Supplemental Information

Supplemental Table legends

Supplemental Table S1. Gene expression changes in T-ALL induced upon Cic inactivation

T-ALL samples obtained from Tmx-treated *Cic*$^{lox/lox}$;*hUBC-CreERT2*$^{+/T}$ mice (n=5) were compared to WT thymuses (n=3) by RNAseq. This tables shows the 55 overexpressed and 181 downregulated genes (log_2FC of ≥ +3 or ≤ – 3; adjusted p-value <0.05).

Supplemental Table S2. 32 gene Cic LOF signature (CIC_LOF_4)

The 32 gene signature (Cic LOF signature; CIC_LOF_4) was selected based on genes that were upregulated at least with a log_2FC of 1.5 and that harbored at least one Cic binding sequence (CBS) with a conservation score of > 0.875 in their promoter.

Supplemental Table S3. 143 gene Cic LOF signature (CIC_LOF_3)

The 143 gene signature (Cic LOF signature; CIC_LOF_3) was selected based on genes that were significantly upregulated and that harbored at least one Cic binding sequence (CBS) with a conservation score of > 0.875 in their promoter.

Supplemental Materials and Methods

Generation of conditional loss-of-function mice for the *Cic* locus.

A targeting vector (Gene Bridges GmbH) (Supplemental Fig. S1A) was prepared from *Cic* sequences present in BACs bMQ66B11 and bMQ66B1212 and assembled via Red/ET recombination. A single *loxP* site and an *FRT-neo-FRT-loxP* resistance cassette

were simultaneously inserted into the *Cic* genomic locus to flanking exons 2-6 for Cre-mediated excision. The resulting targeting vector was linearized with NotI and electroporated into G4 ES cells. Recombinant ES cell clones were selected in 200 μg/ml hygromycin, expanded and identified by Southern blot analysis (see below). Three out of 60 clones (5%) scored positive for the expected homologous recombination event (Supplemental Fig. S1B,C). Two of these ES cell clones (ESMD23.166 and ESMD23.199) were microinjected into C57BL/6 blastocysts to generate chimeric mice. Male chimeras were crossed with C57BL/6 females to obtain germline transmission of the targeted allele. These mice were subsequently crossed with a Flp deleter strain (Rodriguez et al. 2000) to remove the Neomycin selection cassette flanked by *frt* sequences (Supplemental Fig. S1A). The resulting $Cic^{+/lox}$ mice were routinely genotyped by PCR using the following primers: 5'-gaaagaggcccacttcaag-3' (*Cic*-WT forward); 5'-gggaagagtgggacatgag-3' (*Cic-Δ2-6* forward); 5'-ttatacgaagttatgcggcg-3' (*Cic-lox* forward); 5'-gcaccatagccacgatttaa-3' (common reverse primer); 5'-ttcttctgagcgggactct-3' (*Cic-Neo* reverse).

Southern blot analysis

Genomic DNA was isolated using standard procedures and the following protocols were used for the analysis of each allele: to assess homologous recombination of the Cic^{lox} targeting vector, DNA was digested with KpnI (right arm) or EcoRI (left arm) and probed with a 543 bp DNA fragment corresponding to sequences located 7427 bp upstream of exon 2 (probe A), or with a 528 bp DNA fragment corresponding to sequences located 5779 bp downstream of exon 6 (probe B), respectively. These digestions resulted in diagnostic bands of 13.9 kbp for the wildtype and 11.5 kbp for the recombinant alleles

(probe A), or 11.6 kbp for the wildtype and 6.3 kbp for the recombinant alleles (probe B), respectively (Supplemental Fig. S1A-C).

Transformation assays

Immortal *Cic*$^{+/+}$ and *Cic*$^{\Delta 2\text{-}6/\Delta 2\text{-}6}$ MEFs were stably infected with empty retroviruses, or retroviruses expressing H-Ras^{G12V}, adenoviral E1A or H-Ras^{G12V}-IRES-E1A as described (Drosten et al. 2017). For focus formation assays, 10.000 stably transduced cells were mixed with 4.9x10^5 WT MEFs and seeded on 10 cm plates. Focus formation was scored after 14 days. To assay growth in soft agar, a bottom layer was first prepared by adding 1 ml DMEM containing 10% FBS and 0.75% agar per well in 6-well plates. After that, 10.000 stably transduced cells were added as a top layer in 1.5 ml DMEM containing 10% FBS and 0.3% agar. Colony formation in soft agar was also scored after 14 days.

qRT-PCR

Total cellular RNA was extracted with the RNeasy Mini Kit (QIAGEN) and reverse-transcribed using Super Script II Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) following the manufacturer's instructions. The qRT-PCR assays were performed with a FAST7500 Real-Time PCR System using Power SYBR Green PCR Master Mix (Applied Biosystems) with the primers indicated below. *β-Actin* was used for normalization in mouse cells and *GAPDH* in human cells.

Gene	Species	Technique	Forward	Reverse
Etv1	*mouse*	qRT-PCR	gtaacaggaagatggcggat	ggactccattccaagcctaa
Etv4	*mouse*	qRT-PCR	cgactcagatgtccctggat	gcctgtccaagcaatgaaat
Etv5	*mouse*	qRT-PCR	gggaaatctcgatcagaggac	ggagcatgaagcaccaagtt
Actb	*mouse*	qRT-PCR	gacggccaggtcatcactattg	aggaaggctggaaaagagcc

Kras	*mouse*	qRT-PCR	atgactgagtataaacttgtggtgg	tcacataactgtacaccttgtcc
Cic	*mouse*	qRT-PCR	tttggccgtaaggtgttctc	gaccgtagcctttgctgaag
ETV4	*human*	qRT-PCR	gctcgctgaagctcaggt	tccttcttgatcctggtggt
GAPDH	*human*	qRT-PCR	gttccaatatgattccaccca	gatctcgctcctggaagatg

ChIP assay

ChIP assays were performed as described (Maraver et al. 2012). Briefly, $2x10^7$ Flp-In T-REx 293 cells stably transfected with pcDNA5/FRT/TO alone or pcDNA5/FRT/TO expressing either WT or mutated (Δ2-6 deletion) GFP-tagged human CIC-S were cross-linked for 15 min at room temperature. After washing, cells were sonicated at high intensity during 30 cycles, with 30 s ON and 30 s OFF per cycle (Bioruptor Plus, Diagenode), followed by centrifugation for 15 min at 14,000 rpm at 15°C. For each condition, 200 µg of lysate was incubated overnight with 2 µl of anti-GFP antibody (Abcam, ab290) and immunoprecipitated by incubation with 20 µl of protein A/G beads (Santa Cruz Biotechnology) during 1h at 4°C in a rotating platform. After reverse crosslinking, DNA fragments were recovered by phenol/chloroform extraction and qRT-PCR was carried out in a 7500 Fast Real-Time PCR System (Applied Biosystems) using Power SYBR green PCR Mastermix (Applied Biosystems) with the primers indicated below. A region of the *CDK1* promoter known to lack CIC binding sites was used as negative control.

Gene	**Species**	**Technique**	**Forward**	**Reverse**
ETV1	*human*	ChIP-qRT-PCR	caaccacgtgaccaagaag	gcgctccgctaggagatt
ETV4	*human*	ChIP-qRT-PCR	cttctctcttttctctcggttc	ccaatcagaatgtaggggttg
ETV5	*human*	ChIP-qRT-PCR	aagtgcttcactgactcagctaa	cattggccaatcagcaca
CCND1	*human*	ChIP-qRT-PCR	ttcctgaatgaacgcgctc	ggatggtcggtcaggagaaa
CDK1	*human*	ChIP-qRT-PCR	ggccttcaacgtatgaattagc	agttggtattgcacataagtct

T-ALL Patient Samples and Sequencing Analysis of *CIC*

T-ALL lymphoblast specimens were collected at the time of initial diagnosis from children with T-ALL enrolled on Dana-Farber Cancer Institute protocol 05-001 (https://clinicaltrials.gov/ct2/show/NCT00400946). Samples were collected with informed consent and institutional review board approval. Following Ficoll-Paque purification, genomic DNA was extracted using the DNEasy kit (Qiagen). Targeted exon sequencing for a panel of genes that are recurrently somatically mutated in human cancer, including *CIC*, was performed as described (Sholl et al. 2016). Mutation analysis for single nucleotide variants was performed using MuTect v1.1.4. Analysis for insertions/deletions was performed using the SomaticIndelDetector tool. Analysis for multiple CNVs in the same codon was performed using the GATK ReadBackPhasing Tool. Mutation calls were made for sequence alterations in which the mutation was supported by a minimum of 10 individual sequencing reads, and where the mutation was predicted to result in a non-synonymous alteration of amino acid sequence, or predicted to disrupt a start codon, stop codon, or splice site. Mutations were excluded if they were present at >0.1% frequency in reference populations in the NHLBI GO Exome Sequencing Project ESP (http://evs.gs.washington.edu/EVS/) or in the gnomAD (http://gnomad.broadinstitute.org/) databases.

RNA Sequencing

RNA was extracted using the RNeasy Mini Kit (QIAGEN). 1 μg of total RNA was used to generate the libraries using the TruSeq Stranded mRNA Sample Preparation kit (Illumina). The adapter-ligated library was completed by 8 cycles of PCR with Illumina PE primers. The resulting purified cDNA library was applied to an Illumina flow cell for

cluster generation and single-end sequenced on an Illumina instrument by following manufacturer's guidelines.

RNA-seq data processing

50bp single-end reads were analyzed with the nextpresso pipeline (http://bioinfo.cnio.es/nextpresso/) as follows: The sequencing quality was verified with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the mouse genome (GRCm38/mm10) with TopHat-2.0.10 (Trapnell et al. 2012), using Bowtie 1.0.0 (Langmead et al. 2009) and Samtools 0.1.19 (Li et al. 2009), allowing two mismatches. Read counts for genes were obtained with HTSeq (Anders et al. 2015) and differential expression was calculated with DESeq2 (Love et al. 2014) using the *Mus musculus* GRCm38/mm10 transcript annotations from: https://ccb.jhu.edu/software/tophat/igenomes.shtml.

Gene Set Enrichment Analysis (GSEA)

GSEAPreranked (Subramanian et al. 2005) was run to perform a gene set enrichment analysis of the Notch signaling pathway (Liberzon et al. 2015), and of the published $TLX1^+$ ($HOX11^+$), $LYL1^+$ and $TAL1^+$ gene sets (Ferrando et al. 2002) on mouse data, using the RNAseq gene list ranked by log_2FC obtained with DESeq2 and setting 1000 gene set permutations. For the human samples, standard GSEA was used to explore the enrichment of the Cic LOF signature obtained from mouse data (1000 phenotype permutations). In both cases, we considered only those gene sets with significant enrichment levels (FDR q-value < 0.25).

Heatmaps

All heatmaps presented in this study were built with GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/index.html).

Identification of Cic binding sites in promoter regions

The presence of the Cic binding sites (CBS; Kawamura-Saito et al. 2006) was identified for significantly upregulated genes as follows: genomic coordinates for these genes were first downloaded from the UCSC Genome Browser (Karolchik et al. 2012). For each one of these genes, a promoter region was defined as -3kbps,+500bp from the transcription start site. Each promoter region was analyzed for the presence of the motif (Kawamura-Saito et al. 2006), and when the motif was found, a conservation score was calculated as the mean of the conservation scores of each nucleotide of the motif in that region. Single nucleotide conservation scores, previously calculated with phastCons (Siepel et al. 2005), were obtained from the UCSC Genome Browser (see http://hgdownload.cse.ucsc.edu/goldenPath/mm10/phastCons60way/). We focused on genes with a high degree of CBS motif conservation in the promoter region (conservation score > 0.875) to avoid introducing false positive candidates.

TCR clonality analysis

MiXCR was used to quantify clonotypes from RNAseq reads (Bolotin et al. 2015). Overrepresentation of TCR clonotypes in T-ALL samples relative to a normal T cell sample was calculated following a published method (Brown et al. 2017).

Accession numbers

The data for whole-transcriptome analysis has been deposited in the GEO DataSets under ID code GSE89952.

Supplemental references

- Anders S, Pyl PT, Huber W. 2015. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166-169.
- Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. 2015. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12:** 380-381.
- Brown SD, Hapgood G, Steidl C, Weng AP, Savage KJ, Holt RA. 2017. Defining the clonality of peripheral T cell lymphomas using RNA-seq. *Bioinformatics* **33:** 1111-1115.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.
- Karolchik D, Hinrichs AS, Kent WJ. 2012. The UCSC Genome Browser. Curr. Protoc. *Bioinformatics* **Chapter 1:** Unit 1.4.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SMAtools. *Bioinformatics* **25:** 2078-2079.
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov M, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1:** 417-425.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.
- Maraver A, Fernandez-Marcos PJ, Herranz D, Cañamero M, Muñoz-Martin M, Gómez-López G, Mulero F, Megías D, Sanchez-Cabayo M, Shen J, et al. 2012. Therapeutic effect of g-secretase inhibition in KrasG12V-driven non-small cell

lung carcinoma by derepression of DUSP1 and inhibition of ERK. *Cancer Cell* **22:** 222-234.

- Rodriguez CI, Buchholz F, Galloway J, Sequerra R, Kasper J, Ayala R, Stewart AF, Dymecki SM. 2000. High-efficiency deleter mice show that FLPe in an alternative to Cre-loxP. *Nat Genet* **25:** 139-140.
- Scholl LM, Do K, Shivdasani P, Cerami E, Dubuc AM, Kuo FC, Garcia EP, Jia Y, Davineni P, Abo RP, et al. 2016. Institutional implementation of clinical tumor profiling in an unselected cancer population. *JCI Insight* **1:** e87062.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionary conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15:** 1034-1050.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545-15550.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562-578.