The *luciferase* activity was then determined using the One-Glo Luciferase Assay (Promega) on a Tecan Infinite® 200 PRO plate reader (Tecan). Cell viability were measured by fluorescence reading and used for normalization. Primers for cloning the RARE elements are as follows:

Sense: 5’-TCG AGT GAA CTT TCG GTG AAC CCT ACC CTC GAG TGA ACT TTC GGT GAA CCC TAC CCT CGA GGG TAG GGT TCA CCG AAA GTT CAC -3’

Antisense: 5’-AGC TGT GAA CTT TCG GTG AAC CCT ACC CTC GAG GGT AGG GTT CAC CGA AAG TTC ACT CGA GGG TAG GGT TCA CCG AAA GTT CAC -3’

EGFP knock-in experiments

The knock-in of *EGFP* gene was performed using Golden Gate Assembly as reported by Yamamoto et al.(15) Briefly, sgRNA targeting near the last codon of *ALDH1A2* was cloned into the Cas9 vector pX330A-1x2 (Addgene #58766). PITCh-gRNA was cloned into the modified Cas9 vector from PX330S-2-PITCh (Addgene #63670). Then, a donor vector containing the region of EGFP-2A-puro flanked by 5’ and 3’ microhomology sequences was constructed using pCRIS-PITChv2-FBL (Addgene #63671). The PCR primer pairs were designed to amplify the microhomology with correct reading frames spanning the cleavage sites as well as to cover both of MluI restriction sites beyond PITCH sequences. The PCR fragments and pCRIS-PITChv2-FBL (Addgene #63671) were then digested with MluI and ligated. The reconstructed plasmids with correct inserted orientation were validated by Sanger sequencing. Jurkat or K562 cells were co-transfected with sgRNA-Cas9 vectors and donor plasmids in a 2:1 ratio using a NEON® Transfection System (Life Technologies) followed by puromycin selection. The homozygous clones were selected following previous reports.(16)

Confocal microscopic imaging

Images of EGFP-fused ALDH1A2 protein localization in live cells were captured using an Olympus FV1000 TIRF confocal microscope (C4). Mito Tracker Red CMXRos (Thermo) was used to stain mitochondria. To visualize the nuclei, the cells were stained with Hoechst

33342. Colocalization of Mito Tracker Red and EGFP signals was analyzed using the IMARIS 9.5 software with an automated intensity thresholding setting.

Antioxidant treatment

The cells were treated with 500uM freshly prepared N-Acetyl-L-cysteine (Sigma-Aldrich) at 24hrs post virus infection and then ROS level was detected at 72hrs after infection. The data was analyzed using the FlowJo software.

Metabolomic profiling

For metabolite extraction, the cells were washed twice with 5 % (w/w) mannitol solution at room temperature and then harvested in 800ul methanol and 550ul diluted Internal Standard Solution (HMT). Cell extract was transferred into a microtube followed by centrifugation at 2,300xg at 4 degrees. The supernatant was transferred into centrifugal filter units and centrifuged at 9,100xg at 4 degrees for 5 hrs. The extracted metabolites were analyzed by a capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS) Basic Scan conducted by Human Metabolome Technologies (HMT). The number of metabolites was normalized to cell volume, as detected by a Scepter 2.0 cell counter (Millipore).

Extracellular flux analysis

Seahorse XF24 Cell Culture Microplates (Agilent) were coated with Cell-Tak (Corning) before the experiment. The cells were washed with PBS and cell numbers were counted by Scepter 2.0 cell counter. The cells were re-suspended in Seahorse assay medium (Agilent) and then 250,000 cells were seeded into each well of XF24 plates followed by centrifugation. Plates were incubated at 37°C in a non-CO2 incubator for 30 mins before

the analysis by Seahorse XF24 (Agilent). Glucose was used at 10 mM, oligomycin at 2.5 μM, and 2-Deoxy-D-glucose (2-DG) at 50 mM for glycolysis stress assay. Oligomycin and carbonyl cyanide-4-(trifluoromethoxy) phenylhydrazone (FCCP) were used at 2 μM, rotenone and antimycin A were used at 0.5 μM for mitochondrial stress assay. Extracellular acidification rate (ECAR) measurements from glycolysis stress tests are normalized non-glycolytic ECARs. Oxygen consumption rate (OCR) measurements from mitochondrial stress tests are normalized to basal respiration rate, and Percent glycolytic ATP was calculated with values of glycolytic ATP obtained from the glucose-induced proton production rate and OxPhos ATP obtained from oligomycin decrease in OCR. All materials were obtained from Agilent Technologies.

Apoptosis assay

Cells was stained using Hoechst 33258 or Propidium iodide and APC Annexin V (Biolegend). Apoptotic population was measured within the Annexin V positive and PI/Hoechst negative gating.

Nutrient starvation assay

Dox-inducible *ALDH1A2* knockout Jurkat cells were treated with 1μM doxycycline or DMSO for 24 hours. The cells were then washed with PBS and resuspended in regular, glucose-free or glutamine-free RPMI 1640 medium (all by Thermo Fisher Scientific), each of which was supplied with 10% dialyzed fetal bovine serum (Thermo Fisher Scientific) and contained DMSO or doxycycline.

NADH/NAD$^+$ ratio measurement

Cells washed twice with PBS were seeded into 96 wells plate and then lysed with 0.2N NaOH with 1% DTAB. To measure NAD+, lysed cells were treated with 0.4N HCl and incubated at 60°C for 15 minutes. To measure NADH, lysed cells were incubated at 60°C for 15 minutes. Samples were then equilibrated to room temperature. 0.5M Trizma® base was then added to HCl treated samples for NAD^+ measurement. Luminesenece signals were measured using a Tecan Infinite M200 PRO Microplate Reader using the NAD/NADH-Glo™ Assay Kit (Promeaga).

ATP quantification

The cells were resuspended in glucose-free RPMI-1640 medium (Thermo Fisher Scientific) supplemented with 10% dialyzed FBS (Thermo Fisher Scientific). The ATP levels were measured using the CellTiter Glo kit (Promega). Data were first normalized to the Cell Titer Glo level at 0 min and then normalized to the values from the cells cultured in regular media.

Zebrafish transgenesis

Zebrafish studies were conducted according to the recommendations from the Institutional Animal Care and Use Committee (IACUC) and all protocols were approved by the Committee at the National University of Singapore (NUS). The human *ALDH1A2* gene was cloned under the zebrafish *rag2* promoter into the I-SceI meganuclease-based vector to establish transgenic lines. The fluorescence protein (mCherry) was introduced into the same vector backbone. The *rag2-myr-mAKT2* construct was kindly provided by Alejandro Gutierrez (17). All constructs were linearized using the I-SceI enzyme (New England Biolabs). The buffered mixture was co-injected into zebrafish embryos at the one-cell stage to create *Tg(rag2: ALDH1A2)* line. A control *Tg(rag2: mCherry*) line was established by

injecting *rag2-mCherry* only. Zebrafish thymus was harvested and stained with CellROX Deep Red reagent (Thermo Fisher Scientific) for ROS detection according to the manufacturer's instruction. Fluorescence signals were analysed with a BD™ LSR II flow cytometer using BD FACSDiva™ software. Compensation was performed to detect DeepRed signal to be differentiated from mCherry signal. The primers used for genotyping were as follows:

rag2 forward: 5'-GTTTGTTAATGCTTCTCTTGCGTAT-3';

hALDH1A2 reverse: 5'-CTACAGGAATGGTCATCCCATGAA-3';

mAKT2 reverse: 5'-GAAGGCTGGCTCCACAAACGTGG-3'.

Statistic analysis

For all data in which the values were normalized to control samples (Figure 1C, 3D, 3E and 4C; and Supplemental Figures 2C, 3C and 4D), error bars in control samples were considered following the propagation of uncertainty:

$$\sigma_{f(Foldchange)} \approx |f| \sqrt{\left(\frac{\sigma_{control}}{control}\right)^2 + \left(\frac{\sigma_{treatment}}{treatment}\right)^2 - 2\frac{\sigma_{(control,\ treatement)}}{control \times treatment}};$$

Based on this, variation in control samples was found to be equal to zero in Figures 3D, 3E, Supplementary Figure 2C and 3C, and thus there was no error bars. In Figures 1C, 4C and Supplemental Figure 4D which showed qRT-PCR results, the following formula was considered and error bars were shown.

$$\sigma_{f(2^{-\Delta\Delta Ct})} \approx \left|2^{-\Delta\Delta Ct} \times \ln 2 \times \sigma_{\Delta\Delta Ct}\right|;$$

Statistic comparisons were performed by two-tailed Student's t test (for bar charts, dot plots and line charts) or Gehan-Breslow-Wilcoxon test (for survival curve analysis) using Graphpad Prism.

Supplementary Tables

Supplementary Table 1. Consensus clusters of T-ALL Target cohort

Supplementary Table 2. List of genes significantly regulated by *ALDH1A2* depletion in Jurkat cells

Supplementary Figure Legends

Supplementary Figure 1. The *ALDH1A2* gene is directly activated by the TAL1 complex in T-ALL cells. (A) A scheme with filtering criteria used to select genes directly and positively regulated by the TAL1 complex in the T-ALL cells. **(B)** A heatmap indicates 13 genes selected after the second criteria. **(C)** Unsupervised consensus clustering analysis showing *ALDH1A2, TAL1/2* expression in the 262 primary T-ALL cell samples, as reported by Liu, Yet al (12), which were classified into seven stable clusters: ETP, HOXA (2 clusters), TLX, NKX, TAL-RA and TAL-RB T-ALL subgroups. Genetic abnormalities of the driving transcription factors defined by whole genome sequencing were highlighted and correlated among the clusters. The red, blue or green bars represent the expression level of *ALDH1A2, TAL1* or *TAL2* for each individual sample. **(D)** Evidence from the CCLE database showing *ALDH1A2* mRNA expression levels across different cancer cell lines. **(E)** The mRNA expression levels of *TAL1* and *ALDH1A2* in different human hematopoietic cells (two samples each) were analyzed by RNA-seq using the dataset reported by Casero D et al (18). Expression values are shown by transcripts per million (TPM). Hematopoietic stem cells (HSC), lymphoid-prime multipotent progenitors (LMPP), common lymphoid progenitors (CLP), CD4-CD8- double-negative (DN) T-cells (Thy1-3), CD4+CD8+ double-positive (DP) T-cells (Thy4), CD4+ single-positive T-cells (Thy5), CD8+ single-positive T-cells (Thy6). **(F)** Microarray gene expression data from the Gene Expression Commons (10) showing the relative expression levels of the mouse *Aldh1a2* gene in murine hematopoietic cells and bone marrow stroma cells.

Supplementary Figure 2. The short isoform of ALDH1A2 has enzymatic activity and is localized in the cytoplasm in T-ALL cells. (A) Western blot analysis showing the

molecular weight of the purified recombinant ALDH1A2 proteins (long and short isoforms). **(B)** The amount of NADH was measured by a PicoProbe kit after the incubation of all *trans* retinal and NAD+ for each of the ALDH1A2 isoforms. Enzymatic kinetics were generated based on the fluorescence readings. Error bars represent SD from technical replicates. **(C)** Scheme of the luciferase reporter system (top). Luciferase activity was measured in Jurkat cells after shRNA knocking down (bottom). Error bars represent SD from biological replicates. **p<0.01 using two-tailed Student's t test. **(D)** Scheme of the generation of the *EGFP*-fused *ALDH1A2* gene using the PITCH system. PCR-amplified donor homology arms were cloned and co-transfected with a CRISPR/Cas9 plasmid to target the 3' endogenous *ALDH1A2* gene locus. **(E)** DNA sequence showing successful insertion of the *EGFP* gene into the endogenous *ALDH1A2* gene locus. **(F)** The text indicates Pearson coefficient of the pixel-intensity correlation as well as the percentage of colocalization between two signals in Jurkat and K562 cells.

Supplementary Figure 3. ALDH1A2 supports leukemia cell survival in T-ALL cells in culture. **(A)** mRNA expression level of *ALDH1A1, ALDH1A2* and *ALDH1A3* in various T-ALL cell lines measured by RNA-seq (one sample each). **(B)** IC_{50} values for WIN 18,446 for two *TAL1/ALDH1A2*-positive cell lines and two *TAL1/ALDH1A2*-negative cell lines are shown. **p<0,01 using two-tailed Student's t test. **(C)** Cell viability was measured by CellTiter Glo assay after shRNA transduction in Jurkat cells treated with or without all-*trans* retinoic acid (ATRA) at 0, 1, 3 or 10 µM. Error bars represent SD from technical replicates. *p<0.05, **p<0.01 and *** p<0.001 using two-tailed Student's t test. ns, not significant.

Supplementary Figure 4. ALDH1A2 affects the metabolic pathways in T-ALL cells. (A) A volcano plot between -log10 p-values and log2 fold-change values demonstrating the differentially expressed genes after *ALDH1A2* depletion. Downregulated (n=96) and upregulated genes (n=19) are highlighted in blue or red color, respectively. The glycolysis or amino acids metabolism related genes are annotated in the figure. **(B,C)** Gene Ontology analysis showing pathways enriched for the genes downregulated (blue) and upregulated (red) after *ALDH1A2* depletion in Jurkat cells. Pathways are ranked by a combined score. p-value for all pathways showing in the figure is less than 0.05. **(D)** qRT-PCR validation of the genes downregulated or upregulated by *ALDH1A2* depletion in Jurkat cells. mRNA expression of each gene was normalized to *ACTINB* expression. Error bars represent SD from technical replicates. *p<0.05, **p<0.01 and *** p<0.001 using two-tailed Student's t test. ns, not significant. **(E)** Flow cytometry analysis showing cell apoptosis under nutrient starvation condition. Hoeschst and Annexin V staining of control (Dox-) or *ALDH1A2* depletion (Dox+) cells cultured under regular (Control), glucose starvation (Glu-) or glutamine starvation (Gln-) for 24 hours are showed. Q3 population (Hoechst-negative and Annexin V positive) indicates apoptotic cells. **(F)** Quantification of the percentage of apoptotic population upon *ALDH1A2* depletion under the nutrient starvation treatment. Error bars represent SD from technical replicates. **p<0.01 using two-tailed Student's t test.

Supplementary Figure 5. ALDH1A2 supports cellular aerobic glycolysis and energy production in T-ALL cells. (A) Cell viability was measured by CellTiter Glo assay for the control (sgRNA #1, Dox-) and *ALDH1A2*-depleted (sgRNA #1, Dox+) Jurkat cells cultured in glucose-free medium. Cell viability compared with the viability before Dox treatment (0

h) is shown. Error bars represent SD from technical replicates. **p<0.01 and *** p<0.001 using multiple t test. ns, not significant.

Supplementary Figure 6. ALDH1A2 reduces the level of reactive oxygen species (ROS) in T-ALL cells. (A) qRT-PCR analysis showing the expression of *ALDH1A2* short transcript (red) or full-length (blue) in various cell samples. Error bars represent SD from technical replicates. mRNA expression of each gene was normalized to *ACTINB* expression. **(B)** Jurkat cells were transduced with shGFP (control) or shALDH1A2 (#3 or 5) and then treated with NAC. Total ROS levels were measured by flow cytometry analysis after CellROX staining and shown by mean fluorescence intensity (MFI) from single sample.

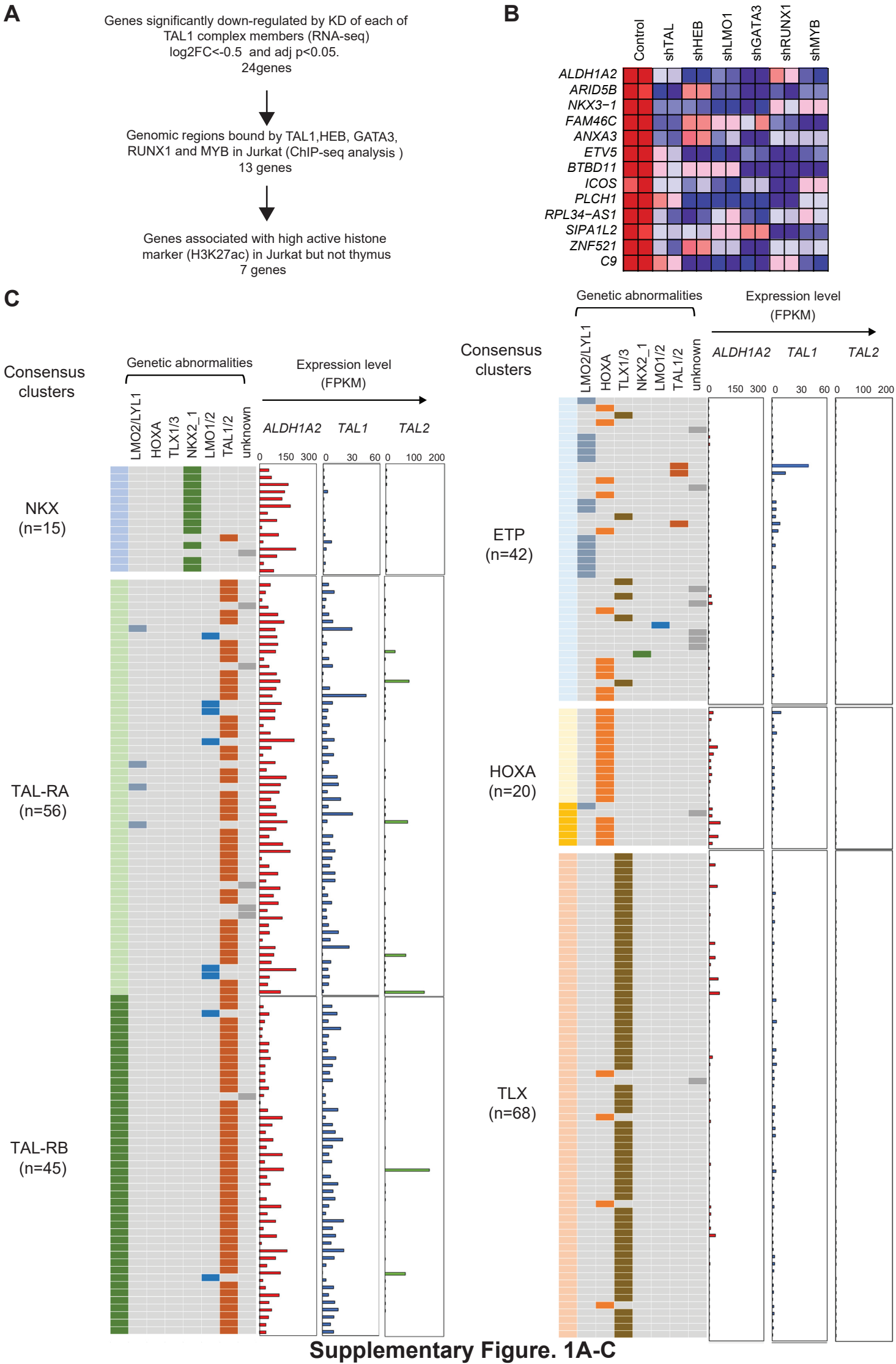
Supplementary Figure 7. *ALDH1A2* overexpression accelerates tumorigenesis in a zebrafish model. (A) Scheme of transgenic zebrafish generation. The human *ALDH1A2* gene (short isoform), *mCherry* (fluorescent marker), *myr-Akt2* or *EGFP* gene was separately cloned under the zebrafish *rag2* promoter. The constructs (*rag2-hALDH1A2* and *rag2-mCherry*) or (*rag2-myr-mAkt2* and *rag2-EGFP*) were injected with the I-SceI enzyme into the one-cell stage zebrafish embryos to establish the transgenic line Tg (rag2-hALDH1A2) or Tg (rag2-myr-mAkt2), respectively. The control zebrafish Tg (rag2-mCherry), which was injected with only the *rag2-mCherry* construct, was also established. **(B)** PCR was performed using genomic DNAs extracted from the *rag2-mCherry, rag2-hALDH1A2* and *rag2-myr-mAkt2* fish. PCR primers targeting the *rag2* promoter (forward), *hALDH1A2* gene (reverse) or *mAkt2* gene (reverse) were used. **(C)** Representative images of the *ALDH1A2* transgenic fish at weeks 6 and 13 post fertilization. The thymus is marked by the fluorescent mCherry signal. **(D)** Representative images of single transgenic (*rag2-myr-mAkt2*) and

double transgenic (*rag2-myr-mAkt2; rag2-hALDH1A2*) animals at 9 weeks. A scale indicates 4 mm.

Supplementary References

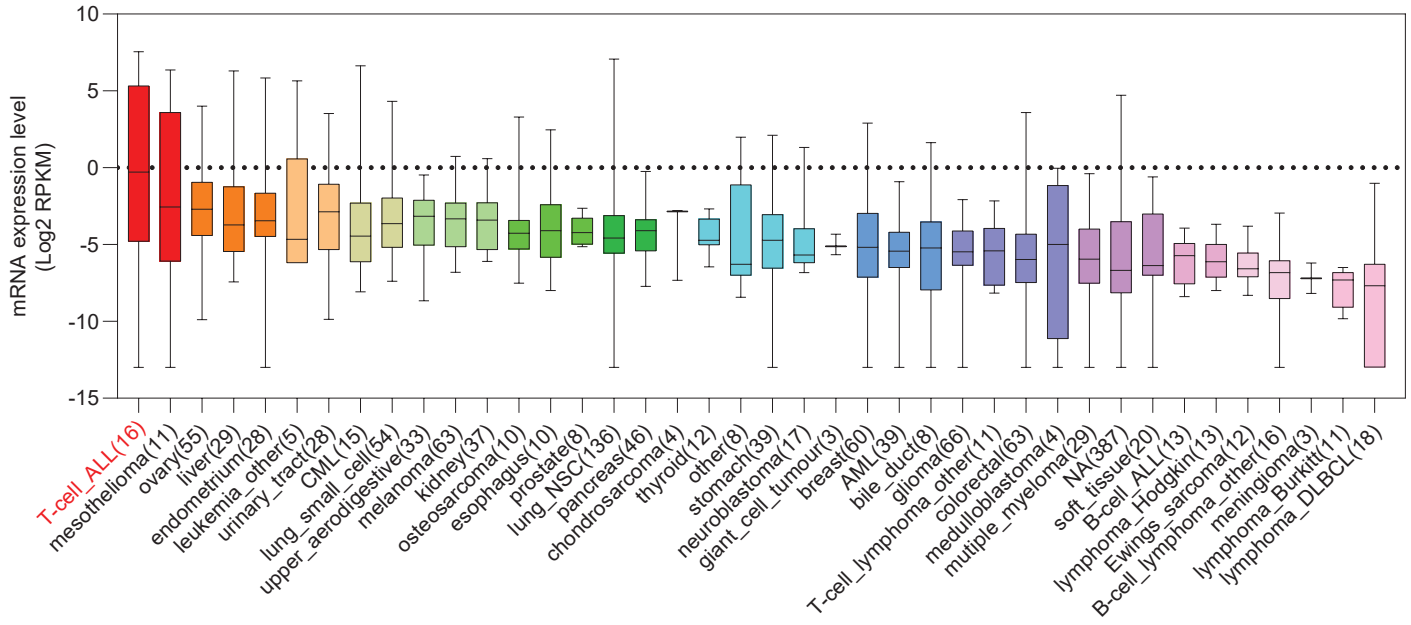
1. Sanda T, Lawton LN, Barrasa MI, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell. Aug 14 2012;22(2):209-221. doi:10.1016/j.ccr.2012.06.007. Cited in: Pubmed; PMID 22897851.
2. Mansour MR, Abraham BJ, Anders L, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science. Dec 12 2014;346(6215):1373-1377. doi:10.1126/science.1259037. Cited in: Pubmed; PMID 25394790.
3. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. Oct 2010;28(10):1045-1048. doi:10.1038/nbt1010-1045. Cited in: Pubmed; PMID 20944595.
4. Abraham BJ, Hnisz D, Weintraub AS, et al. Small genomic insertions form enhancers that misregulate oncogenes. Nat Commun. Feb 9 2017;8:14385. doi:10.1038/ncomms14385. Cited in: Pubmed; PMID 28181482.
5. Vahedi G, Kanno Y, Furumoto Y, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature. Apr 23 2015;520(7548):558-562. Epub 2015/02/18. doi:10.1038/nature14154. Cited in: Pubmed; PMID 25686607.
6. Leong WZ, Tan SH, Ngoc PCT, et al. ARID5B as a critical downstream target of the TAL1 complex that activates the oncogenic transcriptional program and promotes T-cell leukemogenesis. Genes Dev. Dec 1 2017;31(23-24):2343-2360. doi:10.1101/gad.302646.117. Cited in: Pubmed; PMID 29326336.
7. Ngoc PCT, Tan SH, Tan TK, et al. Identification of novel lncRNAs regulated by the TAL1 complex in T-cell acute lymphoblastic leukemia. Leukemia. Oct 2018;32(10):2138-2151. Epub 2018/04/15. doi:10.1038/s41375-018-0110-4. Cited in: Pubmed; PMID 29654272.
8. Qian M, Zhang H, Kham SK, et al. Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. Genome Res. Feb 2017;27(2):185-195. Epub 2016/12/03. doi:10.1101/gr.209163.116. Cited in: Pubmed; PMID 27903646.
9. Aries IM, Bodaar K, Karim SA, et al. PRC2 loss induces chemoresistance by repressing apoptosis in T cell acute lymphoblastic leukemia. J Exp Med. Dec 3 2018;215(12):3094-3114. Epub 2018/11/09. doi:10.1084/jem.20180570. Cited in: Pubmed; PMID 30404791.
10. Seita J, Sahoo D, Rossi DJ, et al. Gene Expression Commons: an open platform for absolute gene expression profiling. PLoS One. 2012;7(7):e40321. Epub 2012/07/21. doi:10.1371/journal.pone.0040321. Cited in: Pubmed; PMID 22815738.
11. Kernfeld EM, Genga RMJ, Neherin K, et al. A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. Immunity. Jun 19 2018;48(6):1258-1270 e1256. Epub 2018/06/10. doi:10.1016/j.immuni.2018.04.015. Cited in: Pubmed; PMID 29884461.
12. Liu Y, Easton J, Shao Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet. Aug 2017;49(8):1211-1218. doi:10.1038/ng.3909. Cited in: Pubmed; PMID 28671688.
13. Paik J, Haenisch M, Muller CH, et al. Inhibition of retinoic acid biosynthesis by the bisdichloroacetyldiamine WIN 18,446 markedly suppresses spermatogenesis and alters retinoid metabolism in mice. J Biol Chem. May 23 2014;289(21):15104-15117. Epub 2014/04/09. doi:10.1074/jbc.M113.540211. Cited in: Pubmed; PMID 24711451.

14. Balkan W, Colbert M, Bock C, Linney E. Transgenic indicator mice for studying activated retinoic acid receptors during development. Proc Natl Acad Sci U S A. Apr 15 1992;89(8):3347-3351. Epub 1992/04/15. doi:10.1073/pnas.89.8.3347. Cited in: Pubmed; PMID 1314386.
15. Sakuma T, Nakade S, Sakane Y, Suzuki KT, Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat Protoc. Jan 2016;11(1):118-133. Epub 2015/12/19. doi:10.1038/nprot.2015.140. Cited in: Pubmed; PMID 26678082.
16. Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Rep. Apr 5 2016;15(1):210-218. Epub 2016/04/08. doi:10.1016/j.celrep.2016.03.001. Cited in: Pubmed; PMID 27052166.
17. Gutierrez A, Grebliunaite R, Feng H, et al. Pten mediates Myc oncogene dependence in a conditional zebrafish model of T cell acute lymphoblastic leukemia. J Exp Med. Aug 1 2011;208(8):1595-1603. Epub 2011/07/06. doi:10.1084/jem.20101691. Cited in: Pubmed; PMID 21727187.
18. Casero D, Sandoval S, Seet CS, et al. Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. Nat Immunol. Dec 2015;16(12):1282-1291. doi:10.1038/ni.3299. Cited in: Pubmed; PMID 26502406.

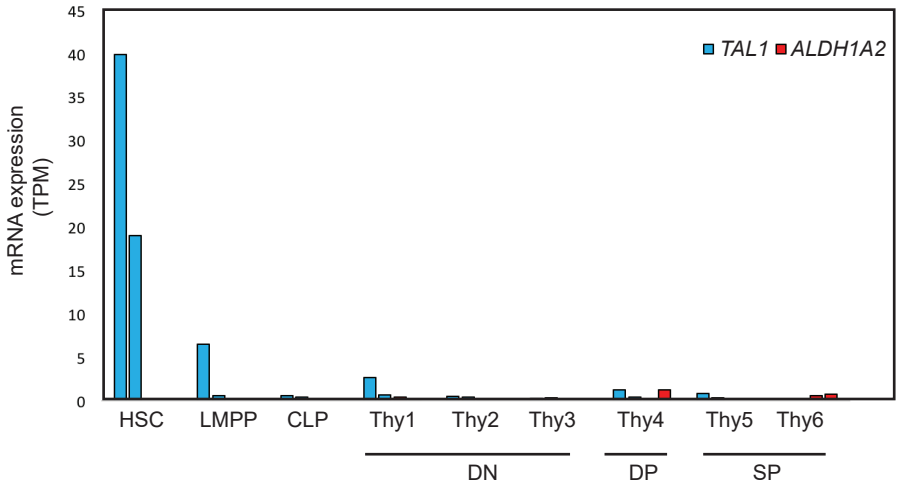


Supplementary Figure. 1A-C

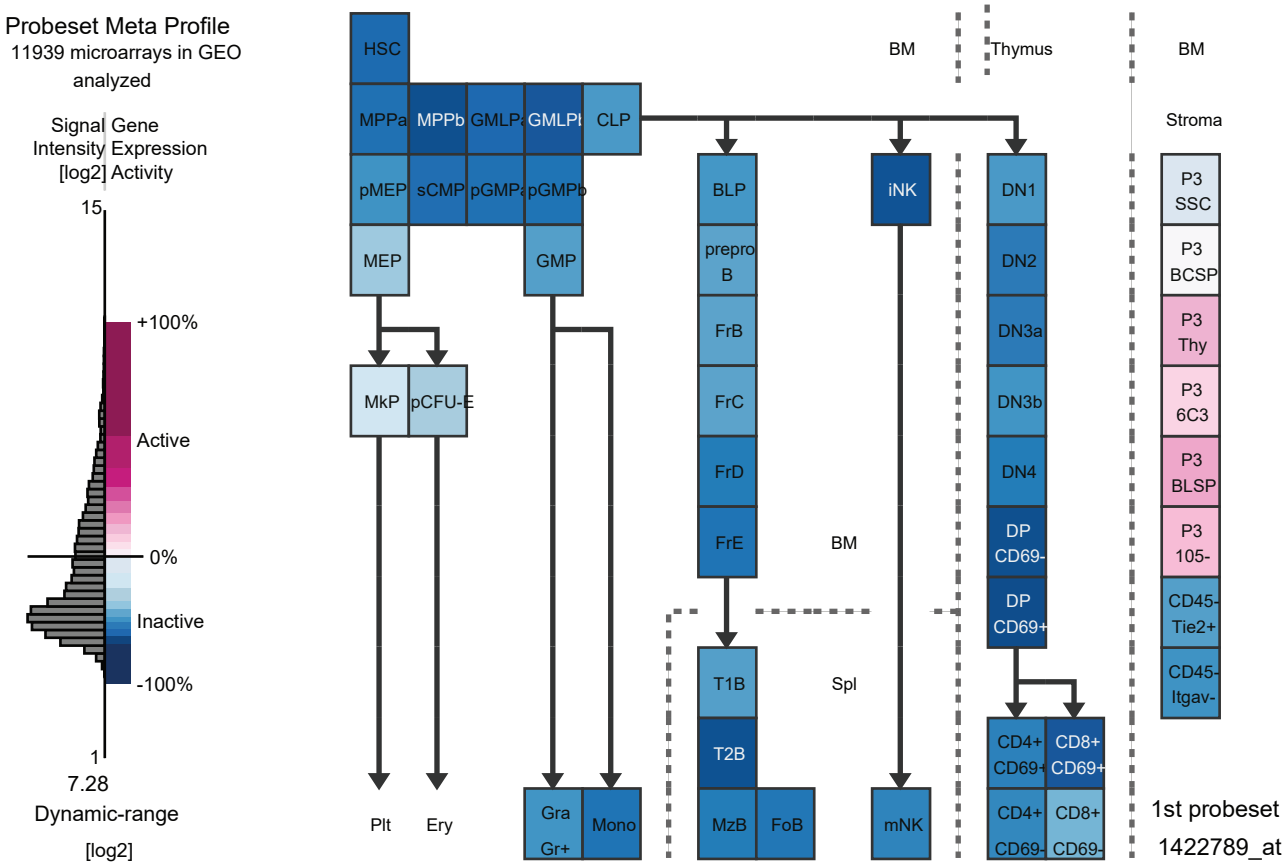
D



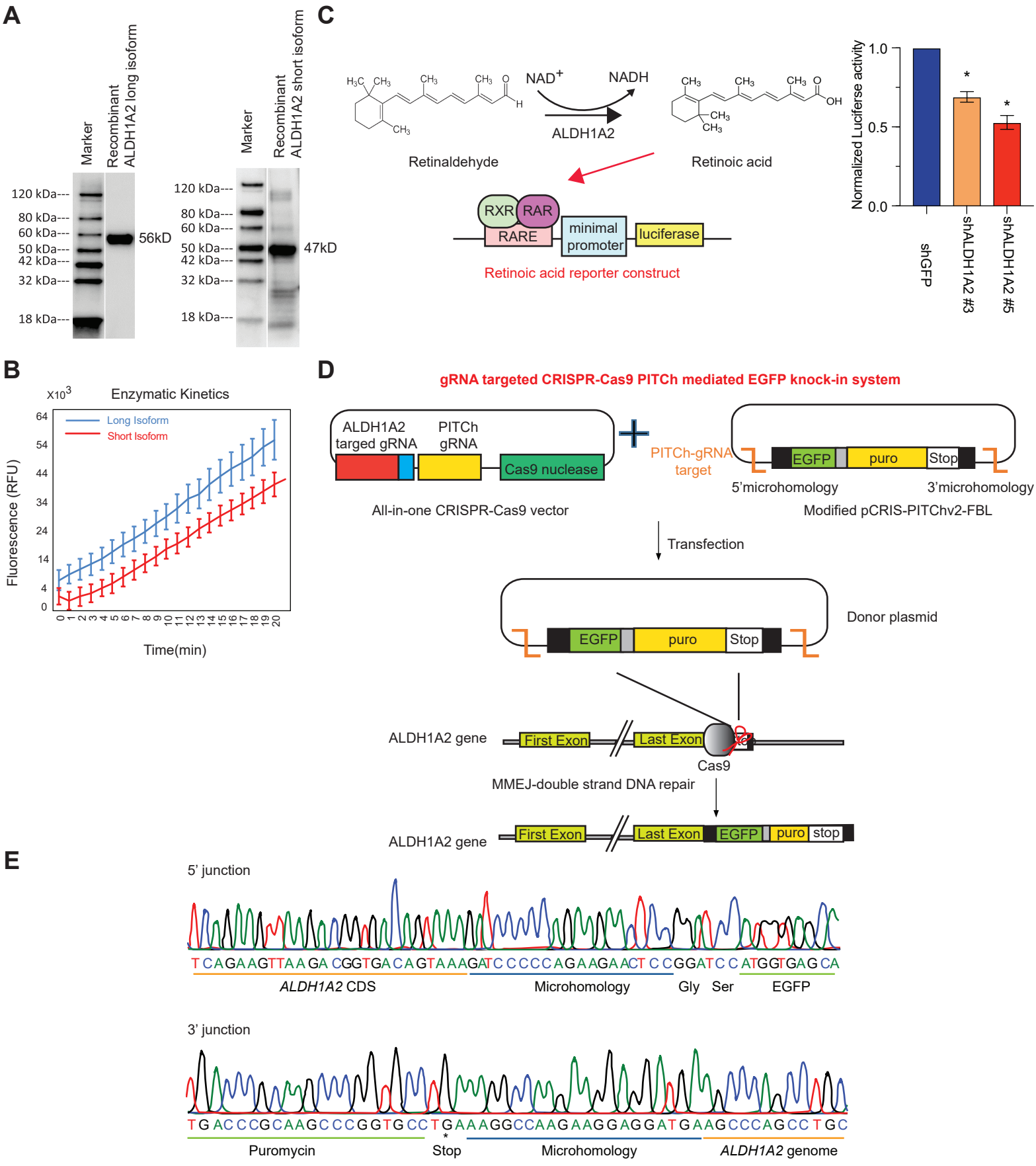
E



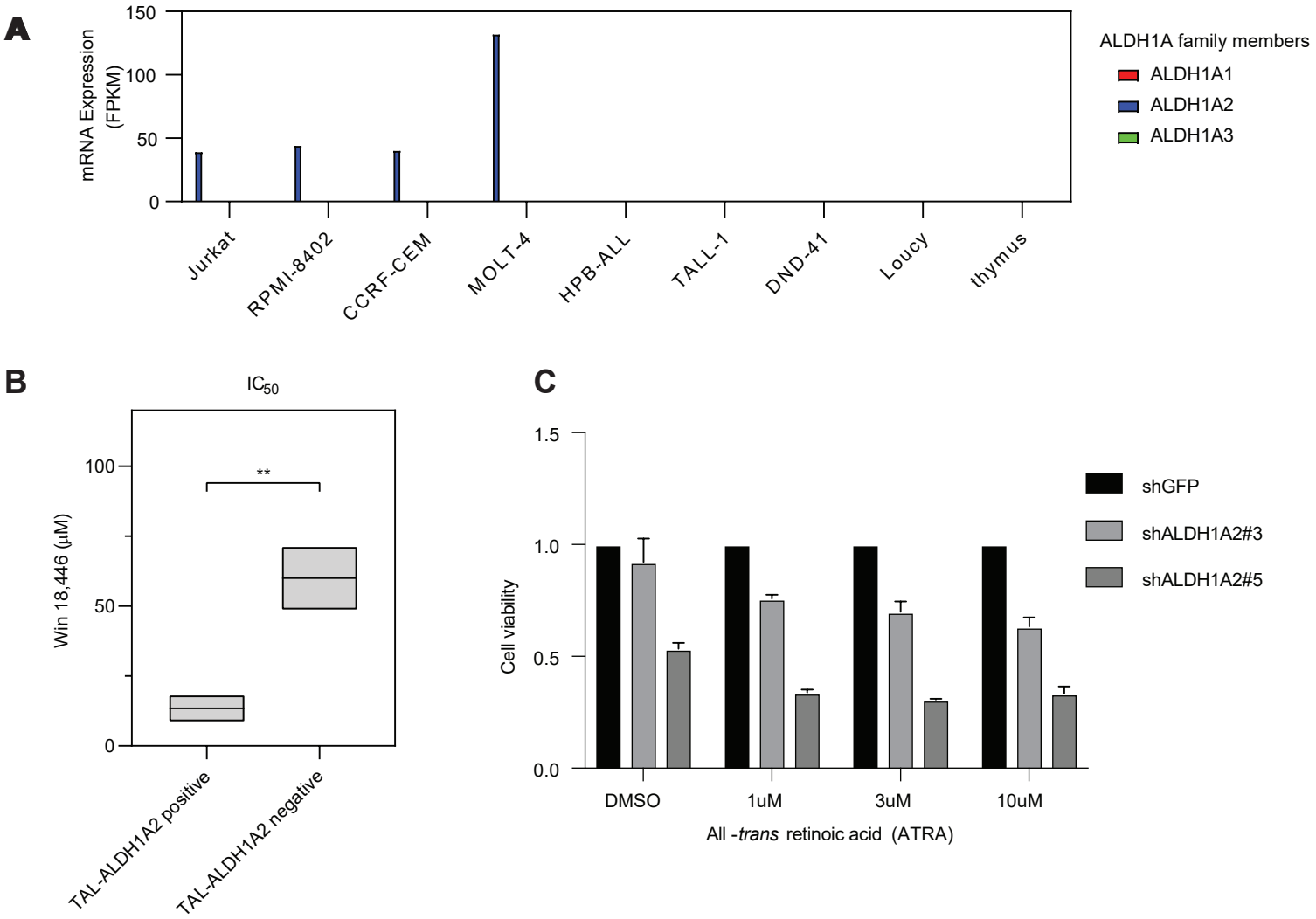
F



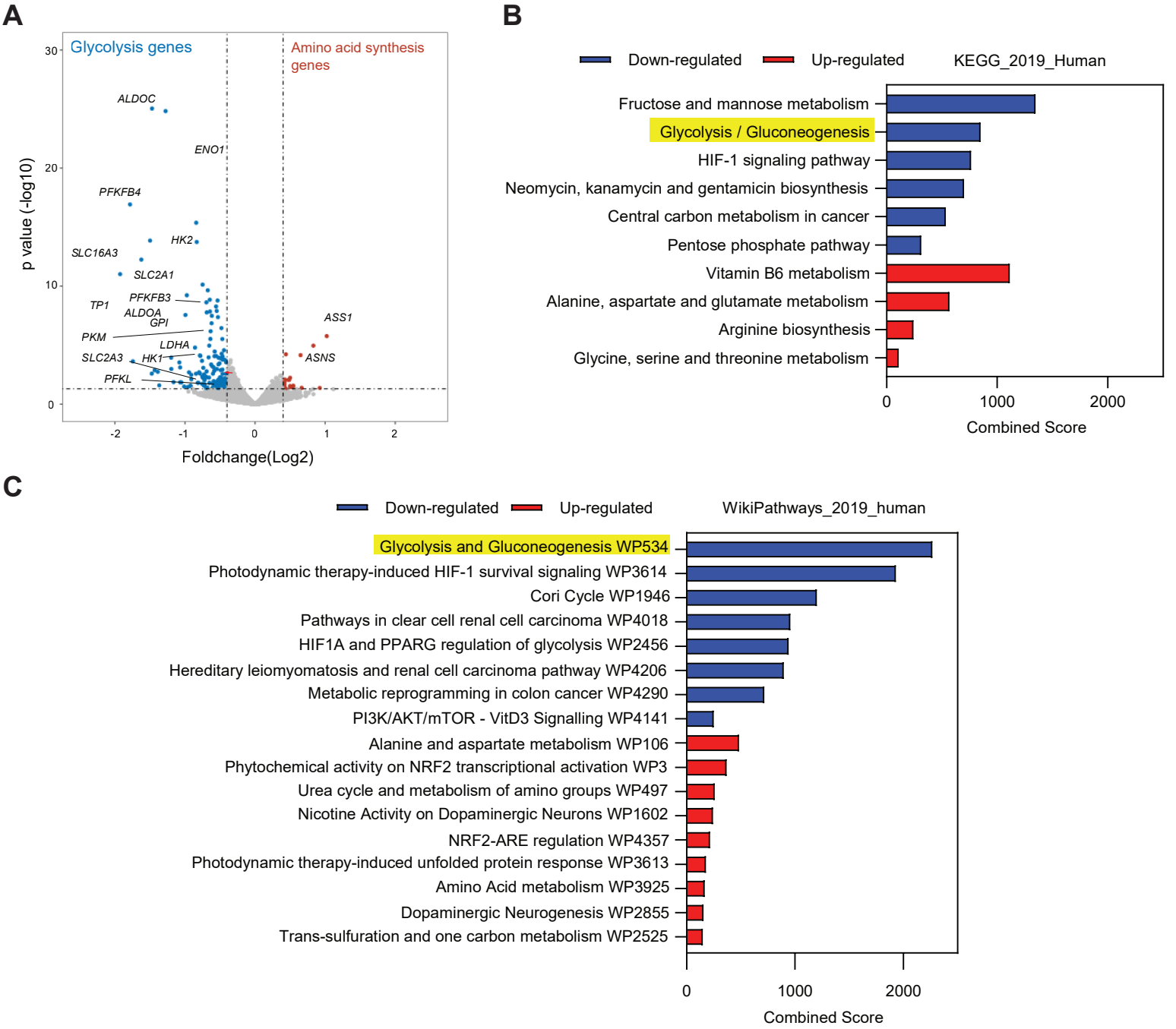
Supplementary Figure. 1D-F



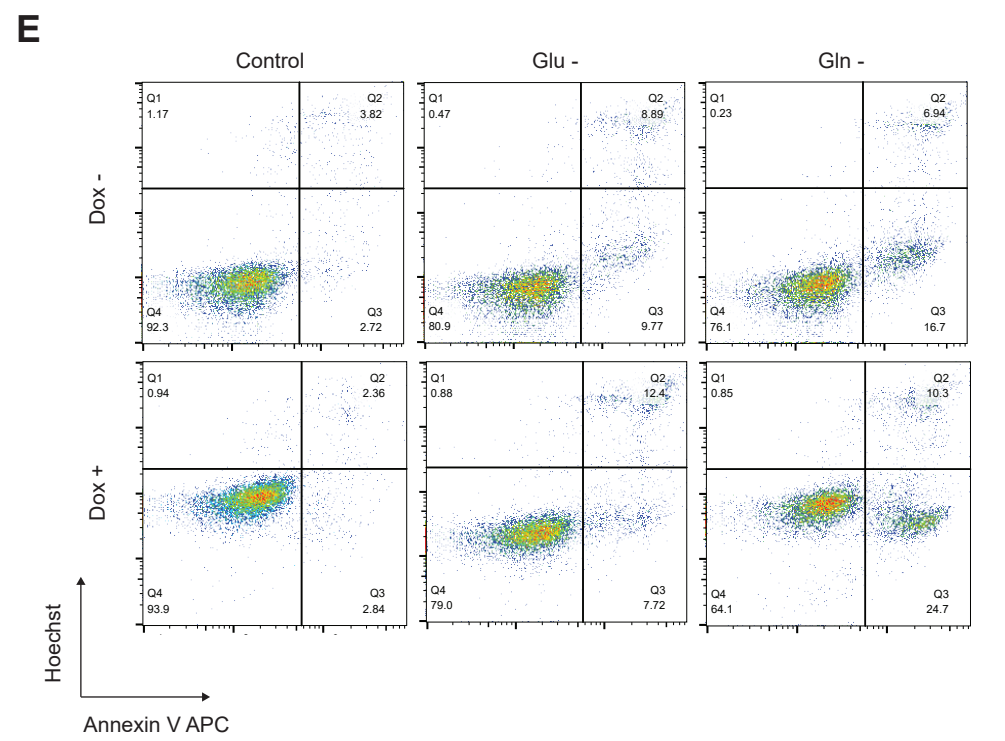
Supplementary Figure. 2

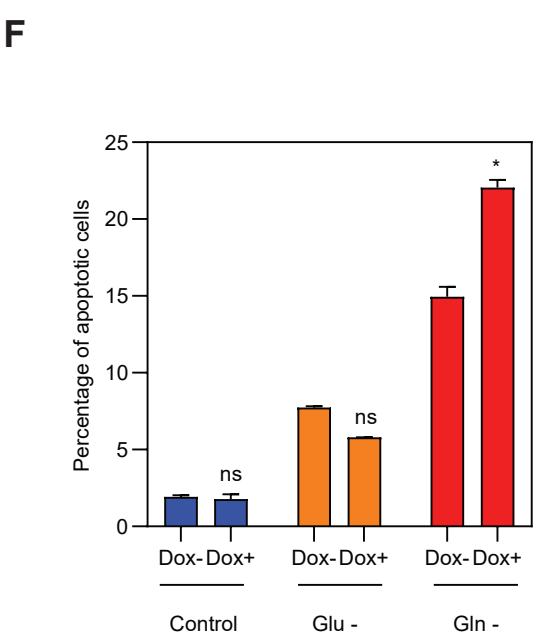


Supplementary Figure. 3



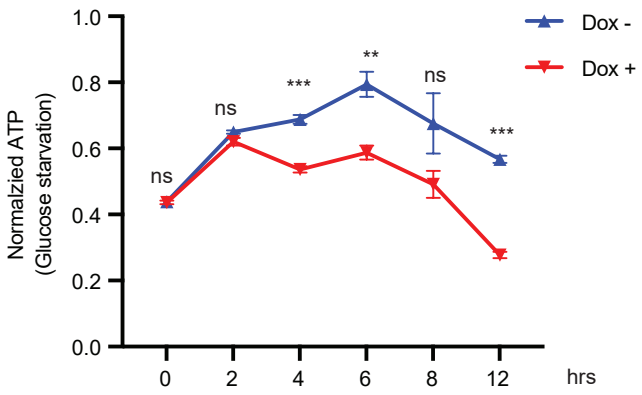
Supplementary Figure. 4A-D



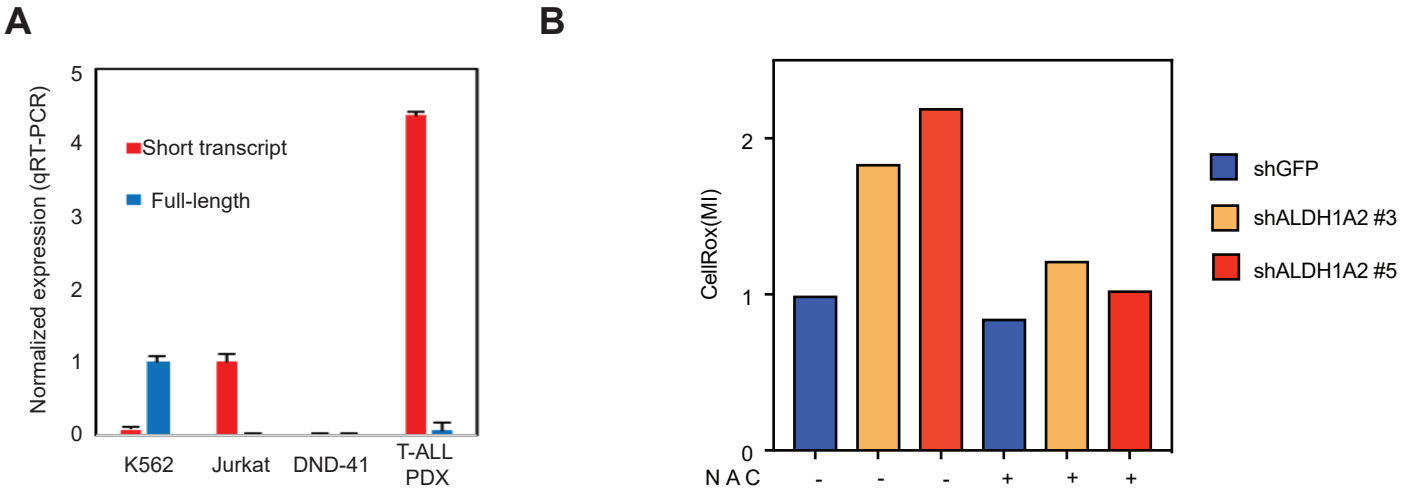


Supplementary Figure. 4E-F

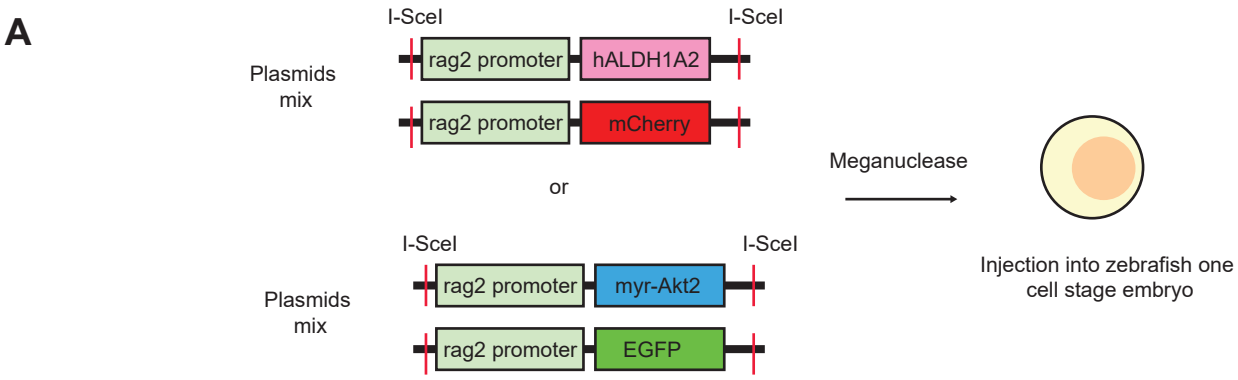
A

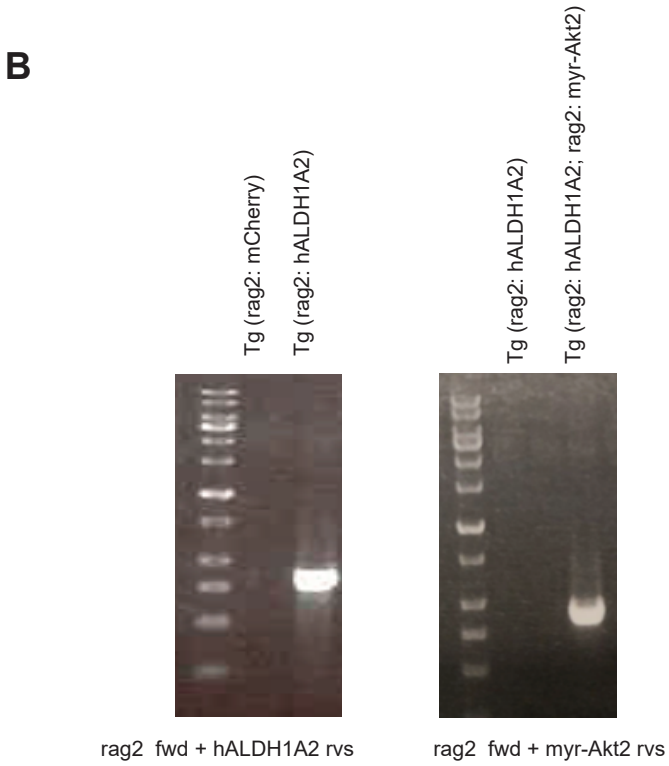


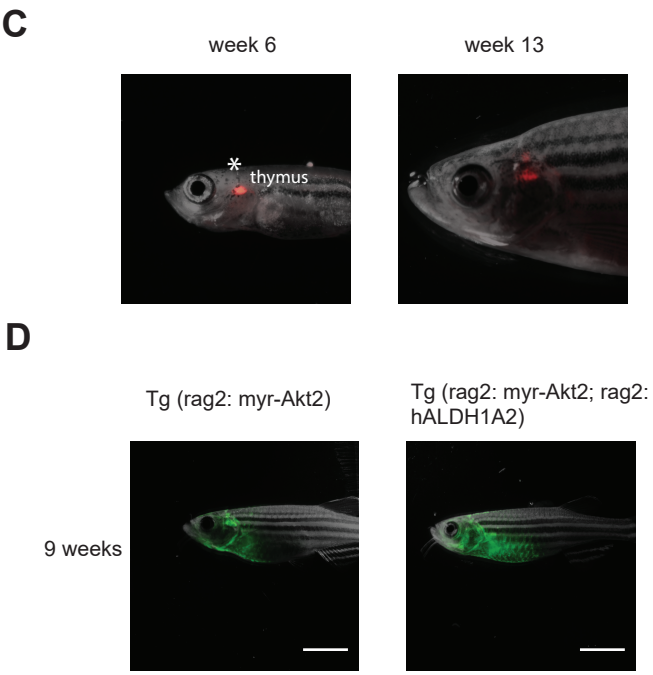
Supplementary Figure 5



Supplementary Figure 6







Supplementary Figure 7