ROQUIN-1 binding to the transcriptome of T cells reveals conserved targets in the control of lymphopoiesis

Vigo Heissmeyer

vigo.heissmeyer@med.uni-muenchen.de

Ludwig-Maximilians-University of Munich https://orcid.org/0000-0002-2263-8545

Elaine Wong

Ludwig-Maximilians-University

Giulia Cantini

Helmholtz Munich

Gesine Behrens

Ludwig-Maximilians-Universität München

Timsse Raj

Ludwig Maximilian University of Munich https://orcid.org/0000-0002-5065-8092

Taku Ito-Kureha

The University of Tokyo https://orcid.org/0000-0002-6239-9227

Lambert Moyon

Helmholtz Zentrum München

Chenge Xin

Department of Histoembryology, Genetics and Developmental Biology, Shanghai Key Laboratory of Reproductive Medicine, Shanghai Jiao Tong University School of Medicine

Kai Hoefig

Helmholtz Munich https://orcid.org/0000-0002-0701-454X

Meng Xu

Helmholtz Munich

Lisa Kifinger

Ludwig-Maximilians-Universität https://orcid.org/0000-0001-6422-057X

Arlinda Negraschus

Ludwig Maximilians University of Munich

Nina Kronbeck

Ludwig-Maximilians-University Munich

Elizaveta Tkacheva

Ludwig-Maximilians-University Munich

Silvia Monticelli
Institute for Research in Biomedicine https://orcid.org/0000-0002-5909-8802
Jana Schor
Young Investigators Group Bioinformatics and Transcriptomics, Department Molecular Systems Biology, Helmholtz Center for Environmental Research – UFZ, Permoserstraße 15, 04318 Leipzig, G
Gergely Csaba
Ludwig-Maximilians-University of Munich
Ralf Zimmer
Institute of Bioinformatics and Practical Informatics, Department of Informatics, Ludwig-Maximilians-Universität München (LMU), Munich, Germany
Marcin Lyszkiewicz
Universität Ulm
Annalisa Marsico
Helmholtz Zentrum München

Article

Keywords: RNA-binding proteins, Roquin, transcriptome-wide interactions, inhibitor of differentiation, lymphocyte development

Posted Date: July 15th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6931797/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

Defining the RNA targets of ROQUIN-1 is essential for understanding post-transcriptional gene regulation in immunity, autoimmunity, and cancer immunotherapy. We systematically mapped transcriptome-wide ROQUIN-1/RNA interactions in mouse and human lymphocytes. Novel targets were enriched in signaling molecules, transcription factors, epigenetic regulators, and mediators of immune cell differentiation. Cross-species analysis revealed conservation at binding sites, suggesting evolutionary preservation of ROQUIN-mediated regulation. Notably, ROQUIN-1 bound all mRNAs of the Id1–Id4 gene family—transcriptional regulators critical for lymphocyte differentiation. Functional perturbation through Id1-3 overexpression impaired thymocyte and B cell development. Similarly, inactivation of Roquin-1/2 increased Id3 expression in thymocytes and developing B cells, delayed double-negative thymocyte progression and blocked B cell development before the pro-B cell stage. Uncovering Roquin's role in controlling gene expression driving lymphocyte development, our dataset also provides a valuable resource for functional genomics to dissect RNA-based regulation in immunity and offers mechanistic insights into ROQUIN function across species.

Introduction

The RNA-binding proteins (RBPs) Roquin-1 and Roquin-2, encoded by the essential genes *Rc3h1* and *Rc3h2*, are critical regulators of immune homeostasis[1, 2, 3]. Loss-of-function or hypomorphic mutations in mice and humans are associated with spontaneous T cell activation, cytokine overproduction, autoimmunity and autoinflammatory syndromes[2, 3, 4]. Conversely, inactivation of Roquin encoding genes in tumor-specific $CD8^+$ T cells enhances antitumor immunity by promoting cytotoxicity, reprogramming of the cellular metabolism and resistance to exhaustion[5, 6, 7, 8]. In T cells, Roquin-1 and Roquin-2 function redundantly to control fate decisions, inducing decay of mRNA targets through recognition of defined stem-loop structures and additional interactions with effectors of mRNA deadenylation and decapping[3, 9, 10, 11]. Roquin proteins are substrates of the Malt1 paracaspase, a component of the CBM (Carma1–Bcl10–Malt1) complex[12]. Malt1 functions as a scaffold to activate NF-κB on the one hand but also affects immune activation through cleavage of target proteins on the other. Constitutive activation of Malt1 protease activity causes lethal inflammation in mice[13], which can be rescued by mutation of the Malt1-cleavage sites in Roquin-1[14], underscoring the physiological importance of this regulatory axis. Deletion of Roquin genes disrupts quiescence and skews T cell differentiation, promoting T_H1, T_H17, or T follicular helper (T_{FH}) phenotypes, while converting regulatory T (T_{reg}) cells to follicular regulatory T (T_{FR}) cells[3, 5, 12, 15, 16]. The phenotypes of mice with B-cell-specific combined ablation of Roquin encoding alleles have not been reported so-far. In fact, homozygous hypomorphic mutations in Roquin-1 have been associated with lupus-like autoimmunity or with a hemophagocytic lymphohistiocytosis syndrome in mouse or man, respectively[2, 4, 5].

Despite our considerable knowledge of the physiologic importance of Roquin proteins, we do not know the gene regulation that underlies these phenotypes. There is little information about site-specific

binding of Roquin on established target mRNAs and we do not understand how Roquin's interaction with *cis*-elements encoded in mRNAs controls the expression of targets. This has only been formally established for the recognition of the constitutive decay element (CDE) or alternative decay element (ADE), stem-loop structures that occur in the 3´-UTR *cis*-elements of *Nfkbid*, *Tnf* and *Tnfrsf4/*Ox40, where recognition was associated with Roquin-induced deadenylation, decapping or translation inhibition of targets[4, 9, 10, 17]. Although it is likely that the deregulation of specific targets drives changes in Roquin-deficient T and B lymphocytes and ultimately causes autoimmunity and pathology, the identity of these genes is unknown, and it is unclear how they contribute to the observed phenotypes. These questions were also not answered by previous crosslinking and immunoprecipitation approaches, performed exclusively in non-lymphoid cells[9, 18, 19]. To address this gap, we performed high-resolution mapping of endogenous Roquin-1 binding sites across the transcriptomes of primary $CD4^+$, $CD8^+$ T and B lymphocytes. The novel targets reveal a critical involvement of Roquin proteins in lymphopoiesis and our datasets provide a foundational resource for understanding the post-transcriptional programs regulated by Roquin and how their disruption drives immune pathology.

Results

Transcriptome-wide analysis of Roquin-1 binding across lymphocytes.

To identify targets and binding sites for the endogenous Roquin-1 protein in the transcriptome of lymphocytes, we used the iCLIP2 protocol on *in vitro* expanded mouse T_H1 cells[20]. Characteristic truncations are detected after cDNA sequencing, aligning the 5' end of the fragments to the first nucleotide downstream of the crosslink, since peptide-adducts block the reverse transcription defining binding sites at near-nucleotide resolution[21]. The same procedure was performed in parallel on control extracts of 4´-hydroxy-tamoxifen treated *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd4*-Cre-ERt2 T cells (named iDKO) (Fig. S1A). We produced three technical replicates within batch 1 and in the independent batch 2 (Fig. 1A and Table S1), for which we also generated size-matched input control samples. We defined Roquin-1 binding sites in a multi-step process: i) Analyzing the two batches separately and processing the raw sequencing data into reads mapped to the reference genome. ii) Calling crosslink sites with the peak caller PureCLIP[22] on merged reads from all technical replicates in each batch and refining them with a previously established pipeline[23]. iii) Applying a filter on the called peaks by computing the ratio of read coverage (Cov) in wildtype (WT) and knockout (iDKO) samples at each putative binding site and discarding all sites not surpassing a defined threshold (see *Methods*) (Fig. 1A). Comparing the independent batches before and after iDKO correction, we demonstrate the efficacy of this procedure in eliminating binding noise, present as intronic enrichment before applying the filter (**Fig. 1B**). To compile a reproducible list of binding sites, we unified the iDKO-corrected binding sites from batch 1 and batch 2 by measuring the support of binding sites in one batch based on an increased read coverage at corresponding positions in the other (**Fig. 1A**). The final set of binding sites was reproducible across

batches, as highlighted by the strong and significant correlation of crosslink coverage (**Fig. 1C**). In total we determined 19,759 binding sites of Roquin-1 in the transcriptome of $CD4^+$ T cells that occurred in 2,136 genes. We found binding to 2,118 mRNAs in $CD4^+$ T cells overlapping with the binding in $CD8^+$ as well as in B cells (**Table S2** and Fig. S2A). The binding in $CD4^+$ T cells was strongly enriched in the 3′-UTR of mRNAs (64%, 12565 sites in 1,718 genes) with few interactions observed in the 5′-UTRs (2%) and some binding in the coding sequence (18%) or introns (16%) (**Fig. 1D**).

We inspected the Roquin-1 binding sites for sequence and structural determinants (**Fig. 1E** and **1F**). To do so, we compiled a set of Roquin-1-bound sequences as windows of 100 nucleotides around binding sites and a set of sequences from peaks called on the Input control sequences, to filter out unspecific patterns. Involving sequence and sequence-structure deep learning models for RBP binding prediction, we applied pysster[24], which classifies bound vs. unbound sequences by learning relevant motifs and structural features. Given Roquin's preference for structured RNA elements, we first trained pysster to distinguish Roquin-1-bound sequences from controls using only predicted RNA structure as input to the model. We found an enrichment in the bound sequences of seven structural patterns harboring stem-loops with 3 to 5 nucleotide-long loops and stems of variable length (**Fig. S1B**), in agreement with previously reported Roquin-bound elements[9, 18, 19]. Next, we used the seven structural patterns identified by pysster to re-scan Roquin-bound sequences, mapping back their genomic positions and searching for enriched sequence preferences within these regions using RSAT[25]. This structure-guided approach allowed us to inspect sequence patterns at stem-loops bound by Roquin-1. We found ten putative sequence motifs with RSAT all significantly enriched in the test set in comparison to the control background. Among them, M0004, showed strongest enrichment within true peaks (26%), and was verified against RBP motif databases to match Roquin-1 (*Rc3h1*). This motif (CTNTG) resembled the core CDE sequence tri-loop sequence with a closing C-G base-pair (CTGTG) (**Fig. S1C**). In line with the previous observation showing unchanged affinity for a tri-loop harboring a central mutation (CTCTG)[26], our data suggests that Roquin-1 accepts also other bases than purine at the second position of the tri-loop *in vivo* (TNT instead of pyrimidine-purine-pyrimidine (YRY)). We then investigated the co-occurrence of sequence and structural patterns. A subset of four tri-loops - K11, 23, 39 and a tetraloop K32- with a subset of sequence motifs was selected (**Fig. 1F**) from the full set of pairs (**Fig. S1D**). Some sequences (M0001) show no or few structural enrichments, possibly matching linear sequences or stems of the hairpins (pair M0002 - K23), some align with the hairpin (M0004 – K32) and less defined long T-rich stretches overlapped with whole loops (M0007 – K11). Together, our results suggest that a sequence consensus motif other than the CDE or U-rich loops[9, 18] might be lacking, and that Roquin-1 may interact with *cis*-elements through additional structured elements of variable sequence content.

We also inspected binding sites with the established regulation of known targets like the 3′-UTR of *Nfkbid* (Fig. 1G) or *Tnfrsf4* (Fig. S2B). In fact, iDKO-corrected and batch-unified binding sites overlapped with the mapped Roquin-response element (RRE) and were positioned within or close to the predicted stem-loops (SLs), that are required or contribute to Roquin-mediated post-transcriptional regulation of these 3′-UTRs[9, 17]. Using *in vitro* expanded $CD8^+$ T cells, *in vitro* expanded $CD4^+$ T cells skewed into

different T cell subsets (T_H0, T_H17 and iT_{reg}) or *in vitro*-derived germinal center B cells[27] confirmed reproducibility of the iCLIP2 procedure by displaying highly consistent Roquin-1 binding in the *Nfkbid* 3′-UTR across different lymphocytes. The only exception was the first truncation site in the 3′-UTR, which coincided with the previously mapped SL1 of the *cis*-element, which was strongly bound in T_H1, T_H0, T_H17 and iT_{reg} cells, but showed much less or no binding in $CD8^+$ T or B cells (Fig. 1G). We transduced GFP expressing reporters containing wildtype or a stem loop 1-mutated version (mutSL1) of the RRE from the *Nfkbid* 3′-UTR into different lymphocytes from *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;Cre-ERt2 mice. Compared to corresponding wild-type cells, the GFP expression revealed a strong derepression for the WT reporter in all 4′-hydroxy-tamoxifen-treated iDKO samples (Fig. 1H, compare black and red histograms in top row). The WT and mutSL1 3′-UTRs showed very similar GFP expression in iDKO cells of the different lymphocytes (Fig. 1H, compare red histograms in upper with lower panels, and Fig. S2C). In contrast, WT cells revealed that the SL1 mutation derepressed GFP much stronger in $CD4^+$ and much less in $CD8^+$ and B cells (Fig. 1H, compare black histograms in upper and lower panels). This indicates that Roquin-mediated repression is more dependent on its binding to SL1 in $CD4^+$ T cells, correlating with stronger peaks at this position in the $CD4^+$ subsets as compared to $CD8^+$ or B lymphocytes (Fig. 1G). Together these data validate site-specific iCLIP2 interactions of Roquin-1 on a well-established target and *cis*-element and suggest differential accessibility as well as regulatory contributions of individual binding sites in different lymphocyte lineages and subsets.

Roquin-1-bound 3′-UTRs are repressed at different extents

We then performed a pathway enrichment analysis on the top 10% of 3′-UTR targets, ranked by binding score using a predefined subset of GO[28, 29] and KEGG[30] terms relevant for T lymphocytes[31]. We find significant enrichment in genes associated with the pathways of TGFb, mTOR and TCR signaling or with the functional categories of leukocyte transendothelial migration and unfolded protein response as well as others (Fig. 2A). The highest enrichment of genes and significance was found for T cell differentiation, underscoring previously observed phenotypes in T_H1, T_H17, T_{FH} and T_{FR} differentiation[2, 3, 4, 5, 12, 14, 15, 32].

We also performed mRNA sequencing of $CD4^+$ T cells from WT (*Cd4*-Cre-ERt2) or iDKO (*Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd4*-Cre-ERt2) mice. The cells were isolated, treated with 4′-hydroxy-tamoxifen and cultured under T_H1 conditions. RNA was extracted before or after 1h, 2h or 4h of anti-CD3/anti-CD28 restimulation (Fig. S3A). In the unstimulated samples we find 1780 differentially expressed genes (DEGs) in iDKO compared to WT T cells, with a smaller fraction of them (339 genes or 19%) being directly bound by Roquin-1 as determined by iCLIP2 (Fig. 2B). Upon stimulation of WT or iDKO T cells several thousands of genes became differentially expressed compared to WT 0h. T cell activation increased the number but not the percentage of direct Roquin-1 targets within the DEGs of iDKO T cells. These data indicate that large numbers of genes are regulated in an indirect manner upon genetic inactivation of

Roquin and contribute to the observed phenotypes. Analyzing the relative expression of all Roquin-1-bound mRNAs (Table S2) in unstimulated iDKO compared to WT T cells, we find a fraction that was not regulated and among the deregulated (>2-fold, -log10 p-value> 0.05) genes (red dots) we find only 3 mRNAs downregulated (Fig. 2C). Consistent with the function of Roquin proteins as repressors, the majority of deregulated target genes exhibited increased expression. Among these were the previously proposed Roquin targets *Zc3h12a*, *Icos*, *Tnfrsf4 (Ox40)*, *Ctla4*, *Nfkbid* and *Irf4*[5, 9, 10, 12, 33] but also the ligand *Tnfsf14*/Light or the receptor *Tnfrsf25*/D3, Unc119 as well as the transcription factor Myb or the proprotein convertase furin as well as its target *Lfng/*Lunatic fringe that operate in the Notch pathway and play a role in neurogenesis (Fig. 2C). Given the extensive indirect regulation of non-targets in Roquin-deficient cells (Fig. 2B) as well as our observation that a large fraction of Roquin-1 bound targets are not deregulated upon inactivation of Roquin (Fig. 2C), we addressed direct 3′-UTR–mediated regulation by cloning a large set (>60) of 3′-UTRs, inserting them 3′to the GFP coding sequence into a retroviral vector. We analyzed regulation of the reporter by endogenous Roquin levels in a MEF cell line that allows 4´-hydroxy-tamoxifen-induced deletion of Roquin-1/2 encoding alleles[9] (Fig. S3B). In this screen (Fig. 2D and E) we tested Roquin-1 targets with strongest upregulation on the mRNA level (*Pim2* and *Polr1c*; Fig. 2C), several known or proposed targets (*Nfkbid*, *Zc3h12a*, *Irf4*, *Ctla4*), we included *Ccr2* as a target, which showed downregulation upon inactivation of Roquin (Fig. 2C), a selection of bound mRNAs with high truncation sites scores as well as high read coverage and a negative control lacking a 3′-UTR (noUTR). We also included the *Ifng* and *Il2* 3′-UTRs. *Ifng* was previously proposed to be a Roquin target, and was shown to contribute to the lupus-like phenotype of *sanroque* mice[34]. However, the 3′-UTR of *Ifng* lacked significant binding or truncation sites in our analysis (Fig. S3C). In contrast, the 3'-UTR of *Il2*, a previously proposed Regnase-1 target[35], showed specific binding by Roquin-1 (Fig. S3D). Of note, *Tnf*, the prototypic Roquin target, reavealed strong 3′-UTR binding (Fig. S3E). The screening results revealed no targets that underwent downregulation (Fig. 2E). About half of the tested targets showed more than 1.2-fold upregulation upon inactivation of Roquin, with the majority showing less than 2-fold induction. Only a small subset of 3′-UTRs exhibited a very profound, more than 2 and up to 16-fold induction. Among the strongest Roquin-regulated targets were *Nfkbid*, *Nfkbiz* and *Zc3h12a* as well as a novel targets *Map3k8* (aka Tpl2 or Cot) and *Lfng (*Fig. 2D and 2E*)*. We confirmed Roquin-regulated expression of *Lfng* and *Map3k8* genes on the protein and RNA levels (Fig. S3F and S3G). Comparing parameters of Roquin-1 binding to 3′-UTRs with the observed regulation in our test set, we determined a moderate positive correlation between the PureCLIP binding score (BS) and fold regulation of the reporter (Fig. 2F), while the correlation only with the crosslink counts was negligible (Fig. S3H). The strongest regulated genes (*Map3k8*, *Nfkbiz*, *Zc3h12a* and *Nfkbid*) have in common that their mRNAs are subject to profound translational inhibition[9, 36, 37]. Of note, the negative control or the *Ifng* 3′-UTR did not respond to the inactivation of Roquin. We hypothesized that Roquin-1-bound 3′-UTRs that showed little regulation by endogenous Roquin expression levels may nevertheless respond to Roquin overexpression. In fact, doxycyclin-induced expression of Roquin-1 decreased the expression of 4 selected 3′-UTRs linked to *Gfp* mRNA that showed almost no induction upon inactivation of Roquin encoding alleles (Fig. S3I). These data indicate that Roquin represses the expression of its targets through different post-transcriptional

programs and to different extents depending at least in part on the level of expression of Roquin itself. Many targets of Roquin are involved in T cell differentiation and activation including signal transducers (*Il2*, *Ptger4*, *Map3k8, Tnfrsf4* etc.), regulators of metabolism (*Polr1c*, *Mat2a*, *Mthfd2*, etc.), other RBPs (*Zc3h12a/d*, *Rbm39*, *Hnrnpl*, *Hnrnpa1*) transcription factors (*Nfkbid*, *Hivep2 etc.*) or epigenetic modifiers (*Tet3* etc.) that may contribute to the indirect effects associated with Roquin loss-of-function.

Roquin-1 binds to orthologous regions in human and mouse transcripts

We extended our iCLIP2 analyses to include human $CD4^+$ and $CD8^+$ T cells (**Tables S1** and **S2**), generating two biological replicates for $CD4^+$ T cells, each with two technical replicates, and two technical replicates for $CD8^+$ T cells. Each experiment was performed alongside its respective Input control. We applied the same computational pipeline used for the mouse data, except for iDKO filtering. In $CD4^+$ T cells, we identified 33,492 binding sites across 2,969 target genes, overlapping with 2,492 transcripts also found in $CD8^+$ T cells (**Table S2**). Sequence motif analysis of these sites revealed 19 enriched sequences, with a subset resembling those observed in the mouse (**Fig. S4A**, highlighted in blue, **Table S3**). Focusing on 3′-UTR binding sites, we assessed the overlap between the human and mouse ROQUIN-1–binding sites in $CD4^+$ T cells—as well as their conservation. We used UCSC LiftOver[38] to map mouse binding sites to the human genome and vice versa. Only 685 binding sites from the mouse and 1632 binding sites from the human datasets could not be aligned. This approach allowed us to identify both species-specific or orthologous sites where binding occurred in orthologous genes at matching or non-matching genomic coordinates (**Fig. 3A** and **Table S4).** We identified 3,169 mouse-specific binding sites and 9,674 sites unique to human. The majority of binding events occurred in orthologous genes: in mouse, 6,237 sites mapped to different and 3,159 to matching coordinates; in human, 9,422 mapped to non-matching and 3,242 to identical coordinates (**Fig. 3B)**. Considering these 3′-UTR targets, 59% of mouse and 40% of human targets overlapped, respectively (**Fig. 3C)**. To further inspect the degree of conservation of ROQUIN-1 binding sites from mouse to human, we first investigated sequence conservation at binding sites with PhyloP scores. Overall, ROQUIN-1 binding sites show higher sequence conservation than background, suggesting constraint related to their functional importance (**Fig. 3D**). The degree of constraint differed across the subsets of alignment with human binding sites. Binding sites at matched genomic positions between orthologous genes in human and mouse showed the highest sequence conservation—significantly higher than random control sequences and Roquin-1 mouse peaks in 3′-UTRs. Less sequence conservation was observed for sites in orthologous genes with different coordinates (**Fig. 3D**). We also evaluated structural conservation of mouse binding sites using LocARNA structural alignment[39] across 60 vertebrates, where higher base-pairing support scores indicate greater structural constraint (**Table S5**). Base-pairing positions within binding sites exhibited significantly higher structural conservation than other base-pairing positions in local control regions (**Fig. 3E**, left panel), independently from their level of sequence conservation (**Fig. 3E**, right panel).

We inspected species-specific interactions of ROQUIN-1 with the mRNA *of ZC3H12A/REGNASE-1*, a shared target of Regnase-1 and Roquin-1, for which ROQUIN-1 binding sites are elusive. We projected a consensus secondary structure of 26 eutheria species (**Table S5**) based on LocARNA and annotated mouse (purple) and human (blue) PureCLIP binding sites (**Fig. 3F**). The predicted secondary structure displayed 10 stem-loops, of which the 9 most 3′-located SLs contained conserved, compatible or covariant bases within base-paired sequences. ROQUIN-1 bound to 9 stem-loops (SL2-9) with crosslinks mostly at sequences that are not base-paired, and there was a very good overlap between mouse and human binding sites. Importantly, SL4 and SL8 are well supported by structural conservation, and have been shown to be critical for Regnase-1–mediated regulation[40, 41]. Importantly, combined neutralization of their secondary structures by antisense oligonucleotides (ASO) has been shown to dampen experimental autoimmune encephalomyelitis (EAE) in mice[41], for which our data now involve a contribution by Roquin.

Together, our comparative iCLIP2 analysis shows a large overlap of target 3′-UTRs and extensive conservation of binding sites in mouse and human, implying a conserved gene regulatory program of ROQUIN-1 in both species. Site-specific binding to the *ZC3H12A* 3′-UTRs in mouse and human provides additional support for the conserved ROQUIN-1/RNA interaction.

Roquin-1 binds to and regulates transcripts of the Id family

In addition to orthologous binding sites, we noticed that several Roquin-1 targets including *Nfkbid*/*Nfkbiz*, *Zc3h12a*/*Zc3h12d*, *Ier3*/*Ier5*, *Socs1*/*Socs2*/*Socs3* and *Id1*/*Id2*/*Id3*/*Id4* are paralogs, suggesting regulatory relationships, since these gene duplications occurred. We focused on the Inhibitor of DNA binding (Id) proteins that heterodimerize with E-box–binding transcription factors and thereby prevent their binding to DNA[42]. The E-box binding transcription factors control a range of cell fate decisions during lymphocyte differentiation, and Id proteins determine their activity through a tightly balanced stoichiometry[43]. The 3′-UTRs of *Id1*, *Id2* and *Id3* mRNAs showed specific and strong Roquin-1 binding in lymphocytes (Fig. 4A), while *Id4* was not detected in lymphocytes. We cloned and tested the 3′-UTR in the context of GFP reporters in MEF cells after inducible deletion of Roquin-1 encoding alleles (Fig. 4B and Fig. S3B). All three 3′-UTRs showed upregulation upon inactivation of Roquin (Fig. 4B). Interestingly, although the *Id4* mRNA was not detected in lymphocytes, it showed binding of Roquin-1 in MEF cells with somewhat lower read density. Its 3′-UTR nevertheless conferred moderate Roquin-mediated repression in reporter assays (Fig. S5A-B). We used shorter regions harboring accumulated Roquin-1 binding sites and high read density as 3′-UTR reporters. Since these shorter fragments (Fig. 4B, *Id1* nt 355-563; *Id2* nt 408-745 and *Id3* nt 257-546) retained regulation, we propose that they contain the respective Roquin-response elements (Fig. 4B). Since we also detected binding of human ROQUIN-1 in the 3′-UTRs of the T cell-expressed mRNAs of ID2 and ID3 (Fig. S5C-D), we utilized LocARNA analyses to determine structural conservation within the response elements across 112 vertebrate species (Fig. S5E-G and Table 5). We then tested the effect of point mutagenesis designed to disrupt individual stem-loops within the predicted secondary structures. Testing different individual stem-loops per 3′-UTR, we found

that all three Roquin-response elements contained at least one stem-loop (SL), with a UAU tri-loop configuration that was critical for Roquin-mediated regulation of Id1 (SL4), Id2 (SL8) and Id3 (SL4) (Fig. 4B and 4C and Fig. S5E-G). Of note, other stem-loops also contributed to the observed Roquin-mediated repression (i.e SL5 in Id1 and SL2 in Id3) (Fig. 4B and C). Investigating physiologic Roquin-dependent regulation of Id1-3 in MEFs or $CD4^+$ T cells we found increased expression of Id factors in both iDKO MEFs and T cells. Derepression of Id1 and Id3 was prominent at both transcript and protein level, while the derepression of Id2 in T cells after Roquin inactivation was more pronounced at the protein level (Fig. 4D-H). Together, these data show binding and regulation of the mRNAs of all Id1-4 paralogs, again suggesting strong structural conservation within specific Roquin-recognized *cis*-elements.

Increased Id protein expression interferes with lymphocyte development

We then asked whether upregulation of individual Id proteins, which may arise in the context of Roquin-loss-of-function can alter fate decisions in hematopoietic progenitors. We ectopically overexpressed three of the Id proteins (Id1, Id2 and Id3) or a control (empty vector) in wild-type CLP cells. Such transduced cells were co-cultured with stromal OP9-DL1 cells in the presence of SCF, IL-7 and Flt3L to promote T lineage commitment[44, 45]. Expression of any of the tested Id factors impaired T cell development beyond the $CD25^+$ $CD117^+$ ($CD44^+$) DN2 stage (Fig. 5A and 5B, Fig. S6A). Similarly, Id1, Id2 and Id3 or control vector transduced LSK cells were co-cultured with OP9 cells to assess B cell, myeloid and NK commitment. Here, each of the tested Id factors strongly inhibited the development of the B lineage ($CD19^+$), whereas the development of both NK ($NK1.1^+$) and myeloid cells ($CD11b^+$) was less affected (Fig. 5C and 5D).

To address potential Roquin-mediated regulation of Id proteins during hematopoietic lineage choice, we took advantage of 4´-hydroxy-tamoxifen-inducible deletion, where we inactivated Roquin-1 and Roquin-2 encoding alleles in CLP cells and subsequently co-cultured them in an analogous manner with OP9 cells. Similar to ectopic expression of Id1, Id2 or Id3, conditional deletion of Roquin-1 and Roquin-2 encoding alleles resulted in an almost complete block in the B cell lineage, with lesser effects on NK and myeloid development (Fig. 5E). Since CLP progenitors retain both B and T lineage potential, it was possible to simultaneously assess how conditional deletion of Roquin-1 and Roquin-2 affects their lymphocyte commitment *in vitro*. We took advantage of a 2:1 mixed OP9:OP9-DL1 co-culture system that allows for simultaneous B and T lineage commitment and tested for the presence of $CD19^+$ B cell or DN2-DN4 T cell committed progenitors. While CLPs with from mice harboring either floxed alleles or the Cre-ERt2 transgene retained the capacity to commit to B and T lineages, conditional deletion of Roquin-1/2 resulted in a 20-fold reduction in B cell frequency (Fig. S6 B-D, day 5). A fraction of B-cell-committed progenitors was initially detectable on day 3 but appeared strongly reduced on days 5 and 7 of coculture. At the same time, deletion of Roquin encoding alleles had marginal effects on T cell progenitor

commitment and expansion (Fig. S6D). Together these data show a cell-type intrinsic impairment in lymphopoiesis of progenitor cells that overexpress Id or lack Roquin protein expression.

Impaired thymocyte development upon inactivation of Roquin

Since in vitro culture conditions are only able to determine effects on early (DN1-3) thymocyte differentiation, we determined the effect of Roquin inactivation on lymphopoiesis *in viv*o. To this end, we combined the floxed alleles encoding for Roquin-1 and Roquin-2 with the *Cd127*-Cre mouse line (*Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre), active in CLPs[46]. Of note, these mice showed pronounced activation of lymphocytes in the periphery, similar to *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;CD4cre mice and an even stronger genetic burden, likely the result of the previously described autoinflammation as reported before[3, 12, 15]. In these mice we analyzed early thymocyte development which, at several different stages, is placed under the control of ID and E-box binding proteins[47]. Flow cytometry with CD4 and CD8 markers revealed a reduction in single positive SP4 and SP8 thymocytes, suggesting reduced thymic output of mature T cells (Fig. 6A). Discriminating double-negative differentiation with CD44 and CD25 markers and analyzing Roquin expression by RT-qPCR in WT mice, we found that *Rc3h2* (Roquin-2) transcripts were especially increased at the DN2/DN3 stages (Fig. 6B). Consistent with a role of Roquin proteins in modulating these stages of development, *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice showed increased DN3 and decreased DN4 thymocyte frequencies (Fig. 6C). A known target of E2A is Rag1[48], we therefore analyzed thymocyte extracts with immunoblots and showed a decrease of Rag1 protein expression in Roquin-deficient thymocytes (Fig. 6D). Pre-gating on DN3/DN4 cells allowed us to further discriminate the substages DN3a-c and DN4a-c according to the expression of CD28 and CD25 markers, showing a significant decrease in the frequency of DN4a cells (Fig. 6E). Of note, most likely due to systemic inflammation in *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice, the sizes of the thymi were smaller causing a general reduction in the absolute numbers of thymocytes at all stages (Fig. S7A-D). The strong expansion that occurs at the DN4a stage is driven by cell autonomous pre-TCR signaling[49], which in turn requires E2A and HEB transcriptional activity to induce preTCRa gene transcription[50]. We hypothesized that decreased Roquin activity in DN4 cells could cause increased ID3 expression that inhibits E-box protein induced pre-TCRa expression. Indeed, measuring pre-TCRa expression by flow cytometry, we found that Roquin-deficient DN4 thymocytes did not efficiently induce the expression of the pre-TCRa chain (Fig. 6F). Employing tamoxifen-treatment of *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;Cre-ERT2 mice enabled us to sort DN4 cells with acute deletion and revealed that preTCRa mRNA expression was reduced in DN4 thymocytes lacking Roquin, which showed an increase in *Id3* mRNA expression (Fig. 6G). These findings suggest that thymocyte development is shaped at multiple stages through Roquin-mediated regulation of Id proteins and E-box binding transcription factors.

Inactivation of Roquin blocks B cell development at the pre-pro-B cell stage

Given the impact of Roquin-mediated regulation of Id proteins, we analyzed B cell development in the bone marrow of *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice via 26-parameter multicolor, spectral flow cytometry and UMAP representation. We discriminated at least 9 subsets of cell types (Fig. 7A). Potentially due to the systemic inflammation in *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice, we also observed reductions in nonlymphoid cells including the fraction of red blood cells (Fr. 4), myeloid precursors (Fr. 9), CD11b^{+} (Fr. 8) and CD11c^{+} dendritic cells (Fr.3) as compared to WT counterparts (Fig. 7B). In the lymphocyte compartment, NK1.1+ cells (Fr.1) were reduced, particularly NK cells, and to a lesser extent NKT cells (Fig. S7D). CD4+ T lymphocytes (Fr. 7) were also reduced, while CD8+ T cells (Fr. 2) were increased and g/d T cells (Fr.6) were unaffected (Fig. 7B and Fig. S7D). The most striking phenotype was the complete absence of CD19^{+} B cells (Fr. 5) (Fig. 7B and Fig. S7D). We recapitulated developmental progression into mature and immature B cells using B220 and IgM surface expression, and found a complete absence of mature, transitional, immature and pre-B cells, with a small fraction and low absolute numbers remaining in the pro-B cell gate (Fig. 7C and Fig. S7E). In WT mice the Roquin-1/2 encoding mRNAs of *Rc3h1* and *Rc3h2* increased during developmental progression from the pre-pro-B cell to pro- and pre-B cell stages (Fig. 7D). We then pre-gated on lin^{-}; CD43^{+} B220^{+} cells and discriminated CD19^{-} pre-pro B and CD19^{+} pro-B cells in *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice. The results showed an almost complete absence of pro-B cells, and the few detected pro-B cells showed much lower CD19 expression levels (Fig. 7E). Tamoxifen-induced deletion of Roquin-encoding alleles in *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;CreERt2 mice enabled us to analyze acute effects of Roquin inactivation in B cell precursors that would not develop upon *Cd127*-Cre–mediated deletion. Coinciding with the complete block of commitment to the B cell lineage in *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice, we found significantly increased *Id3* and *Id1* mRNA expression upon tamoxifen-induced deletion of Roquin encoding alleles in pre-pro-B cells (Fig. 7F and 7G), consistent with previous reports showing a block of B cell development at the pre-pro-B cell stage upon E2A inactivation[51, 52, 53] or Id1 or Id3 overexpression[54, 55, 56]. Together, these data reveal that Roquin post-transcriptionally enables the commitment to the B cell lineage, which can be explained through negative regulation of Id1 and Id3 leading to increased E2A activity.

Overall, we found that deletion of Roquin encoding alleles led to a block in T and B cell development at the pre-pro B and DN4 stage, respectively (Fig. 7H), and that this was mediated at least in part through the Roquin-dependent regulation of Id proteins.

Discussion

This study identifies ROQUIN proteins as a key post-transcriptional regulators of lymphocyte development by providing a comprehensive map of ROQUIN-1 mRNA targets and binding sites across murine and human T and B lymphocytes. Our data highlight the critical importance of understanding

ROQUIN-1's direct RNA interactions for several reasons: First, site-specific binding of ROQUIN-1 to the 3'-UTRs of *Nfkbid*, *Zc3h12a*, and *Tnfrsf4* establishes the molecular basis for their previously described regulation[9, 12, 17, 36]. We extend this analysis now to the family of Id proteins including *Id1*, *Id2*, and *Id3*, identifying distinct Roquin-response elements featuring stem-loop structures essential for Roquin-mediated control. Among the 3'-UTR targets, we observe varying levels of regulatory strength, with increased regulation correlating with higher binding site scores. These data show that different *cis*-elements encode differential regulation. The reason for the difference in regulatory strength is currently unknown but may involve degrees of cooperative interactions with other RNA-binding proteins (RBPs)[57], cooperativity of stem-loop folding[58] or involve different interactions with specific components of the deadenylation, decapping, or translational machinery[9] for subsets of these targets. Cooperative ROQUIN binding has not yet been demonstrated. Based on the observed crosslink patterns of Roquin-1 in the *cis*-elements investigated in this study, we envision that upon interaction of ROQUIN-1 with individual high-affinity CDE-type stem-loops, it may attract more Roquin proteins to bind to variable, lower affinity secondary structures in the vicinity. Such local enrichments of Roquin binding events could blur distinct motif definition beyond the observed CDE-type or U-rich loops. Second, as demonstrated for *Zc3h12a*, ROQUIN-1 binds two stem-loop elements that were also proposed to be targeted by Regnase-1. Preclinical strategies employing antisense oligonucleotides (ASOs) to disrupt these stem-loops have shown potential for treating autoimmunity[41]. Our findings suggest that such approaches must now consider ROQUIN proteins as additional regulatory factors. Third, we show that the majority of genes that become differentially expressed after inactivation of Roquin-1/2 in T cells are indirectly affected – which has led to misinterpretations. A prominent case is *Ifng*, whose elevated expression contributes to lupus-like autoimmunity in *sanroque* mice with hypomorphic Roquin-1 function. Although *Ifng* has previously been suggested to be a direct ROQUIN target, we found no evidence of Roquin binding to *Ifng* in $CD4^+$ or $CD8^+$ T cells in our iCLIP2 or reporter assay. We therefore propose a so-far unknown upstream regulator as direct Roquin target that affects IFN-γ expression on the epigenetic, transcriptional, or post-transcriptional levels. Finally, our cross-species iCLIP2 analysis reveals evolutionary conservation of ROQUIN targets, binding sites, and likely regulatory mechanisms—suggesting that our mouse and human datasets will be critical to interpret disease-linked variants of concern especially within 3'-UTRs. The discovered conservation underscores the value of incorporating evolutionary context into future computational efforts to define ROQUIN-responsive *cis*-elements and investigations of the composition of its response elements. In sum, our detailed characterization of ROQUIN-1's RNA-binding landscape provides a mechanistic framework for deciphering the code of Roquin-mediated post-transcriptional regulation including target specificity, transcriptome encoded regulatory modes and cellular context.

We describe a new role of Roquin proteins in lymphocyte development. Molecularly, we propose that deletion of Roquin encoding alleles antagonizes the function E-box transcription factor such as E2A, E2-2 and HEB in cells that express Roquin and Id proteins. Accordingly, the Roquin-1/2 double-knockout (DKO) phenotype may phenocopy an E-box binding factor deficiency as for example seen in the block of B cell development in E2A and Roquin deficient mice[51, 53]. The observed phenotypes may be explained by dysregulation of established E2A/HEB targets including *Rag1*[48] and *pre-TCRα*[50], but may potentially

also involve upstream changes in the topology of Igh and Rag1/2 loci[59, 60]. Supporting this model, we show upregulation of *Id1* and *Id3* in pre-pro-B cells following Roquin inactivation. Downregulation of *Id1* has been proposed as a prerequisite for B cell commitment[61], and in line with this, both E2A deficiency and Roquin-1/2 co-deletion arrest development at the pre-pro-B cell stage. While other RBPs—such as Regnase-1, Zfp36L1/Zfp36L2, and downstream effectors like Cnot3 or Upf1—also modulate lymphocyte development[62, 63, 64], the developmental arrest observed upon Roquin inactivation occurs at unique stages. Specifically, the block in B cell development precedes that seen in *Zfp36l1/l2*, *Upf1*, or *Cnot3* knockout models. Interestingly, combined ablation of Regnase-1 and Regnase-3 in hematopoietic progenitors ($Zc3h12a^{-/-}$;$Zc3h12c^{-/-}$) also results in a failure to generate B cells; however, this block occurs before common lymphoid progenitors at the stage of multipotent progenitors[65]. Notably, B cell development remains unaffected when Regnase-1 is inactivated using the same *Cd127*-Cre driver employed in our study context of Regnase-3 deficiency ($Zc3h12a^{fl/fl}$;$Zc3h12c^{-/-}$; *Cd127*-Cre) [65]—highlighting the distinct developmental process in which Roquin exerts its critical function.

The Id proteins have established roles in controlling processes including angiogenesis, neurogenesis, hematopoiesis and tumorigenesis. Given the specific functions of Id proteins in many cell types including stem cells, neurons myeloid, erythroid, T, B, NK and ILCs, by impacting on their differentiation and plasticity, it is tempting to speculate that Roquin shapes diverse developmental trajectories across different cell types through this axis. Conversely, this newly uncovered regulation of Id proteins raises the possibility that Roquin's ability to prevent autoimmunity may critically depend on this pathway in T cell function and development and thereby involve peripheral as well as central tolerance mechanisms.

Online Methods

Mice

All mice used in this study were on a C57BL/6 background. All animals were housed in a specific pathogen-free barrier facility under a 12h/12h dark/light regime at 20-24°C and at a humidity of 45-65% in accordance with the Helmholtz Zentrum München and the Ludwig-Maximilians-Universität München institutional, state and federal guidelines. All experimental procedures involving male or female mice were performed in accordance with the regulations of and were approved by the local government (Regierung von Oberbayern reference numbers 55.2-2532.Vet_02-18-10, 55.2-2532.Vet_02-19-68, ROB-55.2-2532.Vet_02-24-39 and ROB-55.2-2532.Vet_02-24-83). *Roquin-1/2*$^{fl/fl}$ mice are transgenic for the Roquin-1-encoding gene *Roquin-1*[1] and the Roquin-2-encoding gene *Rc3h2*[3]. Transgenic *Roquin-1/2*$^{fl/fl}$ mice were crossed with either Cre-ERt2[67], *Cd4*-Cre-ERt2[68]or *Cd127*-Cre[46] transgenic mice. Data collection and analysis were not performed blind to the conditions of the experiments. Experimental groups were assigned according to genotype.

In vivo deletion of Rc3h1/2 floxed alleles

Male and female Cre-ERt2 or *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;Cre-ERt2 mice (age > 8w) were fed a total dose of 20mg tamoxifen (Sigma) in corn oil by oral gavage twice a day on two consecutive days. Mice were sacrificed two days after the last gavage and single cell suspension were prepared and stained for sorting.

Isolation and *in vitro* cultivation of murine primary CD4$^+$ T cells, CD8$^+$ T cells and B cells

Isolation, *in vitro* deletion of floxed alleles and *in vitro* cultivation of primary CD4$^+$ T cells was performed as described[57]. Briefly, total CD4$^+$ T cells were isolated from spleen and lymph nodes using the EasySep^{TM} Mouse CD4 T Cell Isolation Kit (STEMCELL) according to manufacturers' instructions. To induce deletion of Rc3h1/2 floxed alleles, the isolated CD4$^+$ T cells were incubated with 4´-hydroxy-tamoxifen (1 µM; Sigma-Aldrich; cat# H7904) for 24h prior to activation. CD4+ T cells were activated with α-CD3 (0.1 µg/mL; cl. 2C11, in house production), α-CD28 (1 µg/mL; cl. 37.5N, in house production) on anti-hamster antibody (50 µg/mL; MP Biochemicals) pre-coated plates under the following conditions: *T_H1 differentiation*: 10 µg/mL α -IL-4 (cl. 11B11, in house production), 10 ng/mL IL-12 (BD Pharmingen). For *T_H0 conditions*, no cytokines were added. *T_H17 differentiation:* α-CD3 (0.5 µg/mL; cl. 2C11, in house production), α-CD28 (2.5 µg/mL; cl. 37.5N, in house production), 10 µg/mL α –IFN-g (cl. XMG1.2.1, in house production), 10 µg/mL α -IL-4, 20 ng/mL rmIL-6 (R&D System), 5 ng/mL rhTGF-b1 (R&D System); *T_{reg} differentiation:* α-CD3 (0.5 µg/mL; cl. 2C11, in house production), α-CD28 (2.5 µg/mL; cl. 37.5N, in house production), 10 µg/mL α –IFN-g (cl. XMG1.2.1, in house production), 10 µg/mL α -IL-4, 100 IE/mL hIL-2 (Novartis), 2 ng/mL rhTGF-b1 (R&D System).

40-48h after start of activation, the cells were resuspended, cultured in medium supplemented with recombinant hIL-2 (200U/mL, Proleukin S, Novartis) and expanded for 2-3 days, as indicated.

For CD8$^+$ T cells, total CD8$^+$ T cells were isolated from spleen and lymph nodes using the EasySep^{TM} Mouse CD8 T Cell Isolation Kit (STEMCELL) according to manufacturers' instructions. CD8+ T cells were activated with α-CD3 (0.5 µg/mL; cl. 2C11, in house production), α-CD28 (1 µg/mL; cl. 37.5N, in house production) on anti-hamster antibody (50 µg/mL; MP Biochemicals) pre-coated plates with recombinant hIL-2 (700U/mL, Proleukin S, Novartis). After 40-48h from start of activation, the cells were resuspended, cultured in medium supplemented with recombinant hIL-2 (700U/mL, Proleukin S, Novartis) and expanded for 3 days.

For iGB cell culture, B cells were expanded in coculture with 40LB feeder cells that express CD40L and BAFF as previously described[27]. Briefly, splenocytes were cocultured with 40LB cells in B cell medium (RPMI Glutamax (Thermo Fisher cat. 61870-010) supplemented with 10% FCS, 100 units/ml penicillin, 100 µg/ml streptomycin, 10mM HEPES, 50 µM b-Mercaptoethanol, 1x MEM non-essential amino acids (100x), 1x Sodium pyruvate (100x)) in the presence of 1 ng/ml recombinant IL4 (Biolegend, cat. 574302). On day 3, the expanded B cells were replated onto fresh 40LB cells with 10 ng/ml recombinant IL21 (Biolegend, cat. 574504) and cultured until harvest on day 6.

Isolation and *in vitro* cultivation of human CD4$^+$ and CD8$^+$ T cells

Blood from healthy donors with approval by the Ethics Committee of the LMU München (project no. 19-819) was heparinized and collected into leukocyte reduction system cones. Blood cells were diluted with PBS (Gibco) and lymphocytes were separated by gradient centrifugation (anprotec). CD4+ or CD8+ T cells were isolated by positive selection using magnetic microbeads (Miltenyi Biotec), according to manufacturers' instructions. For activation, T Cell Activation beads (Miltenyi Biotech) were added to T cells at a ratio of 1:2 (bead to T cell) and kept in RPMI 1640 GlutaMAX (Gibco) supplemented with 10% fetal bovine serum (FBS; Anprotec), 10mM HEPES, and penicillin/streptomycin (100 IU/ml; Thermo Fisher Scientific) for 72h at 37°C and 5% CO_2. Fresh medium containing IL-2 (for $CD4^+$: 50 U/ml; for $CD8^+$ 100 U/ml; Biomol) was added to the cells for 3 days at 37°C and 5% CO_2.

Cell Culture of HEK cells, MEF cells and OP9/OP9-DL1 cells

HEK293T (ATCC), hCARLSL;*Rc3h1/2*$^{fl/fl}$;Cre-ERt2 MEF cells[26], *Rc3h2*$^{-/-}$ MEF cells and *Rc3h1/2*$^{-/-}$ clonal MEF cells containing a doxycycline inducible *Rc3h1*-P2A-mCherry reporter[69] were cultured in DMEM medium (Invitrogen) supplemented with 10% (v/v) FCS (BioSell), 1% Pen-Strep (Thermo Fisher) and 1% HEPES, pH7.2-7.5 (Invitrogen) at 37°C and 10% CO_2. OP9 bone marrow stromal cells expressing the Notch ligand Delta-like ligand 1 (OP9-DL1) and OP9 control cells (OP9-GFP) were provided by J.C. Zuñiga-Pflucker (University of Toronto, Toronto, Canada). OP9 and OP9-DL1 cells were grown in freshly prepared minimum essential medium (MEM) alpha (Thermo Fisher) supplemented with 10% FCS, 1 mM sodium pyruvate, 10 mM HEPES and 100 U/mL penicillin-streptomycin (all Gibco).

Calcium-phosphate transfection for generation of retroviral particles

HEK293T cells pre-treated with 25µM chloroquine, were cotransfected with the packaging vector pCL-Eco (Addgene; 12371) and the respective MSCV expression plasmid using calcium phosphate as a transfection reagent. After 6h cells were washed and cultured in fresh medium for an additional 40h. Viral particles were filtered (0.45µM) and mixed with polybrene (10µg/ml) prior to MEF cell, B cell, T cell, OP9 or OP9-DL1 cell transduction.

Transduction of MEF cells for GFP reporter assays

hCARLSL; *Rc3h1/2*$^{fl/fl}$;Cre-ERt2 or *Rc3h1/2*$^{-/-}$;*Rc3h1*-P2A-mCherry MEF cells were seeded into 24 well plates and spin infected with MSCV-pEGFP constructs fused to the indicated 3´-UTRs or mutations thereof (300xg, 32°C, 2h). After a 6h incubation at 37°C the virus containing medium was exchanged for fresh MEF cell medium and cells were allowed to rest overnight. On the next day, transduced hCARLSL;*Rc3h1/2*$^{fl/fl}$;Cre-ERt2 MEF cells were trypsinized, transferred onto 12 well plates (2 wells per reporter construct) and rested for 4-6h. One well per reporter was treated with 0.3µM 4´-hydroxy-tamoxifen for 5d to induce Roquin deletion, the other well received fresh MEF cell medium without tamoxifen every day. Transduced *Rc3h1/2*$^{-/-}$;*Rc3h1*-P2A-mCherry MEF cells were expanded in MEF cell medium for 4 days. Cells were then trypsinized and transferred onto 12 well plates (2 wells per reporter construct) and rested for 4-6h. One well per reporter was treated with 1µg/ml Doxycycline for 16h the

other well received fresh MEF cell medium. On the day of analysis MEF cells were trypsinized, washed with PBS and lysed in Trizol for RNA isolation or transferred onto 96-well v-bottom plates, washed with FACS buffer and acquired on the BD LSR Fortessa.

Transduction of B cells and T cells for GFP reporter assays

For transduction, B cells were spin-infected on day 2 after seeding with retrovirus supplemented with 2 ng/ml IL4 and 8 µg/ml polybrene at 500xg for 1.5 hours at 32 °C, followed by overnight incubation at 37°C and 5% CO_2. On the next day, the virus containing medium was removed and the B cells of each transduction were replated into two plates with 40LB cells in the presence of 1 ng/ml IL4 and with or without 4´-hydroxy-tamoxifen (1 µM) and incubated for 3 days.

$CD4^+$ and $CD8^+$ T cells were transduced 40h after start of activation with anti-CD3/anti-CD28 antibodies. After addition of the virus containing medium, T cells were centrifuged for 1h at 800xg and 18°C and incubated for 4-6h at 37°C. The virus containing medium was removed, the T cells resuspended in T cell medium supplemented with IL-2, splitted into two wells and incubated for 24h with or without 4´-hydroxy-tamoxifen. Afterwards the cells were washed, resuspended in fresh medium containing IL-2 and cultivated for another 2d.

Transduction and *in vitro* deletion of common lymphoid progenitors

$Lin^-Sca\text{-}1^+CD117^{hi}$ (LSK) and $Lin^-CD127^{lo}CD135^{lo}CD117^{lo}$ (CLP) cells were sorted from lineage-depleted bone marrows and cultured overnight in MEM-a supplemented with SCF (50 ng/mL), IL-7 (25 ng/mL), Flt-3L (25 ng/mL) and IL-6 (20 ng/mL) (all Peprotech). For retroviral vector transduction LSK and CLP cells were then transferred into 96-well plates (Sarstedt, Germany) preloaded with retroviral vector attached to RetroNectin (Takara, Japan) according to the manufacturer's protocol (see also DOI 10.1182/blood-2014-07-592667). For tamoxifen-induced deletion, progenitor cells were cultured with identical cytokine-supplemented medium and 4´-hydroxy-tamoxifen (1 µM).

OP9 co-cultures

OP9 and OP9-DL1 co-culture assays were performed essentially as described[44, 45]. Progenitor cells were plated at an initial density of 1 to 20 $x10^3$ cells on subconfluent OP9-GFP or OP9-DL1 monolayers at 5 x 10^4 cells/well in a 24-well plate. For OP9/OP9-DL1 mixed cocultures, cells were mixed in a 2:1 ratio. All cocultures were performed in the presence of 1 ng/mL IL-7, 5 ng/mL Flt3 ligand (Flt3-L) and 10 ng/mL stem cell factor (SCF) (all Peprotech). Fifty percent of the medium was replaced on day 4 of the coculture.

Expression plasmids

The full coding sequences (CDS) of murine *Id1* variant 2, *Id2*, and *Id3* were amplified from murine T cell cDNA via PCR and cloned into the MSCV–IRES–Thy-1.1 expression vector. Cloning was performed using

the Gateway® system (Thermo Fisher Scientific). For Reporter Assays the 3´UTR of the corresponding gene of interest was inserted into a MluI restriction site downstream of the GFP coding sequence (CDS) into the retroviral expression plasmid MSCV-pEGFP. Mutations in the 3´-UTRs were inserted via the QuikChange site directed-mutagenesis procedure (Stratagene). Primer sequences are available upon request.

Flow cytometry and data analysis with FlowJo

Single cell suspensions were stained with fixable blue viability dye (Thermo Fisher Scientific) or fixable viability dye eFluor780 (Invitrogen) in PBS for 20min at 4°C. For the detection of surface proteins cells were stained with the appropriate antibodies in FACS buffer (2% FCS, 1mM EDTA in PBS) for 20min at 4°C. For commercially available intracellular antibodies cells were fixed with Foxp3 Fixation/Perm Buffer (eBioscience) according to manufacturer's protocol for 30min at 4°C, permeabilized with Foxp3 permeabilization Buffer (eBioscience) and stained with antibodies diluted in Foxp3 permeabilization Buffer for 40min at 4°C. For intracellular staining with in-house generated antibodies cells were fixed with 2% formaldehyde for 15min at RT, washed with saponin permeabilization buffer (0.5% saponin, 1% BSA in PBS) and stained with antibodies in saponin buffer for 30min at 4°C. Cell populations were acquired on BD LSR Fortessa or BD FACSCanto II flow cytometry devices. Data were processed using FlowJo software (v10.6.0, BD Bioscience). An exemplified gating strategy is shown in **Supplemental data Fig. S8**.

For multiparameter flow cytometry cells were stained with fixable viability dye, followed by surface staining in FACS buffer supplemented with Brilliant Stain Buffer (Invitrogen) for 40min at 4°C. Cells were washed and fixed using the Foxp3 Fixation/Perm Buffer (eBioscience) following manufacturers' instructions. For intracellular staining antibodies were diluted in Perm Buffer (eBioscience) and staining was performed at 4°C overnight. Cell populations were acquired on the Cytek Aurora device and analysed using FlowJo software (v10.6.0, BD Bioscience). The antibodies used for multiparameter flow analyses are listed in **Supplemental data Table 6**.

Sorting for qPCR analysis: Single cell suspensions from Thymus and bone marrow (BM) of gavaged WT or iDKO mice were generated and surplus populations were depleted using MACS LS columns (Miltenyi) according to manufacturers' instructions. Following biotinylated antibodies were used for depletion: Thymus: anti-CD4 (1:400), anti-CD8 (1:400), BM: anti-CD11b (1:500), anti-CD11c (1:500), anti-TCRγδ (1:200), anti-Gr1 (1:500), anti-NK1.1 (1:200). Depleted cells were stained as described above and sorted using BD Facs Aria III. Antibodies used for sorting are listed in **Supplemental data Table 6**.

Sorting of LSK and CLP cells for OP9/OP9-DL1 co-culture: Lin-Sca-1^{+}CD117^{hi} (LSK) and Lin^{-} CD127^{lo}CD135^{lo}CD117^{lo} (CLP) cells were sorted from lineage-depleted BM cells. Depleted cells were stained as described above and sorted using BD Facs Aria III. Following antibodies were used: Lineage (CD19 (cln.: 6D5), CD11b (cln.: M1/70), CD11c (cln.: N418), NK1.1 (cln.: PK136), TCRb, Gr-1 (cln.: RB6-

8C5)) all PE-Cy7, CD127 PE (cln.: A7R34), CD135 BV421 (cln.: A2F10), Sca-1 BV650 (cln.: D7), CD117 Ax647 (cln.: ACK2) (all Biolegend).

preTCRα staining: Thymocytes were stained with viability dye for 20mins on ice. Cells were resuspended in 2x Fc block in FACS buffer (in house production) and incubated for 5 mins. 2x master mix containing anti-preTCRα (1:100) in FACS buffer was added to the cells and incubated for another 25min at 4°C. Cells were washed and stained with biotinylated anti-IgG1 (1:100) followed by streptavidin-PE (1:1000) and subsequent fixation using the Foxp3 Fixation/Perm Buffer Set (eBioscience) according to manufacturers' instructions. All remaining surface and intracellular antibodies were diluted in Perm buffer containing Brilliant Stain Buffer (Invitrogen) and stained at 4°C overnight. Antibodies used for pre-TCRα staining are listed in **Supplemental data table 6**. Cells were acquired on the Cytek Aurora and analyzed using FlowJo.

Cell lysis, SDS-PAGE and immunoblotting

Cell lysis and SDS-PAGE were done as described [15]. For immunoblotting proteins were transferred to a PVDF membrane and analyzed using primary antibodies followed by horseradish peroxidase (HRP)-conjugated secondary antibodies (Cell Signaling). For protein detection, the Amersham ECL Prime Western Blotting Detection Reagent and X-ray films were used. The following primary antibodies were used: Roquin-1/2 (cl. 3F12, monoclonal, in house production), GAPDH (cl. 6C5, Merck Millipore), Rag1 (cl. AB172637, Abcam), Id1 (18475-1-AP, Proteintech), Id2 (cl. D39E8, Cell Signaling), Id3 (BCH-4/#17-3, Biocheck), Tubulin (cl. 8328, Sigma), Map3k8 (cl. M-20, Santa Cruz), Lfng (cl. D6V2V, Cell Signaling).

RNA isolation and RT-qPCR

RNA isolation was performed by column-based RNA isolation utilizing the NucleoSpin RNA isolation kit (Macherey-Nagel), RNeasy kit (QIAGEN) or Trizol reagent (Ambion) according to manufacturer's protocol. RNA was transcribed into cDNA using the Quantitect RT Kit (Roche) according to the manufacturer´s protocol. To quantify gene expression, the UPL Probe Library System by Roche or Power SYBR™ Green PCR Master Mix (Life Technologies) and the Roche Light Cycler 480 were utilized.

RNA-Sequencing

For RNA-Sequencing $CD4^+$ T cells were taken up in Trizol Reagent (Ambion) and after mRNA extraction according to manufacturer`s protocol libraries were prepared using the Universal RNA-Seq Library Preparation Kit (NuGEN). Illumina Next Generation Sequencing with a length of 100 bp paired-end and 15×10^6 reads/sample was performed at the LAFUGA platform (LMU München) using the Illumina HiSeq1500 sequencer.

iCLIP2

Library preparation was performed as previously described [20]. Briefly cells were UV-crosslinked (254 nm, 2 × 200 mJ/cm²), lysed in buffer containing protease inhibitors, and partially digested with RNase. After centrifugation, lysates were incubated with anti-Roquin-1 antibody (Bethyl; cl. A300-514A) conjugated to

Dynabeads Protein A (Invitrogen). Immunoprecipitated complexes were washed and subjected to 3' dephosphorylation and adapter ligation. After 5' end labeling, RNA-protein complexes were separated by SDS-PAGE and transferred to a nitrocellulose membrane (Whatman). RNA was isolated by proteinase K (Invitrogen) treatment, purified, reverse-transcribed, ligated with a second adapter, and PCR amplified (8–13 cycles). Libraries were size-selected, quantified with a High Sensitivity D1000 TapeStation Kit (Agilent Technologies) on a TapeStation 2200 (Agilent Technologies). Libraries were sequenced with a 50 bp length and 50 million reads per sample on an Illumina HiSeq1500 sequencer at the LAFUGA platform (LMU München).

iCLIP2 read processing and peak calling

Read processing was done following[23]. Raw reads were checked for experimental quality using *FastQC 0.11.9*, while *Flexbar 3.5.0* was employed for demultiplexing and adapter trimming. Individual samples were mapped to the reference genome (GRCm38 / mm10 for mouse, GRCh38 / hg38 for human). Crosslink sites at single-nucleotide resolution were called per single batch on the merge of all the replicates, with the peak caller *PureCLIP*[22] (version 1.3.1) with parameters *bc* set to 1 and *b2p* set to 0.03. Subsequently, identified peaks were filtered and refined into binding site intervals of 9nt in length centered at the cross-link site, following[23] where we require a binding site to have a sufficient number of crosslink events for all the replicates in a batch to be considered.

The iDKO filtering procedure applied on the mouse data was implemented with the following formula

$$\frac{Cov(BS_{WT})}{Cov(BS_{iDKO})} \cdot \lambda \geq c$$

Where $Cov(BS_{WT})$ and $Cov(BS_{iDKO})$ represent the read coverage for the wildtype and knockout samples at binding sites, respectively, while λ is a scaling factor computed as ratio between the number of total reads for the two experiments (0.66 for batch 1 and 1.39 for batch 2) and c is the fold change threshold that determines the filtering cutoff. We retained only the binding sites surpassing a fold change (FC) cutoff of 1.5, corresponding to a 50% increase of experiment signal over iDKO.

This step was omitted in the analysis of human data, and in mouse CD8, T_H0, T_H17 and T_{reg}, MEF and B cells as matching iDKO controls were not available. Finally, confirmed binding sites were annotated with an in-house pipeline using GENCODE VM25 for mm10 and GENCODE v40 for hg38[70].

Merging binding signal from different batches

To account for possible batch effects, binding sites that were determined separately in the two $CD4^+$ T cell iCLIP2 batches are later pooled into a final set after ensuring that each binding site is sufficiently

supported by both batches. More in detail, after merging, we require that each binding site identified in one batch shows read coverage in the other batch exceeding the 70th percentile of that batch's binding site coverage distribution.

Binding sites which overlapped between batches (i.e. results of peaks called at the same positions in the two batches) were solved by identifying the position of the maximum crosslink signal after library size-normalization, and designating that as the center of a new binding site of length 9nts. In case of single-replicate and/or single-batch experiments, this step was omitted.

Structure motifs extraction

We trained the convolutional neural network Pysster[24] as a classifier to distinguish between bound sequences as windows of 100nts centered at Roquin-1 peaks (*positives*) from unbound sequences of the same length, constructed around peaks called on the control Input experiment (*negatives*). To train in a balanced setting, positive sequences were subsetted to match the number of negatives, after ranking them by binding site score and selecting the top 12,000. The final number of sequences in the training dataset was 24,000, with a 1:1 ratio of positives and negatives. Specifically, input sequences converted to an RNA secondary structure notation, from a fixed 4-letter alphabet (H: hairpin, I: internal loop, M: multiloop, S: stem) obtained by running the *forgi* (ver. 2.1.2) library on the dot-bracket strings generated as secondary structure prediction output by the *Vienna RNAFold* (2.5.1) tools.

After hyperparameter tuning, the network was trained with 2 convolutional layers, 40 kernels or filters of length 20, and dropout rates for input, convolutional and dense layers set respectively to 0.1, 0.4 and 0.6, with a batch size of 128 and a patience of 10 epochs for early stopping.

Exploiting Pysster built-in capabilities, the kernel weights of the first convolutional layer were extracted and plotted as position-weight matrices (PWMs) into logos, for interpretation of structural binding motifs. The n=40 kernel logos obtained were classified into *positive* and *negative* motifs, depending on their enrichment in the positive (bound) or negative class, and positive motifs containing stem-loop like structural features (K7, K11, K23, K32, K34, K36, K39, *n=7*) were selected as motifs driving Roquin-1 binding and used for further downstream analyses.

Sequence motifs analyses

The Regulatory Sequence Analysis Tool (RSAT) suite[25] (locally installed rsat-core using the conda package version 2020.02.29) was applied to identify de novo primary sequence motifs enriched within the set of Roquin-1 binding sites sequences as compared to background regions, characterize these motifs, and rescan binding sites sequences to identify their positions. We first describe the sets of sequences used in these analyses, then describe the different steps in detail.

Sequence motif analysis was performed on 3'-UTR binding sites from protein-coding genes, extending each 9-nt site by 9 nt upstream and downstream to form 27-nt intervals, merged if overlapping. To

control for iCLIP2 protocol biases, peak calling was also applied to input controls, excluding regions overlapping true binding sites. Motif discovery focused on the top 10% of binding sites ranked by PureCLIP score to enrich for relevant sequence determinants. Furthermore, mouse binding sites were intersected with the activation map of RNA-structure kernels from the Pysster model, so as to enrich for sequences motifs from relevant structure context, but this did not exclude any of the sequences. Enriched motifs, manipulated each in the form of a matrix of counts, were used to scan the full set of 3'UTR binding sites, to characterize the extent of binding modalities of Roquin and candidate RNA-binding proteins binding in proximity, while confirming their specific enrichment in true peaks compared to control peaks.

The oligo-analysis tool was used to identify enriched motifs within 27-nt binding site sequences by scanning for 5–7 nt oligos on the binding strand. The top 50 enriched oligos were merged (allowing one substitution) into a maximum of 20 patterns, which were re-scanned to build final motifs based on a score ≥3, including 3 flanking nucleotides for context. Enrichment was assessed against background sequences using 3rd- and 4th-order Markov models on both strands, yielding two motif sets per species.

For each species, motif sets were unified using the matrix-clustering tool, applying the "Ncor" similarity metric and clustering based on average motif profiles. Default thresholds were used (column correlation ≥0.7, normalized correlation ≥0.4). Final motifs were re-generated by rescanning the original sequences, retaining matches with a weight score ≥3. The motifs were compared against known RBPs binding motifs from the oRNAment database[71] using the RSAT compare-matrix command, accepting a minimum correlation of 0.7 and normalized correlation of 0.4, not considering the reverse-complemented motifs. The final assignments were done by sorting the motif that matched on their normalized correlation score.

The count matrices from the matrix-clustering step were used to scan the full set of 3'-UTR binding site sequences with matrix-scan, using default parameters (Markov model order 1, pseudo-frequencies 0.01, pseudo-counts 1, decimal 1, P-value ≥0.1), except for strand-specific scanning and a reduced minimum-weight threshold of 0.8 to allow exhaustive motif detection, while still enabling stricter downstream filtering based on matching thresholds.

Orthology status of binding sites and sequence conservation

Evaluating the correspondence between ROQUIN-1 peaks called in the human-data experiments and the Roquin-1 peaks called in the mouse experiments was done by performing two-ways lift-over of genomic coordinates from both the human genome (version hg38) and the mouse genome (version mm10]), using the appropriate chain files from the FTP of the UCSC[72]. Binding sites (BSs) and their lifted-over genomic coordinates were annotated with GENCODE gene and biotype annotations. Orthology relationships were evaluated both at gene level and at biotype level. A table of orthologous genes was built by aggregating Ensembl gene IDs orthologies from BioMart database[73], the OMA Orthology Database[74], and the MGI Mouse Genome Informatics database[75]. A given gene was labeled as bound in

both species if any of its orthologs presented at least one BS in both species, restricting or not to the 3'UTR biotype. This enabled the identification of genes presenting BSs only in one of the species (labeled “species-specific”) and genes presenting BSs in their orthologs (labeled “orthologous”). The orthologous binding patterns were further refined by considering the distance between genomic coordinates of BSs in one species and lifted-over coordinates from the BSs in the other species Setting a maximum distance of 10 nucleotides, BSs from one species were either labeled as “distant” or as “matching”, indicating that the BS from one species occurs in proximity or overlaps with the lifted-over coordinates of a BS identified in the other species.

Nucleotide sequence conservation of BSs was evaluated from the mouse genome viewpoint, using the PhyloP60w scores from the UCSC[72]. The score provides an intuitive interpretation for positive (evolutionary constraint) and negative (evolutionary acceleration) values, while enabling quantification and comparison of this constraint level as measured across multiple regions. Average scores (over the span of each 9 nucleotides region) were calculated for each genomic interval from the different sets. The BedTools suite[76] was used to generate a set of control regions by randomly sampling the number of BSs within bound 3’-UTRs, excluding Roquin binding sites. Differences between sets of regions was statistically evaluated using the Mann-Whitney U Test, using Bonferroni correction of P-values to account for multiple testing.

Structure conservation analysis

We produced RNA-structure based realignment of orthologous sequences so as to evaluate the conservation of consensus structures at Roquin-1 BS locations in the 3’UTR of mouse genes. Multi-z sequence alignments of 60 vertebrates (59 species aligned (see Extended genome as reference and **Supplemental data Table 5**) were downloaded from the UCSC (https://github.com/bxlab/bx-python). We used Roquin-1 BSs from the mouse experiments, annotated within 3’UTR of genes, to query and extract alignment blocks intersecting the genomic coordinates of query BSs. When successful, a query BS was associated with an alignment block, and genomic coordinates of the center of the BS were extracted from all the aligned species sequences. In the case of a species’ sequence not being aligned at the BS location within the block, we extracted the genomic coordinates of the first aligned nucleotides upstream and downstream. The single-base genomic intervals from all species were then extended by 100 nucleotides upstream and downstream, and the genome sequences were queried to build a multi-species FASTA file. These nucleotide sequences were input to mLocARNA 2.0[39] (--struct-weight 200 --max-diff-am 30 --struct-local false --sequ-local false --width 120 --noLP --consensus-structure alifold --min-prob 0.01 --alifold-consensus-dp) to generate ClustalW alignments and consensus secondary structures for each Roquin-1 mouse peak associated with vertebrate sequences. These were analyzed using the aln_conservation_struct function from the ViennaRNA Python API (v2.6.4) [77], which computes, for each base-paired column in the consensus structure, the fraction of aligned nucleotides compatible with canonical base-pairing, providing a measure of structure conservation ranging from 0 to 1.

For a given Roquin-1 BS and its associated mLocARNA alignment[39, 77] and consensus structure, we apply the consensus structure to the 200 nucleotides of the aligned sequence of mouse and extract the structure-conservation values of all base-pairing positions. We distinguish base-pairing positions at the center that overlaps with the BS of interest, from other base-pairing positions, further excluding any positions overlapping with any other Roquin-1 BS. The remaining base-pairing positions are used as control, background positions. We finally calculate the average structure conservation score of the BS base-pairing positions and the average structure conservation score of the background base pairing positions. The distribution of structure conservation scores across BS is finally compared to the distribution calculated from background base-pairing positions. Furthermore, the PhyloP60w sequence conservation scores of base-pairing positions were retrieved, so as to compare the levels of sequence conservation and structure conservation.

Pathway enrichment analysis

A curated set of terms[31] was collected from GO[28, 29] and KEGG[30] used to perform an over-representation analysis (ORA) of Roquin 3′-UTR targets against a background of all genes expressed in $CD4^+$ T cells, according to our RNA-seq data. Significance testing was performed carrying out a hypergeometric test (Fisher's test) and applying the false discovery rate method to control for false positives. Only resulting pathways with FDR < 0.05 were selected for visualization.

Analysis of RNA-Sequencing data

The reads were mapped with STAR[78] (version 2.3.1z4) against the mouse genome (GRCm38). Read counts for genes were derived in two ways: 1. counting read pairs fully contained within annotated genes and 2. counting only read pairs fully consistent with at least one annotated transcript using the ENSEMBL[79] annotation GRCm38.75.

Differential expression analysis of RNA-seq transcripts was performed in R with DESeq2 (1.41.2) to identify significant expression changes for deregulated genes between the two genotypes: wildtype (WT) and Roquin-1/2 double knockout (iDKO), across the four different timepoints (0h, 1h, 2h, 4h). To do so we ran DESeq2 (default parameters) with a *contrast* term set to the combination of genotype and timepoint, comparing to the WT baseline at timepoint 0h. edgeR (4.0.2) was previously used to filter out genes that had too few counts to be considered for significance testing and the *lfcShrinkage* method was applied at the end to obtain more reliable estimates.

Secondary structure prediction and visualization

A Galaxy workflow was created that systematically extracts all potentially homologous RNA sequences based on a query sequence file from available chordate genomes and returns a multiple-sequence-structure alignment of all candidates: This workflow downloads the NCBI Genome Datasets https://www.ncbi.nlm.nih.gov/datasets/, release 14: CLI v14.x genomes of 472 chordates (only those with available annotation, reference, and representative assembly) using the NCBI Datasets command-

line tools https://www.ncbi.nlm.nih.gov/datasets/docs/v2/download-and-install/, release 14: CLI v14.x (API v2alpha). It uses MegaBLAST NCBI https://doi.org/10.1093/bioinformatics/btn322, from NCBI BLAST+ 2.10.1 and nucleotide BLAST https://doi.org/10.1016/s0022-2836(05)80360-2, from NCBI BLAST+ 2.10.1 in parallel expectation cutoff of 0.001 to identify homologous sequences of the known and provided human and mouse *Id1-3* and *Zc3h12a* 3'UTRs. A filter step selects the highest-scoring hit sequence from either human or mouse in the target genome (needs more details), i.e. bitscore * (N_match - N_mismatch - 0.5 * N_gaps) / L_query, where - L_query: length of the original query sequence (column 11 of blast tabular output), N_match: number of matches (column 15), N_mismatch: number of mismatches (column 5), N_gap: number of gaps (column 17). In the final step, the LocARNA tool https://doi.org/10.1371/journal.pcbi.0030065, release 1.9.2.3folds and aligns the selected sequences simultaneously and produces a multiple-sequence-structure alignment for each 3'-UTR. The alignment was inspected and partially curated manually. A set of 112 species for the *Id* family members and of 26 species for the *Zc3h12a* mRNA was selected and the R2R https://doi.org/10.1186/1471-2105-12-3, release 1.0.6 software was used to depict aesthetic consensus RNA secondary structures. For a complete list of species and associated identification numbers, see **Supplemental data Table 5**.

Statistical analysis and experimental design

Statistical analysis was performed with Prism 5.0b, Prism 10.4.1 (GraphPad), Origin and R (version 4.3.1). *P* values were calculated with two-tailed Student's t test, ordinary one – way ANOVA or one-tailed Mann Whitney test, as indicated in figure legends. Statistical significance was indicated. Error bars represent mean of all data points with +/- SEM. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[5]. Data distribution was assumed to be normal, but this was not formally tested. The experiments did not involve randomization of animals/samples or conditions or blinding of investigators. No animals or data points were excluded from analyses.

Declarations

Materials availability:

Materials are available upon reasonable requests.

Data and code availability.

The RNA sequencing data are available under:

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152961.

The iCLIP2 data have been deposited in NCBI's Gene Expression Omnibus[66] and are accessible through GEO Series accession number GSE295259. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE295259).

The bioinformatics scripts and analyses associated to this manuscript have been published under https://github.com/marsico-lab/roquin-iclip.

Author Contribution

E.H.W., T.K., and G.B. performed and G. Cantini and L.M. carried out the bioinformatic analysis of the iCLIP2 data. G.B. and E.H.W. analyzed reporter screen and Roquin-mediated regulation of *Id* genes. T.R. investigated lymphocyte development. C.X. and M.Ł. performed in vitro lymphocyte differentiation experiments, K.P.H., M.X., L.K., A.N., N.K., and E.T., contributed additional experiments, G. Csaba analyzed the mRNA-seq data under supervision by R.Z., S.M. provided critical protocols and conceptual input for human T cell culture. J.S. performed the LocARNA analyses. A.M. supervised the bioinformatic components of the study. V.H. conceived and directed the study and wrote the manuscript together with AM with input from all authors.

Acknowledgment

We would like to thank H-R. Rodewald (DKFZ, Heidelberg, Germany) for providing *Cd127*-Cre mice and Daisuke Kitamura (Tokyo University of Sciences, Chiba, Japan) for providing the 40LB cell line. We thank M. Dreher (Ludwig-Maximilians-Universität) und S. Hartig (Helmholtz Munich) for excellent technical support. The project was funded by the German Research Foundation (DFG) SFB-TRR 338/1-452881907 (project C02 to V.H.), SFB 1054 # 210592381 (project A03 to V.H.), SFB-TRR 355/1 #490846870 (project A06 to V.H.), SFB-TRR267 # 403584255 (project Z02 to A.M.) as well as HE3359/8-1 (#444891219 to V.H.), DFG-DACH (#432656284 to V.H. and SM) and LY 150/2-1 (AOBJ 690849 to M.Ł.), as well as by grants from the Wilhelm-Sander (#2018.082.3 to V.H.) and Krebshilfe (#70113538 to V.H.) foundations. It was also supported by the Federal Ministry of Education and Research (BMBF) as part of the German Center for Child and Adolescent Health (DZKJ) under the funding code 01GL2407A to M. Ł., by the Helmholtz Association under the joint research school “Munich School for Data Science (MUDS)” to G.Cantini and A.M. and by the NBI Cloud within the German Network for BioinformaticsInfrastructure (de.NBI) and ELIXIR-DE.

References

1. Bertossi, A. *et al.* Loss of Roquin induces early death and immune deregulation but not autoimmunity. *J Exp Med* **208**,1749-1756 (2011).
2. Vinuesa, C.G. *et al.* A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature* **435**,452-458 (2005).
3. Vogel, K.U. *et al.* Roquin paralogs 1 and 2 redundantly repress the Icos and Ox40 costimulator mRNAs and control follicular helper T cell differentiation. *Immunity* **38**,655-668 (2013).
4. Tavernier, S.J. *et al.* A human immune dysregulation syndrome characterized by severe hyperinflammation with a homozygous nonsense Roquin-1 mutation. *Nat Commun* **10**,4779 (2019).

5. Behrens, G. *et al.* Disrupting Roquin-1 interaction with Regnase-1 induces autoimmunity and enhances antitumor responses. *Nat Immunol* **22**,1563-1576 (2021).
6. Behrens, G. & Heissmeyer, V. Cooperation of RNA-Binding Proteins - a Focus on Roquin Function in T Cells. *Front Immunol* **13**,839762 (2022).
7. Mai, D. *et al.* Combined disruption of T cell inflammatory regulators Regnase-1 and Roquin-1 enhances antitumor activity of engineered human T cells. *Proc Natl Acad Sci U S A* **120**,e2218632120 (2023).
8. Zhao, H. *et al.* Genome-wide fitness gene identification reveals Roquin as a potent suppressor of CD8 T cell expansion and anti-tumor immunity. *Cell Rep* **37**,110083 (2021).
9. Essig, K. *et al.* Roquin targets mRNAs in a 3'-UTR-specific manner by different modes of regulation. *Nat Commun* **9**,3810 (2018).
10. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**,869-881 (2013).
11. Sgromo, A. *et al.* A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat Commun* **8**,14307 (2017).
12. Jeltsch, K.M. *et al.* Cleavage of roquin and regnase-1 by the paracaspase MALT1 releases their cooperatively repressed targets to promote T(H)17 differentiation. *Nat Immunol* **15**,1079-1089 (2014).
13. O'Neill, T.J. *et al.* TRAF6 prevents fatal inflammation by homeostatic suppression of MALT1 protease. *Sci Immunol* **6**,eabh2095 (2021).
14. Schmidt, H. *et al.* Unrestrained cleavage of Roquin-1 by MALT1 induces spontaneous T cell activation and the development of autoimmunity. *Proc Natl Acad Sci U S A* **120**,e2309205120 (2023).
15. Essig, K. *et al.* Roquin Suppresses the PI3K-mTOR Signaling Pathway to Inhibit T Helper Cell Differentiation and Conversion of Treg to Tfr Cells. *Immunity* **47**,1067-1082 e1012 (2017).
16. Heissmeyer, V. & Vogel, K.U. Molecular control of Tfh-cell differentiation by Roquin family proteins. *Immunol Rev* **253**,273-289 (2013).
17. Janowski, R. *et al.* Roquin recognizes a non-canonical hexaloop structure in the 3'-UTR of Ox40. *Nat Commun* **7**,11032 (2016).
18. Murakawa, Y. *et al.* RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-kappaB pathway. *Nat Commun* **6**,7367 (2015).
19. Song, J. *et al.* Human cytomegalovirus induces and exploits Roquin to counteract the IRF1-mediated antiviral state. *Proc Natl Acad Sci U S A* **116**,18619-18628 (2019).
20. Buchbender, A. *et al.* Improved library preparation with the new iCLIP2 protocol. *Methods* **178**,33-48 (2020).
21. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**,129-141 (2010).

22. Krakau, S., Richard, H. & Marsico, A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol* **18**,240 (2017).
23. Busch, A., Bruggemann, M., Ebersberger, S. & Zarnack, K. iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. *Methods* **178**,49-62 (2020).
24. Budach, S. & Marsico, A. pysster: classification of biological sequences by learning sequence and structure motifs with convolutional neural networks. *Bioinformatics* **34**,3035-3037 (2018).
25. Santana-Garcia, W. *et al.* RSAT 2022: regulatory sequence analysis tools. *Nucleic Acids Res* **50**,W670-W676 (2022).
26. Schlundt, A. *et al.* Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nat Struct Mol Biol* **21**,671-678 (2014).
27. Nojima, T. *et al.* In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. *Nat Commun* **2**,465 (2011).
28. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**,25-29 (2000).
29. Gene Ontology, C. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224** (2023).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**,27-30 (2000).
31. Saenz-Narciso, B., Bell, S.E., Matheson, L.S., Venigalla, R.K.C. & Turner, M. ZFP36-family RNA-binding proteins in regulatory T cells reinforce immune homeostasis. *Nat Commun* **16**,4192 (2025).
32. Cui, X. *et al.* Regnase-1 and Roquin Nonredundantly Regulate Th1 Differentiation Causing Cardiac Inflammation and Fibrosis. *J Immunol* **199**,4066-4077 (2017).
33. Yu, D. *et al.* Roquin represses autoimmunity by limiting inducible T-cell co-stimulator messenger RNA. *Nature* **450**,299-303 (2007).
34. Lee, S.K. *et al.* Interferon-gamma excess leads to pathogenic accumulation of follicular helper T cells and germinal centers. *Immunity* **37**,880-892 (2012).
35. Li, M. *et al.* MCPIP1 down-regulates IL-2 expression through an ARE-independent pathway. *PLoS One* **7**,e49841 (2012).
36. Behrens, G. *et al.* A translational silencing function of MCPIP1/Regnase-1 specified by the target site context. *Nucleic Acids Res* **46**,4256-4270 (2018).
37. Schott, J. *et al.* Translational regulation of specific mRNAs controls feedback inhibition and survival during macrophage activation. *PLoS Genet* **10**,e1004368 (2014).
38. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**,D590-598 (2006).
39. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F. & Backofen, R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**,900-914 (2012).
40. Iwasaki, H. *et al.* The IkappaB kinase complex regulates the stability of cytokine-encoding mRNA induced by TLR-IL-1R by controlling degradation of regnase-1. *Nat Immunol* **12**,1167-1175 (2011).

41. Tse, K.M. *et al.* Enhancement of Regnase-1 expression with stem loop-targeting antisense oligonucleotides alleviates inflammatory diseases. *Sci Transl Med* **14**,eabo2137 (2022).
42. Wang, L.H. & Baker, N.E. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. *Dev Cell* **35**,269-280 (2015).
43. Engel, I. & Murre, C. The function of E- and Id proteins in lymphocyte development. *Nat Rev Immunol* **1**,193-199 (2001).
44. Lyszkiewicz, M. *et al.* Limited niche availability suppresses murine intrathymic dendritic-cell development from noncommitted progenitors. *Blood* **125**,457-464 (2015).
45. Schmitt, T.M. & Zuniga-Pflucker, J.C. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity* **17**,749-756 (2002).
46. Schlenner, S.M. *et al.* Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus. *Immunity* **32**,426-436 (2010).
47. Hidaka, R., Miyazaki, K. & Miyazaki, M. The E-Id Axis Instructs Adaptive Versus Innate Lineage Cell Fate Choice and Instructs Regulatory T Cell Differentiation. *Front Immunol* **13**,890056 (2022).
48. Miyazaki, K. *et al.* The transcription factor E2A activates multiple enhancers that drive Rag expression in developing T and B cells. *Sci Immunol* **5** (2020).
49. Yamasaki, S. *et al.* Mechanistic basis of pre-T cell receptor-mediated autonomous signaling critical for thymocyte development. *Nat Immunol* **7**,67-75 (2006).
50. Takeuchi, A. *et al.* E2A and HEB activate the pre-TCR alpha promoter during immature T cell development. *J Immunol* **167**,2157-2163 (2001).
51. Bain, G. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79**,885-892 (1994).
52. Ikawa, T., Kawamoto, H., Wright, L.Y. & Murre, C. Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* **20**,349-360 (2004).
53. Zhuang, Y., Soriano, P. & Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* **79**,875-884 (1994).
54. Jaleco, A.C. *et al.* Genetic modification of human B-cell development: B-cell development is inhibited by the dominant negative helix loop helix factor Id3. *Blood* **94**,2637-2646 (1999).
55. Kee, B.L., Rivera, R.R. & Murre, C. Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF-beta. *Nat Immunol* **2**,242-247 (2001).
56. Sun, X.H. Constitutive expression of the Id1 gene impairs mouse B cell development. *Cell* **79**,893-900 (1994).
57. Hoefig, K.P. *et al.* Defining the RBPome of primary T helper cells to elucidate higher-order Roquin-mediated mRNA regulation. *Nat Commun* **12**,5208 (2021).
58. Tants, J.N. *et al.* Roquin exhibits opposing effects on RNA stem-loop stability through its two ROQ domain binding sites. *Proc Natl Acad Sci U S A* **122**,e2424434122 (2025).

59. Jhunjhunwala, S. *et al.* The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**,265-279 (2008).
60. Miyazaki, K. & Miyazaki, M. The Interplay Between Chromatin Architecture and Lineage-Specific Transcription Factors and the Regulation of Rag Gene Expression. *Front Immunol* **12**,659761 (2021).
61. Cochrane, S.W., Zhao, Y., Perry, S.S., Urbaniak, T. & Sun, X.H. Id1 has a physiological role in regulating early B lymphopoiesis. *Cell Mol Immunol* **8**,41-49 (2011).
62. Galloway, A. *et al.* RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. *Science* **352**,453-459 (2016).
63. Inoue, T. *et al.* CNOT3 contributes to early B cell development by controlling Igh rearrangement and p53 mRNA stability. *J Exp Med* **212**,1465-1479 (2015).
64. Iwai, N. *et al.* UPF1 plays critical roles in early B cell development. *Nat Commun* **15**,5765 (2024).
65. Uehata, T. *et al.* Regulation of lymphoid-myeloid lineage bias through regnase-1/3-mediated control of Nfkbiz. *Blood* **143**,243-257 (2024).
66. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**,207-210 (2002).
67. Ventura, A. *et al.* Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**,661-665 (2007).
68. Sledzinska, A. *et al.* TGF-beta signalling is required for CD4(+) T cell homeostasis but dispensable for regulatory T cell function. *PLoS Biol* **11**,e1001674 (2013).
69. Rehage, N. *et al.* Binding of NUFIP2 to Roquin promotes recognition and regulation of ICOS mRNA. *Nat Commun* **9**,299 (2018).
70. Mudge, J.M. *et al.* GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* **53**,D966-D975 (2025).
71. Benoit Bouvrette, L.P., Bovaird, S., Blanchette, M. & Lecuyer, E. oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res* **48**,D166-D173 (2020).
72. Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53**,D1243-D1249 (2025).
73. Kinsella, R.J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**,bar030 (2011).
74. Altenhoff, A.M. *et al.* OMA orthology in 2024: improved prokaryote coverage, ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA Ecosystem. *Nucleic Acids Res* **52**,D513-D521 (2024).
75. Baldarelli, R.M. *et al.* Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227** (2024).
76. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**,841-842 (2010).

77. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**,26 (2011).

78. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,15-21 (2013).

79. Yates, A.D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**,D682-D688 (2020).

Figures

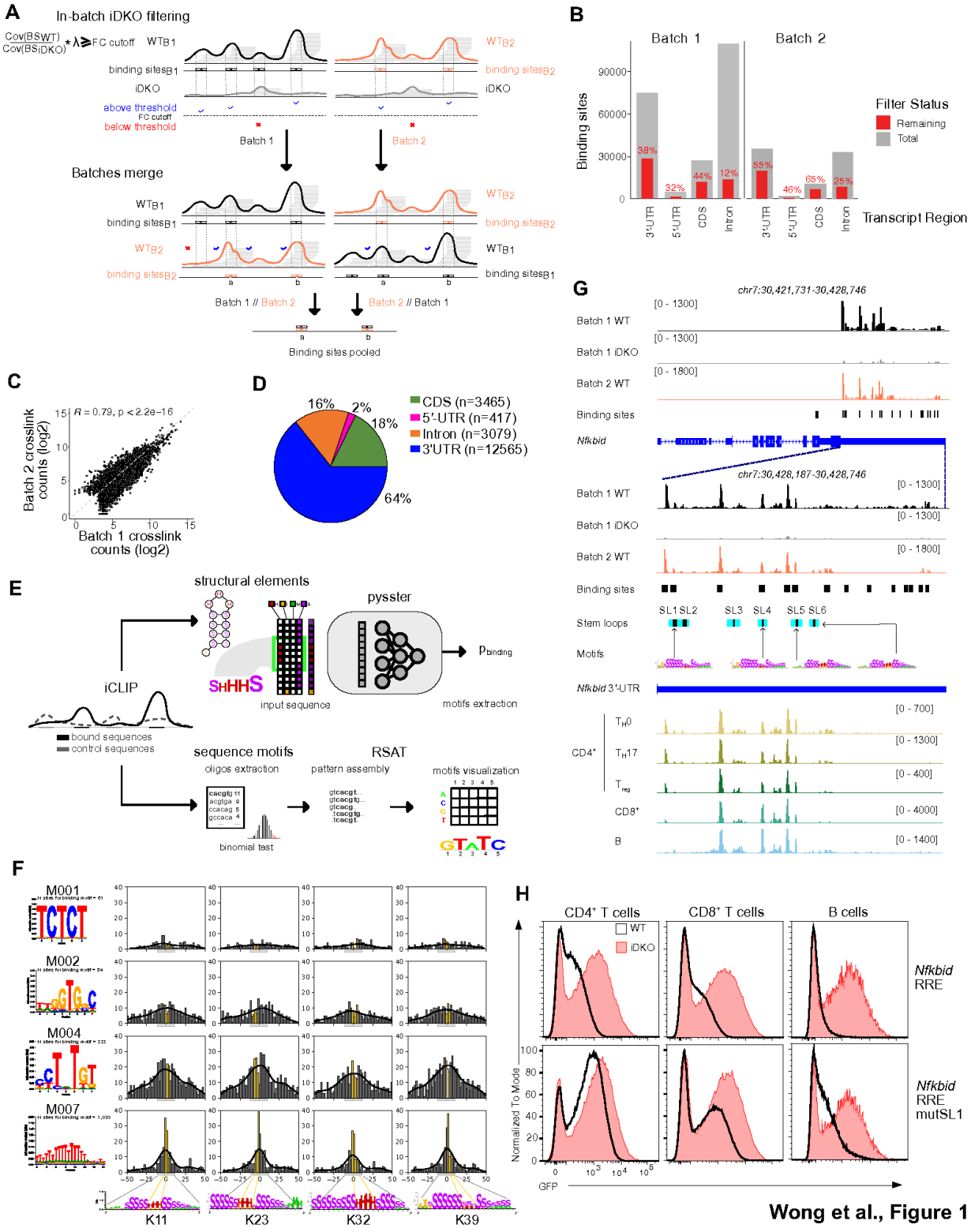


Figure 1

Sequence and structure determinants of Roquin-1 binding to the CD4+ T cell transcriptome. (a) Pipeline for establishing a robust set of binding sites across experiments. (**b**) Proportions of binding sites annotated with protein coding biotypes in total and remaining after application of iDKO filter in each of the two batches. (**c**) Correlation of cross-link counts at the final binding sites between the two batches. Correlation measured with the Pearson coefficient R=0.79; p<2.2e-16. (**d**) Number of binding sites and their biotype annotations with relative proportions. (**e**) Pipelines of analyses for sequence (bottom) and structural (top) motif identification. (**f**) Co-occurrence analysis with relative positions between structure and primary sequence motifs for four of the sequence motifs and the top scoring structural motifs. Rows represent primary motifs and columns structural elements. Each density plot displays the frequency counts of the primary motif with respect to the structural annotation, at the center. Light grey bottom bars indicate the position of the structural string, bars highlighted in yellow indicate the hairpin; abbreviations: (H: hairpin, I: internal loop, M: multiloop, S: stem). (**g**) Igv browser of *Nfkbid* (top: whole gene, bottom: 3´-UTR), with visualized iCLIP2 signal in $CD4^+$ T cells, binding sites, previously annotated stem loops responsible for regulation[9], identified structure and sequence motifs, and iCLIP2 signal from lymphocyte subsets and lineages, as indicated. (**h**) $CD4^+$, $CD8^+$ and B cells of *Rc3h1/2*$^{fl/fl}$;Cre-ERt2 mice were transduced with GFP reporters, inserting the *Nfkbid* RRE, encompassing nt 1-263 of the 3´-UTR, downstream of the GFP-CDS without or with a mutation in the first stem loop 1 (mutSL1) [9]. After transduction, the cells were treated with 4´-hydroxy-tamoxifen (iDKO) or left untreated (WT). The GFP fluorescence was analyzed after cultivation for 3 (T cells) or 4 days (B cells). To focus the analysis of GFP fluorescence intensity on cells that underwent deletion of floxed alleles, iDKO B cells or iDKO $CD8^+$ or iDKO $CD4^+$ T cells were pregated for IkB_{NS}^{high} or $Icos^{high}$ expression, respectively. Data in (**h**) are representative of 3 independent experiments.

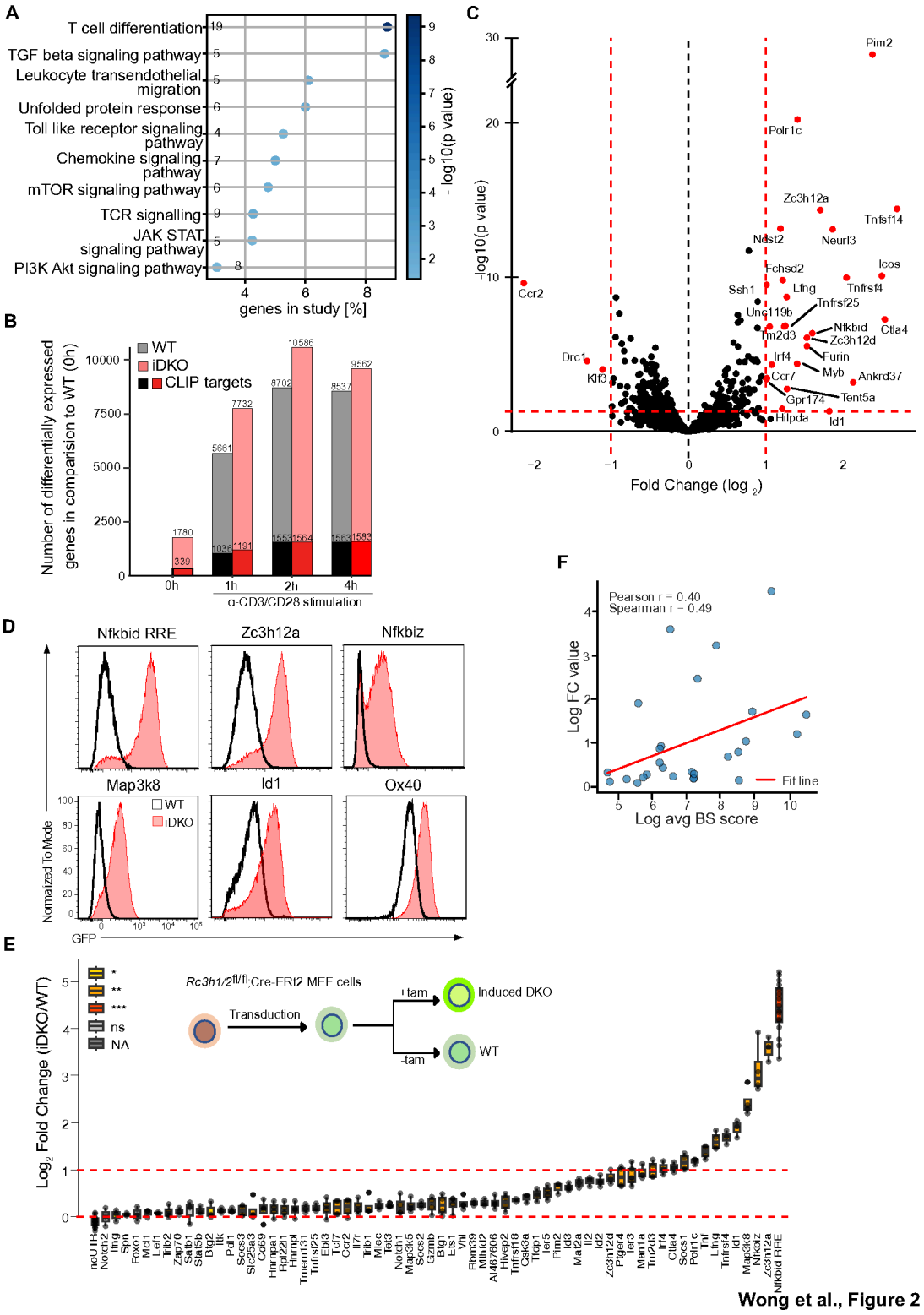
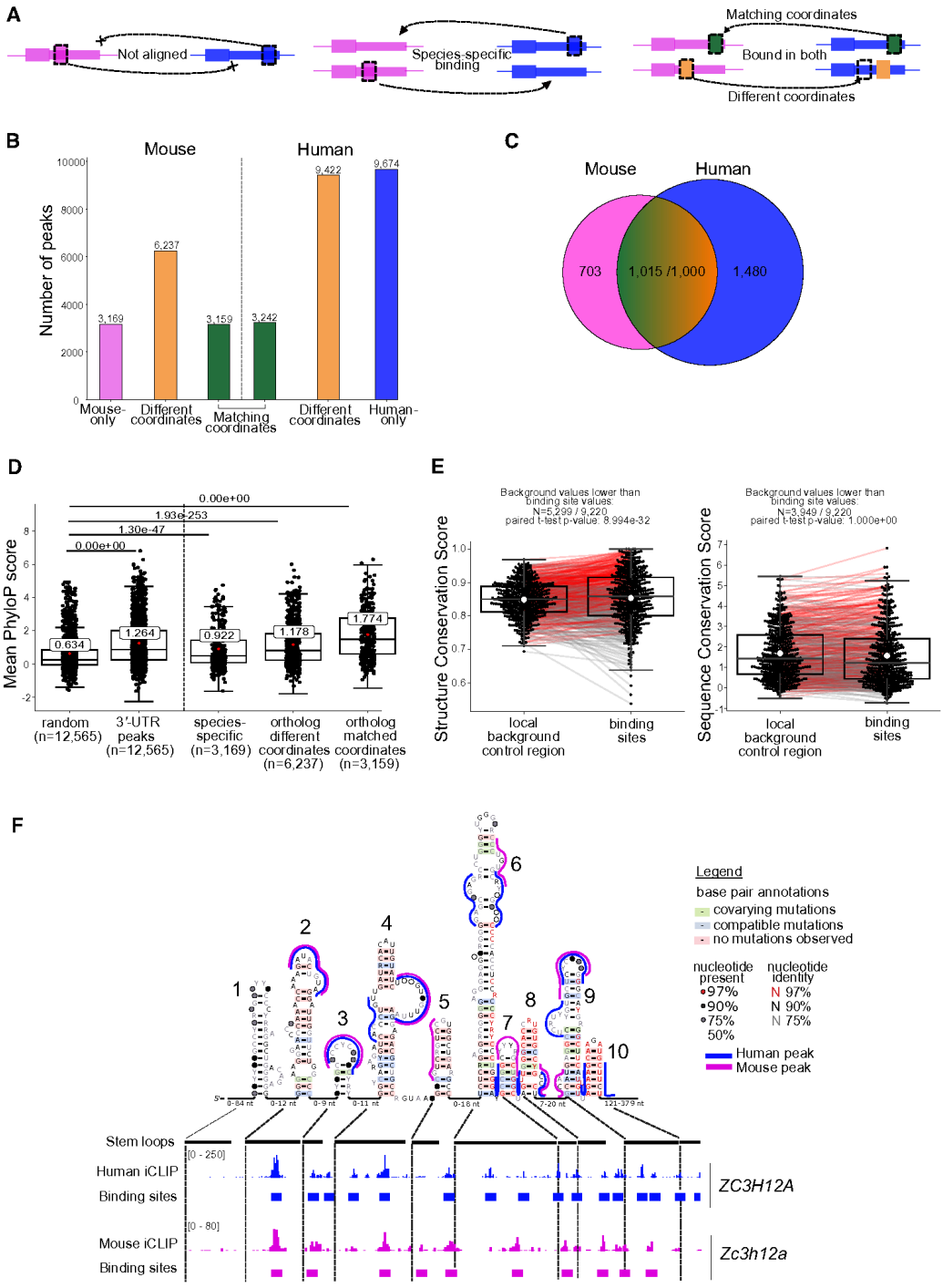


Figure 2

Roquin targeted 3′-UTRs encode different regulatory capacities. (**a**) Pathway enrichment analysis of the top 10% of 3′-UTR targets using T-cell-targeted KEGG and GO terms. Rows are sorted by gene enrichment, reported on the x-axis. Only significant terms are shown, significance levels are color-coded. (**b**) CD4$^+$ T cells from *Cd4*-Cre-ERt2 (WT) and *Rc3h1/2*$^{fl/fl}$;*Cd4*-Cre-ERt2 (iDKO) mice after *in vitro* 4′-hydroxy-tamoxifen treatment were activated with anti-CD3/CD28 under T_H1 conditions and expanded for

2 days. Cells without (0h) or with anti-CD3/28 restimulation for 1h, 2h or 4h were subjected to mRNA sequencing. Number of significantly differentially expressed genes (DEGs, log2(FC)≥1, false discovery rate value ≤0.05) in WT and iDKO $CD4^+$ T cells at the indicated times of stimulation compared to unstimulated WT cells. (**c**) Volcano plot displaying mRNA expression in fold change of iDKO/WT for all iCLIP2 bound genes. (**d,e**) *Rc3h1/2*$^{fl/fl}$;Cre-ERt2 MEF cells were retrovirally transduced with GFP reporter constructs fused to the 3´-UTRs of the indicated transcripts followed by treatment of cells with (iDKO) or without (WT) 4´-hydroxy-tamoxifen. GFP expression changes are depicted as histogram overlays (**d**) or as fold change of iDKO/WT of median MFI values **(e)**. (**f)** Correlation of average binding site score (PureCLIP) vs fold change values for reporter genes. Pearson and Spearman coefficients are reported. The regression line is displayed in red. Data are representative of 3 independent experiments (**d**) or 3-14 independent experiments (**e**). (**e**) Statistical significance was calculated by one-tailed Mann-Whitney test; ns = not significant, $p < 0.001$ = ***, $0.001 < p < 0.01$ = ** and $0.01 < p < 0.05$ = *. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure.

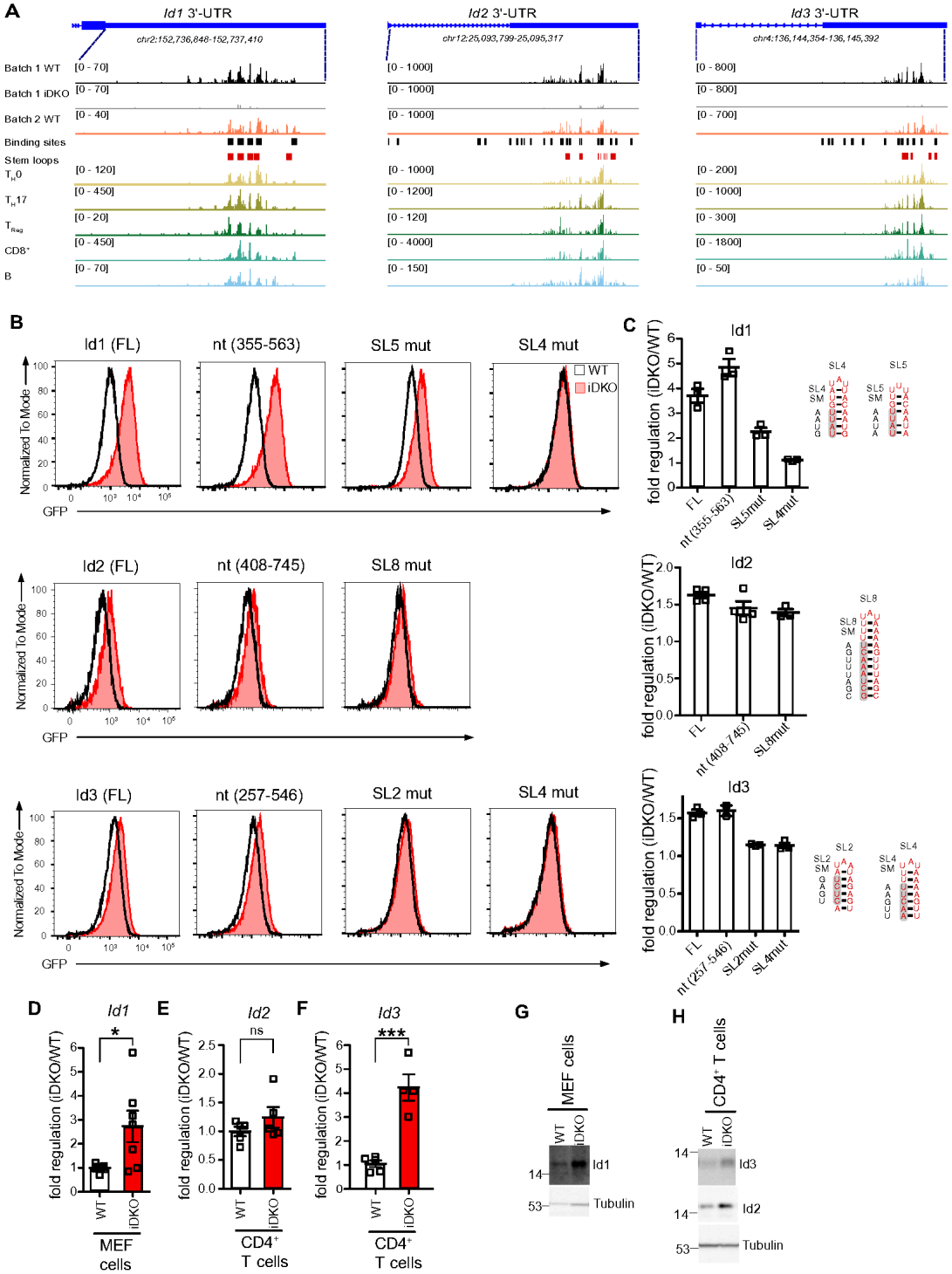


Wong et al., Figure 3

Figure 3

Comparative analysis of ROQUIN-1 binding to mouse and human 3′-UTRs. (**a**) Annotation of binding sites from human and mouse datasets, considering orthologous genes and their 3′-UTRs. (**b**) Barplot of the counts of human and mouse binding sites following the nomenclature from (**a**), separating species specific from orthologous binding, and further separating binding sites with matching genomic coordinates from others. (**c**) Venn diagram showing counts of species-specific and bound-in-common

genes considering protein coding genes with at least one binding site in their 3′-UTR. Gene counts for shared binding differ per species owing to non-unique orthology relationships. (**d**) Sequence conservation of mouse 3′-UTR binding sites, separated or not on their orthologous binding status as defined in (**b**). Control regions were sampled from 3′-UTRs presenting at least 1 binding site, excluding bound regions. (**e**) Structure conservation (left) or sequence conservation (right) scores of mouse 3′-UTR binding sites. RNA secondary structures were investigated by mLocARNA alignment of up to 60 vertebrates. Structure conservation was defined as the proportion of aligned nucleotides compatible with base-pairing, irrespective of sequence conservation. PhyloP60w sequence conservation scores were extracted for positions of both, binding sites and background positions. Conservation was evaluated at paired positions in the consensus structure applied to the mouse sequence, separating positions corresponding to Roquin-1 binding from background by pairing positions within 200 nucleotides intervals around binding sites. (**f**) LocARNA predicted RNA secondary structures and experimentally measured iCLIP2 signals for ROQUIN-1/Roquin-1 in human and mouse in the 3′-UTR sequence of ZC3H12A/Zc3h12a, respectively. Binding sites were annotated onto the secondary structures of the predicted stem-loops.

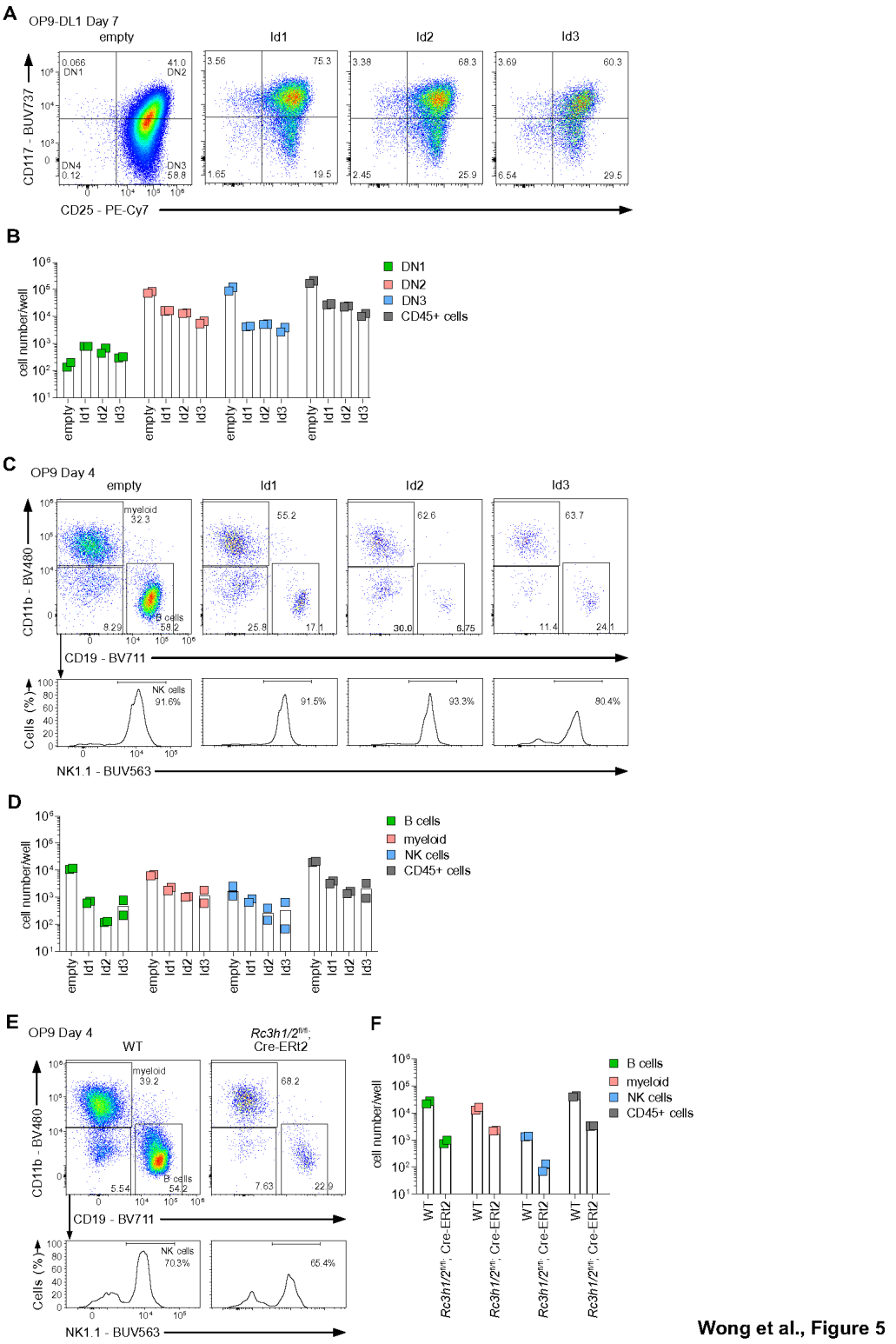


Wong et al., Figure 4

Figure 4

Roquin binds and regulates mRNAs of the Id family. (**a**) Igv browser screenshots of *Id1* (left panel), *Id2* (middle panel) and *Id3* (right panel) 3'-UTRs with visualized iCLIP2 signal in $CD4^+$ T cells, binding sites, predicted stem loops according to LocARNA analysis (**Supplemental data Fig. S5c-e**), and iCLIP2 signal from lymphocyte subsets and lineages, as indicated. **(b-c)** Flow cytometry analysis of GFP expression in *Rc3h1/2*$^{fl/fl}$;Cre-ERt2 MEF cells retrovirally transduced with the indicated GFP-3'-UTR reporter with the full

length (FL), truncated or mutated version of *Id1* (top row), *Id2* (middle row) or *Id3* (bottom row) 3´-UTRs without (WT) or with 4´-hydroxy-tamoxifen (iDKO) treatment. Depicted are exemplary histograms (**b**) and bar diagrams of fold regulation of the GFP reporter levels based on median MFI values in iDKO cells relative to WT cells (**c**). Schematic representation in **(c)** shows conserved stem-loop structures within the 3´-UTR of mouse Id1 (top), Id2 (middle), Id3 (bottom) with indication of introduced stem mutations (SM: disruption of base-pairing by mutation of complementary nucleotides in the 5´stem) used in (**b**). qPCR (**d-f**) or WB results (**g, h**) of the indicated Id family members analyzed in MEF or $CD4^+$ T cells from *Rc3h1/2*$^{fl/fl}$; Cre-ERt2 or *Rc3h1/2*$^{fl/fl}$;*Cd4*-Cre-ERt2 mice, respectively, without (WT) or with 4´-hydroxy-tamoxifen (iDKO) treatment. Data (**b, g-h**) are representative of 3-5 independent experiments or (**c-f**) presented as mean ± SEM of 3-5 independent experiments (**d-e**). Statistical significance was calculated using unpaired student's t- test (two-tailed) ($p < 0.001$ = ***, $0.001 < p < 0.01$ = ** and $0.01 < p < 0.05$ = *, ns: not significant).



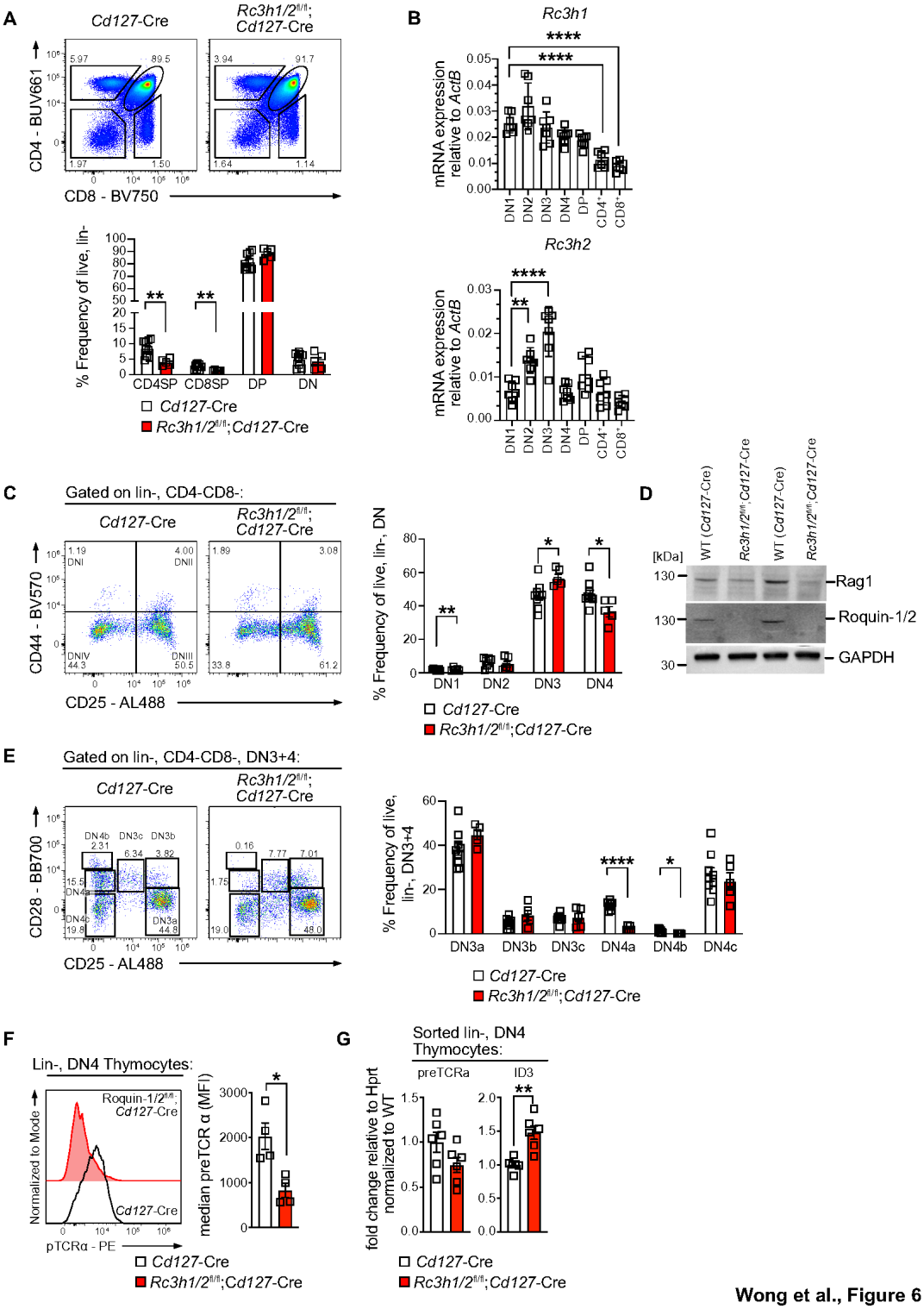
Wong et al., Figure 5

Figure 5

Id protein overexpression or Roquin inactivation interferes with lymphocyte development.

(**a**) Common lymphoid progenitor (CLP) cells were retrovirally transduced to overexpress Id1, Id2 or Id3 open reading frames and were co-cultured with OP9-DL1 (**a-b**) or OP9 (**c-d**) cells. Progression of thymocyte development was investigated with CD117 and CD25 markers indicating DN2

(CD117hi/CD25hi) and to DN3 (CD117low/CD25hi) transition stages. (**b**) Absolute numbers of cells recovered from OP9-DL1 CLP co-cultures at day 7. (**c-d**) CD11b and CD19 markers were used to discriminate myeloid (CD11b^{hi}) or B cell development (CD19^{hi}), while NK1.1 was used to characterize NK cells in the CD11b^{lo}CD19^{lo} gate. (**d**) Quantification of absolute numbers from cells shown in (**c**). (**e-f**) Common lymphoid progenitors of *Rc3h1/2*$^{fl/fl}$; Cre-ERt2 mice were treated with 4´-hydroxy-tamoxifen to induce deletion of Roquin- 1/2 at the CLP stage. Cells were co-cultured with OP9 cells to analyze B lineage development as shown in (**c-d**). (**a,c,e**)The data are representative of two experiments.



Wong et al., Figure 6

Figure 6

Inactivation of Roquin in CLPs impairs thymocyte development.

(a) Flow cytometry plots and quantification of CD4 and CD8 populations in the thymus of 4-6 weeks old *Cd127*-Cre and *Rc3h1*$^{fl/fl}$*;Rc3h2*$^{fl/fl}$*;Cd127*-Cre mice (n ≥ 5, representing at least three independent experiments; bars show mean ± SEM, unpaired student's T-test, two-tailed). **(b)** *Rc3h1* and *Rc3h2* expression in different stages of thymocyte development in WT mice (n = 7, mean ± SEM, one-way ANOVA). **(c)** Flow cytometry plots and quantification of the main stages of negative selection in the thymus (n ≥ 5, representing at least three independent experiments; mean ± SEM, unpaired student's T test, two-tailed). **(d)** Western blot analysis of Rag-1 expression in total thymocytes of 4-6 weeks old *Cd127*-Cre and *Rc3h1*$^{fl/fl}$*;Rc3h2*$^{fl/fl}$*;Cd127*-Cre mice (n = 3 in two independent experiments). **(e)** Flow cytometry plots and quantification of DN3 to DN4 transition in the thymus of 4-6 weeks old *Cd127*-Cre and *Rc3h1*$^{fl/fl}$*; Rc3h2*$^{fl/fl}$*;Cd127*-Cre mice (n ≥ 5, representing at least three independent experiments; mean ± SEM, unpaired student's T test, two-tailed). **(f)** Flow cytometry plots and quantifications of pre TCRα staining on DN4 thymocytes of 4-6 weeks old *Cd127*-Cre and *Rc3h1*$^{fl/fl}$*;Rc3h2*$^{fl/fl}$*;Cd127*-Cre mice (n = 4 in two independent experiments, mean ± SEM, unpaired student's T test, two-tailed). **(g)** qPCR analysis of sorted DN4 thymocytes from of Cre-Ert2 and *Rc3h1*$^{fl/fl}$*;Rc3h2*$^{fl/fl}$*;*Cre-Ert2 fed with tamoxifen (n = 6, representing two independent experiments; mean fold change ± SEM, unpaired student's T test, two-tailed).

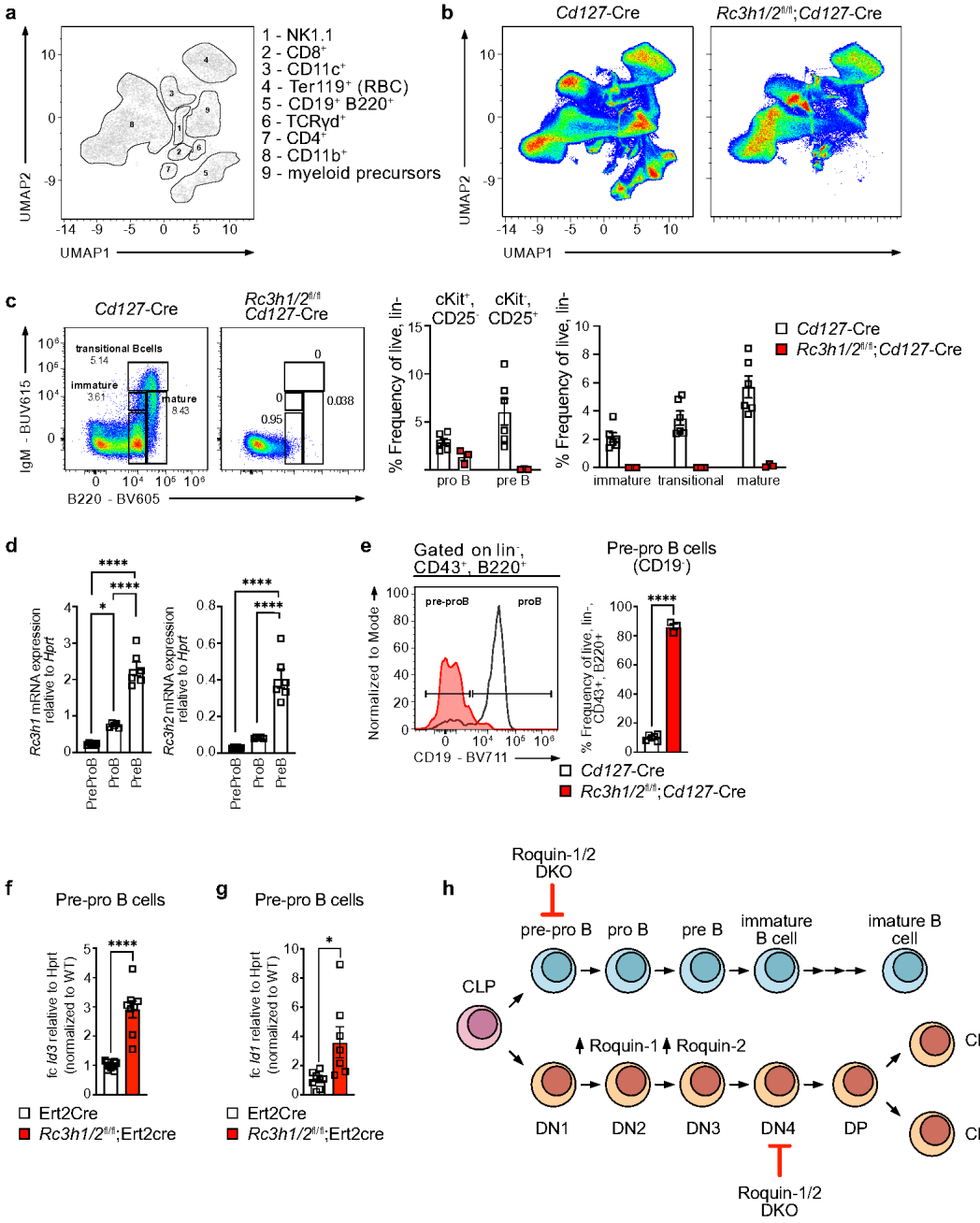


Figure 7

Inactivation of Roquin in CLPs blocks B cell development.

(a) UMAP representation of distinct cell populations in the bone marrow. **(b)** UMAP analysis of cell populations in the bone marrow of 4-6 weeks old *Cd127*-Cre and *Rc3h1*$^{fl/fl}$;*Rc3h2*$^{fl/fl}$;*Cd127*-Cre mice (n ≥ 3, representing at least three independent experiments). **(c)** Representative plots and quantification of

the stages of B cell development in the bone marrow of of 4-6 weeks old *Cd127*-Cre and Roquin-1$^{fl/fl}$; Roquin-2$^{fl/fl}$; *Cd127*-Cre mice (n ≥ 5, representing at least three independent experiments; mean ± SEM, unpaired student's T test, two-tailed). **(d)** *Id3* mRNA expression in pre-pro, pro and pre – B cells sorted from WT mice. **(e)** Representative plots and quantification of pre-pro and pro B cells in the bone marrow (n ≥ 5, representing at least three independent experiments; mean ± SEM, unpaired student's T test, two-tailed). **(f,g)** qPCR analysis of sorted pre-pro B cells from the bone marrow of *Cre-Ert2* and Roquin-1$^{fl/fl}$; Roquin-2$^{fl/fl}$; *Cre-Ert2* fed with tamoxifen (n = 6, representing two independent experiments; mean fold change ± SEM, unpaired student's T test, two-tailed). **(h)** Schematic representation of B cell development in the bone marrow and T cell development in the thymus, with indicated developmental block in the absence of Roquin-1/2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2250516FigureS2.pdf
- 3250516FigureS3.pdf
- 7250516FigureS7.pdf
- TableS3Motifscharacterization.xlsx
- 6250516FigureS6.pdf
- TableS5SpeciesLocarnaanalysis.xlsx
- 8250516FigureS8.pdf
- TableS2Roquintargetsinthedifferentcelllines.xlsx
- TableS6FACSAntibodies.xlsx
- TableS1overviewclipexperiments.xlsx
- 4250516FigureS4.pdf
- TableS4Roquinbindingsitesandorthologies.xlsx
- 1250516FigureS1.pdf
- 5250516FigureS5.pdf
- SupplementaldataFigures.docx