1

A global map of RNA binding protein occupancy guides functional dissection of post-transcriptional regulation of the T cell transcriptome

Authors: Adam J. Litterman[1], Wandi S. Zhu[1], Robin Kageyama[1], Wenxue Zhao[2,3], Noah Zaitlen[2], David J. Erle[2], K. Mark Ansel[1*]

Affiliations:

[1] Department of Microbiology & Immunology and Sandler Asthma Basic Research Center University of California San Francisco, San Francisco, CA

[2] Department of Medicine and Lung Biology Center, University of California San Francisco, San Francisco, CA

[3] School of Basic Medicine (Shenzhen), Sun Yat-Sen University, Guangzhou 510080, People's Republic of China

*For correspondence: Mark.Ansel@ucsf.edu

Abstract: RNA binding proteins (RBPs) mediate constitutive RNA metabolism and gene specific regulatory interactions. To identify RNA cis-regulatory elements, we developed GCLiPP, a biochemical technique for detecting RBP occupancy transcriptome-wide. GCLiPP sequence tags corresponded with known RBP binding sites, specifically correlating to abundant cytosolic RBPs. To demonstrate the utility of our occupancy profiles, we performed functional dissection of 3' UTRs with CRISPR/Cas9 genome editing. Two RBP occupied sites in the CD69 3' UTR destabilized the transcript of this key regulator of lymphocyte tissue egress. Comparing human Jurkat T cells and mouse primary T cells uncovered hundreds of biochemically shared peaks of GCLiPP signal across homologous regions of human and mouse 3' UTRs, including a cis-regulatory element that governs the stability of the mRNA that encodes the proto-oncogene PIM3 in both species. Our GCLiPP datasets provide a rich resource for investigation of post-transcriptional regulation in the immune system.

Introduction

The life cycle of protein coding RNA transcripts involves their transcription from DNA,

5′ capping, splicing, 3′ polyadenylation, nuclear export, targeting to the correct cellular

compartment, translation and degradation (Beelman and Parker, 1995; Martin and Ephrussi,

2009; Reed, 2003). RNA binding proteins (RBPs) coordinately regulate these processes through

interaction with RNA cis-regulatory elements, often in the 5′ and 3′ untranslated regions (UTRs)

35   whose sequences are not constrained by a functional coding sequence (Keene, 2007).

36   Mammalian genomes encode hundreds of RBPs (Castello et al., 2012), and mutations in

37   individual RBPs or even individual binding sites can induce strong developmental, autoimmune

38   and neurological defects in human patients and mouse models (Bassell and Kelic, 2004; Kafasla

39   et al., 2014; Schwerk and Savan, 2015). As much as half of the extensive gene expression

40   changes that occur during T cell activation occur post-transcriptionally (Raghavan et al., 2002),

41   and several RBPs are known to be critical determinants of immune function and homeostasis

42   (Kafasla et al., 2014).

43   Methods like DNase I hypersensitivity and ATAC-seq that query regulatory element

44   accessibility and occupancy without prior knowledge of their protein binding partners have

45   proven themselves as powerful techniques for the systematic mapping of *cis*-regulatory

46   sequences in DNA (Buenrostro et al., 2013; Thurman et al., 2012). Their development has

47   allowed for comparisons in the regulatory structure of diverse cell types (Corces et al., 2016) and

48   across the tree of life (Villar et al., 2015; Wilson et al., 2008). A lack of analogous systematic

49   methods for mapping the transcriptome's cis-regulatory landscape has limited our understanding

50   of post-transcriptional regulatory circuits and the evolution of untranslated regions of transcribed

51   genes.

52   Current methods for regulatory element identification in RNA have focused on specific

53   *trans* factors (Lee and Ule, 2018), although more recent technologies have also analyzed

54   secondary structure (Spitale et al., 2015) and interaction with chromatin (Li et al., 2017)

55   transcriptome wide. Protein precipitation (Baltz et al., 2012) and chemical biotinylation of

56   proteins (Freeberg et al., 2013) have been used to analyze global RBP occupancy in cell lines

57   and yeast, respectively, but difficulty remains in defining RNA regulatory activity in a

58   systematic way. Here, we create global RBP occupancy maps for a human T cell line, Jurkat, and

59   primary mouse T cells. Comparing RBP occupancy for thousands of mRNAs across species

60   identified biochemically shared regulatory sites, which are enriched for phylogenetically

61   conserved sequences. Finally, we used a scalable system of CRISPR dissection to define regions

62   of functional activity in 3' UTRs of mouse and human transcripts of immunological importance.

63   Biochemically derived maps of RBP occupancy are a powerful tool for the interrogation of post-

64   transcriptional gene regulation in the immune system.

65

66   **Results**

67   <u>Transcriptome-wide analysis of RBP occupancy in T cells in two species</u>

68   To achieve transcriptome-wide RBP binding site profiling, we developed a protocol for

69   Global Cross-linking Protein Purification (GCLiPP) suitable for use in mammalian cells and

70   applied this technique in human Jurkat T cells and cultured primary mouse T cells (Figure 1A).

71   GCLiPP is an adaptation of previously described biochemical methods for crosslinking

72   purification of all mRNA-RBP complexes. The key features of GCLiPP include: crosslinking of

73   endogenous ribonucleoprotein complexes using high energy UV light (no photo-crosslinkable

74   ribonucleotide analogues); oligo-dT pulldown prior to biotinylation to enrich for mRNA species;

75   chemical biotinylation of primary amines using a water soluble reagent with a long, flexible

76   linker; brief RNase digestion with RNase T1; and on-bead linker ligation with radiolabeled 3′

77   linker to facilitate downstream detection of ligated products. We used the guanine specific

78   ribonuclease T1 to favor larger average fragment sizes than using an RNA endonuclease with no

79   nucleotide specificity (such as RNase A) and ligated RBP protected fragments into a small RNA

80   sequencing library.

3

81   We called local peaks of GCLiPP sequence read density and measured the distribution of

82   GCLiPP reads within those peaks to assess the reproducibility of the technique. Local read

83   density within individual transcripts was similar between experiments, as GCLiPP fragments

84   yielded highly reproducible patterns in technical replicates (e.g. comparing replicate Jurkat T cell

85   samples, Figure 1B) and across multiple pooled experiments (e.g. comparing CD4+ and CD8+ T

86   cells, Figure 1C). A similar distribution of transcriptome features constituted GCLiPP libraries

87   from both Jurkat and primary T cells, with read coverage strongly enriched within mature

88   mRNAs and long non-coding RNAs (Figure 1D,E). The most striking difference was the greater

89   proportion of reads derived from transposable elements in mouse GCLiPP libraries. This

90   increase is likely due to the greater amount of annotated transposable elements in the mouse

91   genome since the relative coverage of these elements was similar between species.

92

93   GCLiPP read density represents cytosolic RBP occupancy

94   To validate GCLiPP, we systematically examined the relationship between GCLiPP

95   occupancy profiles in human Jurkat cells and enhanced cross-linking immunoprecipitation

96   (eCLIP) analyses of specific RBP binding profiles in K562 cells from the Encyclopedia of DNA

97   Regulatory Elements (ENCODE) project (Sundararaman et al., 2016). We examined pairwise

98   correlations of normalized read density across individual 3' UTRs between GCLiPP and

99   individual RBP eCLIP samples to identify contributions of each RBP, and also compared

100   GCLiPP and the input control for each eCLIP experiment (Figure 2A). eCLIP for many RBPs,

101   such as TIA1 and IGF2BP1, more closely matched GCLiPP read density than the eCLIP control

102   input across the genome (Figure 2B). We also found RBPs, such as PUM2, that exhibited anti-

103   correlations with GCLiPP signal across individual transcripts. In these cases, we typically saw

104   focal RBP binding to specific sites within transcripts (such as UGUA motifs in the case of

105   PUM2) that, while represented in GCLiPP reads, did not dominate the GCLiPP signal (Figure

106   2A, bottom panel). Although the PUM2 eCLIP profile did not correlate to GCLiPP signal

107   genome wide, PUM2 binding sites were still overrepresented in GCLiPP data. This was revealed

108   when we called GCLiPP peaks with CLIPper (Lovci et al., 2013) and compared these peaks with

109   CLIPper called peaks in eCLIP datasets. The observed fraction of PUM2 eCLIP peaks that

110   overlap GCLiPP peaks (0.56) was much greater than the fraction overlapping eCLIP peaks

111   randomly shuffled across the 3' UTRs from which they were derived (Figure 2C, bottom panel).

112   Similar results were obtained for TIA-1 (Figure 2C, top panel) and IGF2BP1 (Figure 2C, middle

113   panel). These enrichments above background binding for IGF2BP1, TIA1 and PUM2 were

114   amongst the highest 8 of the 87 RBPs whose eCLIP signals were examined (Supplementary

115   Figure 1).

116   We performed genome wide correlation analysis for 87 RBPs obtained from eCLIP data,

117   and compared the correlation between eCLIP and GCLiPP with RBP abundance previously

118   determined via mass spectrometry (Baltz et al., 2012). There was an overall significant

119   correlation between RBP abundance and correspondence between RBP eCLIP and GCLiPP

120   profiles (r=0.28, p=.022). However, stratifying RBPs by their predominant cellular localization

121   (Binder et al., 2014) showed that this correlation was driven almost entirely by cytosolic RBPs

122   (Figure 2D). The fraction of eCLIP peaks that overlapped GCLiPP peaks above a shuffled

123   background was also significantly greater for cytosolic versus non-cytosolic RBPs (p=0.003,

124   Supplementary Figure 1 inset). These findings were expected, as the GCLiPP experimental

125   protocol preferentially samples the cytosol by eliminating most nuclear material. In summary, we

126   conclude that GCLiPP read density reflects transcriptome-wide cytosolic RBP occupancy.

127

128   RBP Occupancy of known RNA cis-regulatory elements in primary T cells

129   We examined the GCLiPP profiles at previously characterized cis-regulatory elements of

130   various functional and structural categories in primary mouse T cells. The canonical

131   polyadenylation signal AAUAAA is a known linear sequence motif that binds to a number of

132   RBPs in the polyadenylation complex, including CPSF and PABP (Millevoi and Vagner, 2009),

133   as part of constitutive mRNA metabolism. We examined T cell lineage-defining transcripts with

134   well-resolved GCLiPP profiles (due to their high expression levels), including *Cd3g* (Figure 3A),

135   *Cd3e, Cd4*, and *Cd8b1* (Supplementary Figure 2). The only canonical polyadenylation signal

136   sequences in these transcripts were contained within called GCLiPP peaks, often as the peak

137   with the highest GCLiPP read density in the entire transcript. Interestingly, the GCLiPP profile

138   of *Cd8b1* contained direct biochemical evidence for alternative polyadenylation signal usage

139   (Figure S2C), a phenomenon that has previously been described to be important in activated T

140   cells (Sandberg et al., 2008). GCLiPP peaks appeared in multiple canonical polyadenylation

141   signal sequences in *Cd8b1*, coincident with clear evidence for both short and long 3′ UTR

142   isoform usage indicated by lower RNAseq read counts after the initial canonical polyadenylation

143   signal. A similar pattern was also apparent in *Hif1a* (Figure S1D) and a number of other highly

144   expressed transcripts.

145   Known cis-regulatory elements involved in transcript localization were also represented

146   by local regions of GCLiPP read density. The Beta-actin "zipcode" element is responsible for

147   localization of *Actb* mRNA to the cellular leading edge in chicken embryo fibroblasts (Kislauskis

148   et al., 1994) and contains conserved linear sequence elements separated by a variable linker.

149   These conserved sequence elements are thought to form the RNA/protein contacts in a complex

6

150   involving the actin mRNA and the RNA binding protein Igf2bp1 (previously known as Zbp1)

151   where the non-conserved sequence winds around the RBP (Chao et al., 2010). This sequence

152   corresponds to the center of the second highest peak of GCLiPP read density in the *Actb*

153   transcript (Figure 3B). Some RBPs regulate the half-life and/or translation of the mRNAs that

154   they bind. The mRNA-destabilizing Roquin/Regnase binding site in the 3′ UTR of *Ier3* is a

155   straightforward example of this functional category of RNA/RBP interaction detected as a region

156   of GCLiPP read density (Figure 3C).

157   The insertion of the selenium containing amino acid selenocysteine into selenoproteins

158   represents a unique case of RBP regulation of protein translation. Selenoproteins are redox

159   enzymes that use selenocysteine at key reactive residues (Johansson et al., 2005; Papp et al.,

160   2007) Selenocysteine is encoded by the stop codon UGA, and this recoding occurs only in

161   mRNAs that contain 3′ UTR *cis*-regulatory elements (termed SECIS elements) that bind to RBPs

162   that recruit the elongation factor Eefsec and selenocysteine-tRNA (Berry et al., 1993; Tujebajeva

163   et al., 2000). SECIS elements were prominent peaks of GCLiPP read coverage in selenoprotein

164   mRNAs. For example, the predicted SECIS element (Mariotti et al., 2013) in the 3′ UTR of *Gpx4*

165   was entirely covered by GCLiPP reads (Figure 3D). Indeed, a canonical polyadenylation signal

166   and the full hairpin structure containing the SECIS element account for essentially all of the

167   GCLiPP reads in the *Gpx4* 3′ UTR (Figure 3E). Comparing transcriptome-wide in vivo folding

168   data from icSHAPE (Spitale et al., 2015) and GCLiPP data supports the identification of an RBP

169   bound, structured SECIS element (Figure 3F,G). Furthermore, this analysis suggests that the

170   folded, RBP bound structure is even larger than that predicted by SECISearch 3, with regions of

171   GCLiPP read density and apposed high and low icSHAPE signals spanning almost the entire 3′

172   UTR. Thus, GCLiPP recapitulated previously described cis-regulatory elements that mediate

7

173   constitutive RNA metabolism, transcript localization, regulation of gene expression, and

174   translation, including both structured elements and single-stranded RNA determinants.

175

176   <u>GCLiPP-guided CRISPR dissection of immune gene post-transcriptional regulation</u>

177   We then sought to use our GCLiPP RBP occupancy profiles to guide experimental

178   dissection of the post-transcriptional regulation of immunologically important transcripts. We

179   first focused on CD69, a cell surface C-type lectin protein transiently upregulated on T cells

180   early during activation. CD69 inhibits lymphocyte egress from lymphoid organs, and has been

181   implicated in a variety of other immune cell functions (Cibrián and Sánchez-Madrid, 2017). As

182   *CD69* mRNA is a labile transcript (Santis et al., 1995) we sought to identify cis-regulatory

183   elements in the 3'UTR that regulate stability. First, we designed guide RNAs (gRNA) targeting

184   nucleotide positions 57 and 784 downstream of the stop codon to generate large deletions in the

185   3'UTR by transfecting CRISPR-Cas9 ribonucleoprotein complex (crRNP) in Jurkat cells. From

186   the pool of transfected cells, we generated a clone (3'UTR Δ57-784) that contained two mutant

187   alleles with deletions that spanned positions 22 to 853 and 44 to 833 of *CD69* 3' UTR (Figure

188   4A). As a control, we generated a wildtype (WT) clone from Jurkat cells transfected with a

189   scrambled control gRNA crRNP. Homozygous deletion of most of the *CD69* 3' UTR led to

190   higher basal expression of CD69 protein (Figure 4B, left) and higher expression after stimulation

191   with PMA and Ionomycin (Figure 4B, right). Importantly, the *CD69* transcript in the 3'UTR

192   Δ57-784 clone decayed at a much slower rate, with a half-life of greater than 3 hours compared

193   to 0.36 hours in WT Jurkat cells after global transcriptional inhibition with actinomycin D

194   (Figure 4C). This effect was specific to *CD69* as the half-life of dual-specific phosphatase 2

195   (*DUSP2*), another labile transcript, was similar in WT and mutant clones (Figure 4C).These data

196   indicate that the *CD69* 3'UTR contains destabilizing cis-regulatory elements responsible for the

197   short half-life of the mRNA.

198   To determine whether RBP-occupied sites in the 3'UTR contain cis-regulatory elements

199   that regulate stability, we performed CRISPR-Cas9 dissection of the region (Zhao et al., 2017).

200   Using the GCLiPP profile as a guide, we designed 6 gRNAs along the 3'UTR, transfected them

201   as a crRNP pool into Jurkat cells, and (RT)-PCR amplified the CD69 3' UTR from genomic

202   DNA and RNA from transfected cells (Figure 4D). The dissection led to many distinct short and

203   long deletions (Figure 4E) that possessed destabilizing activity, indicated by a high RNA/gDNA

204   ratio relative to the predicted WT allele as measured by microcapillary gel electrophoresis

205   (Figure 4F). To determine whether certain RBP-occupied regions had greater destabilizing

206   activity than others, we sequenced amplicon fragments to measure the relative abundance of

207   transcripts containing the various deletions (Figure 4— source data 1), analyzed deletions

208   <250bp and calculated relative RNA/gDNA ratios along the 3'UTR. Our analysis revealed

209   varying deletion sizes (Figure 4G) and identified two regions with the highest destabilizing

210   activity that correspond with GCLiPP peaks (protein occupied regions PR1 and PR2 in Figure

211   4D,G-H). This pattern was replicated in duplicate experiments using different crRNP

212   concentrations (Figure 4H). Destabilizing activity was highly significantly concentrated in

213   regions PR1 and PR2. These findings demonstrate that GCLiPP can provide useful hypothesis-

214   generating data for identifying 3'UTR cis-regulatory elements that contribute to post-

215   transcriptional gene regulation.

216

217   Cross-species comparison of GCLiPP reveals patterns of biochemically shared post-

218   transcriptional regulation

219   Next, we sought to compare RBP occupancy in mouse and human T cells. To do so, we

220   performed Clustal Omega sequence alignments of thousands of human 3' UTRs and their

221   corresponding sequences in the mouse genome, and then designed an algorithm to identify

222   correlated peaks of normalized GCLiPP read density along the aligned nucleotides (Figure 5A).

223   Using this approach, we identified 1047 high-stringency biochemically shared GCLiPP peaks

224   derived from 901 3' UTRs (Supplementary table 1). As a class, biochemically shared peaks

225   exhibited significantly higher sequence conservation than the full 3' UTRs in which they reside

226   (Figure 5B). The highly conserved, biochemically shared peak in *USP25* exemplifies this general

227   pattern (Figure 5C, right panel). However, many biochemically shared peaks did not exhibit

228   corresponding increases in local sequence conservation. For example, the *ARRB2* mRNA that

229   encodes β-arrestin, another regulator of T cell migration in response to chemoattractant gradients

230   (Fong et al., 2002), exhibited a common peak of RBP occupancy in Jurkat cells and primary

231   mouse T cells that is roughly equally conserved as the rest of the 3' UTR (Figure 5C, left panel).

232   To examine which RBPs contributed to biochemically shared peaks more than other

233   GCLiPP peaks, we used HOMER motif calling software (Heinz et al., 2010) to identify enriched

234   motifs. Strikingly, of the six linear sequence motifs present in >10% of biochemically shared

235   peaks with $p \leq 10^{-10}$, five resemble well-known regulatory sequences (Figure 5D). The two most

236   common appeared to represent canonical CELF (Timchenko et al., 1996) and PUM (Hafner et

237   al., 2010) binding motifs. Three other identified motifs corresponded to runs of homo-polymers:

238   An A-rich motif that resembled the canonical polyadenylation signal (Proudfoot, 2011); a poly-U

239   containing motif similar to a sequence that has long been known to stabilize mRNAs (Zubiaga et

240   al., 1995) and a poly-C containing motif similar to the C-rich RNAs bound by poly-C binding

241   proteins (Makeyev and Liebhaber, 2002). We used Metascape (Tripathi et al., 2015) to identify

242 categories of biologically related genes enriched among mRNAs that contained biochemically

243 shared GCLiPPpeaks (Figure 5E and Figure 5— source data 2). Interestingly, 3 of the 5 most

244 enriched categories were related to RNA regulation ("regulation of mRNA metabolism," "large

245 Drosha complex," "RNA splicing"), with the broad category "post-transcriptional regulation of

246 gene expression" also in the top 10. Thus, biochemically shared GCLiPP binding sites are

247 generally more well conserved than their local sequence context, are enriched for well-studied

248 RBP binding motifs, and occur preferentially in genes that encode proteins involved in post-

249 transcriptional gene regulation, suggestive of conserved autoregulatory gene expression

250 networks.

251   We hypothesized that functionally conserved destabilizing cis-regulatory elements could

252 be identified by examining biochemically shared GCLiPP peaks in 3' UTRs of labile transcripts.

253 To prioritize candidates, we computed Pearson correlation coefficients for the normalized

254 GCLiPP profiles of 3' UTRs of genes expressed in both Jurkat cells and primary mouse T cells

255 (Figure 6A, black histogram) and examined transcript instability by RNAseq analysis of primary

256 mouse T cells treated with actinomycin D (Figure 6A, red histogram). The proto-oncogene *PIM3*

257 emerged as an outstanding candidate with both strong interspecies GCLiPP correlation and very

258 high transcript instability. Alignment of the GCLiPP profiles of human and mouse *PIM3*

259 revealed a dominant shared peak of GCLiPP read density (Figure 6B). This peak corresponded to

260 a highly conserved region of the transcript that contains a G-quadruplex, followed by a putative

261 AU-rich element (ARE) and a CELF binding motif (Figure 6C). Another conserved region with

262 a G-quadruplex followed by a putative ARE appeared upstream of the biochemically shared

263 GCLiPP peak. We numbered these conserved regions CR1 and CR2 according to their order in

264 the 3' UTR, and hypothesized that CR2 would exert greater cis-regulatory activity than CR1,

11

265   given its RBP occupancy in both species and the relative lack of occupancy in CR1. To test this

266   hypothesis, we performed CRISPR dissections of both the human and mouse *PIM3* 3' UTRs

267   (Figure 6— source data 1). These analyses produced largely concordant patterns of post-

268   transcriptional cis-regulatory activity in the human and mouse 3' with the greatest significant

269   destabilizing effect corresponding to the shared region of GCLiPP read intensity covering the

270   CR2 element (Figure 6D-K). Consistent with this portrait of the entire 3' UTR, when we filtered

271   specifically for mutations that completely deleted either CR1 or CR2, we observed significantly

272   greater expression of transcripts derived from cells with CR2 deleted versus CR1 (Figure 6L,M).

273   Thus, PIM3 is a very unstable transcript with highly concordant RBP occupancy in human and

274   mouse. Functional dissection of the post-transcriptional regulatory landscape of this gene

275   revealed that this biochemical concordance between mouse and human cells is mirrored at a

276   functional level, with the most highly occupied region indicated by GCLiPP read density

277   corresponding to the most destabilizing region of the 3' UTR.

278

279   **Discussion**

280   Interconnected networks of bound RBPs and RNAs form a complex layer of post-

281   transcriptional regulation that affects all biological processes. Understanding these networks

282   remains one of the key challenges in deciphering how the genome encodes diverse cell identities

283   and behaviors. The outcomes of RNA/RBP interactions can be quite varied. RBP occupancy can

284   affect RNA biogenesis, decay, translation, localization, splicing, chemical modification and

285   editing. Developing a roadmap to understand the cis-regulatory elements in each gene will be

286   critical to full elucidation of the post-transcriptional biology of any given transcript. Biochemical

287   procedures like CLIP that utilize RBP-specific immunoprecipitation have facilitated decoding of

12

288   these networks for individual RBPs. However, the number of validated CLIP antibodies remains

289   much smaller than the number of RBPs which are known to associate with mature RNAs (Baltz

290   et al., 2012; Sundararaman et al., 2016). In addition, the relative occupancy of cis-regulatory

291   elements bound by different proteins cannot be directly compared using different pools of

292   immunoprecipitated material. Here, we adapted previously described methodologies to arrive at

293   a technique, GCLiPP, to provide such a global RBP occupancy roadmap, and applied this

294   technique to interrogate post-transcriptional cis-regulatory activity in human and mouse T cells.

295   Systematic comparison with eCLIP data for 87 individual RBPs (Sundararaman et al.,

296   2016) indicated that GCLiPP roughly represented a weighted average of all potential eCLIP

297   experiments for cytosolic RBPs. Presumably, this includes proteins that are not appreciated as

298   having RNA binding activity, and those for which no specific affinity reagents are currently

299   available. GCLiPP peaks overlapped eCLIP peaks at a frequency much greater than would be

300   expected by chance, and overall GCLiPP read density correlated with eCLIP read density in a

301   manner that corresponded with the relative abundance of a given RBP in purified cellular

302   mRNPs (Baltz et al., 2012). Nevertheless, the eCLIP peaks for some low abundance RBPs were

303   significantly enriched in GCLiPP profiles. The strongest correlations were observed for abundant

304   cytosolic RBPs, and the correspondence between eCLIP and GCLiPP was only apparent for

305   cytosolic, but not non-cytosolic, RBPs. This result was expected. COMPARTMENTS

306   annotations (Binder et al., 2014) indicate that most of the RBPs classified as non-cytosolic are

307   mainly located in the nucleus, and the GCLiPP protocol includes pelleting nuclei after a gentle

308   detergent based cellular lysis, followed by an enrichment for polyadenylated RNA. Both of these

309   steps would be predicted to selectively eliminate nuclear RNPs associated with primary

13

310 transcripts. Future iterations of GCLiPP could be modified to intentionally enrich for nuclear

311 RBPs to examine the regulatory landscape of mRNA biogenesis.

312 The GCLiPP datasets described here provide a rich resource for the annotation and

313 experimental dissection of cis-regulatory function in mRNAs. GCLiPP detected RBP occupancy

314 at many known cis-regulatory regions, including canonical polyadenylation signals and elements

315 that control mRNA localization, translation and stability, providing a biochemical correlate of

316 functional activity. In addition, we demonstrated that GCLiPP can guide discovery of novel cis-

317 regulatory elements. Dissection of the *CD69* 3'UTR revealed regions of destabilizing activity

318 that correspond to RBP occupied sites detected by GCLiPP. Our findings corroborate a previous

319 study that identified region PR2 as a potential destabilizing region that contained AREs (Santis et

320 al., 1995). We expanded on this work and further identified a new destabilizing region (region

321 PR1) that also exhibits GCLiPP evidence of RBP occupancy. Further assessment of these post-

322 transcriptional cis-regulatory regions may provide novel insights into immune cell biology.

323 CD69 is an important regulator of T cell differentiation (Martín et al., 2010), migration (Shiow et

324 al., 2006) and metabolism (Cibrian et al., 2016). Because of these effects, CD69 is considered a

325 potential target of therapeutic treatments for autoimmune and inflammatory disorders (González-

326 Amaro et al., 2013). Generation of post-transcriptionally modified T cells expressing high levels

327 of CD69 could provide insight into its role in T cell biology and inflammatory disorders.

328 We leveraged the matched datasets from similar cell types expressing many shared

329 transcripts to perform a cross species comparison of the post-transcriptional regulatory

330 landscape. As might be expected, the sequences of 3' UTR regions that appeared as peaks of

331 RBP occupancy in both species were in general more conserved than the full length 3' UTRs in

332 which they occurred. These biochemically shared peaks were enriched in well-known RBP-

14

333   binding cis-regulatory sequences including PUM motifs, CELF motifs and canonical

334   polyadenylation signals. Surprisingly, though, we also found clear biochemically shared peaks

335   with relatively poor sequence conservation. These regions retain RBP occupancy despite an

336   evident lack of strong selective pressure on their primary sequence, perhaps due to highly

337   degenerate and/or structural determinants of RBP occupancy. RNAs with conserved structure

338   and RBP binding but poorly conserved primary sequence have been reported before, and they are

339   enriched in gene regulatory regions (Seemann et al., 2017; Weinreb et al., 2016). Finally, we

340   noted that transcripts with biochemically shared peaks tended to encode proteins that were

341   themselves involved in post-transcriptional gene regulation. This pattern is consistent with

342   previous suggestions that auto-regulatory or multi-component feedback loops may be a

343   conserved mode of post-transcriptional gene regulation (Kanitz and Gerber, 2010).

344   Our dissection of the human *PIM3* and mouse *Pim3* 3'UTRs demonstrates the utility of

345   GCLiPP for decoding biochemically shared and functionally conserved post-transcriptional

346   regulation. The PIM family of serine/threonine kinases exert profound regulatory effects on

347   MYC activity, cap-dependent translation independent of MTOR, and BAD mediated antagonism

348   of apoptosis (Narlik-Grassow et al., 2014). Post-transcriptional regulation of PIM kinases is

349   important, as proviral integrations in the *Pim1* 3' UTR are highly oncogenic (Nawijn et al.,

350   2011). *Pim3* mRNA was abundant but highly labile in T cells, with a turnover rate in the top 2%

351   of expressed mRNAs. PIM family members contain multiple ARE like repeats of AUUU(A), but

352   the specific sequences responsible for rapid mRNA decay have not been described and cannot be

353   predicted from the primary sequence alone. The *PIM3* 3'UTR contains two phylogenetically

354   conserved regions with very similar predicted ARE sequences. Of these regions, we predicted

355   that greater regulatory activity would be exerted by the region with GCLiPP evidence for RBP

15

356   occupancy in both human and mouse cells. CRISPR dissection bore out this prediction in both

357   species. The inactive conserved region may be structurally inaccessible to RBP occupancy, or it

358   may be occupied and exert regulatory activity only in other cell types or signaling conditions.

359   These possibilities further highlight the utility of unbiased biochemical determination of RBP

360   occupancy for annotating the regulatory transcriptome. The datasets reported here will accelerate

361   the annotation of cis-regulatory elements operant in T cell transcripts. In general, GCLiPP can be

362   combined with other unbiased biochemical assays and genetic analyses to yield a roadmap for

363   the dissection of post-transcriptional regulatory networks.

364

365

366   **Materials and Methods**

367   <u>Cells</u>

368   Jurkat cells were grown in RPMI supplemented with fetal bovine serum (Omega). Primary CD4$^+$

369   and CD8$^+$ mouse T cells were isolated from C57BL/6J mouse peripheral lymph nodes and spleen

370   using positive and negative selection Dynabeads, respectively, according to the manufacturer's

371   instructions (Invitrogen). All mice were housed and bred in specific pathogen-free conditions in

372   the Animal Barrier Facility at the University of California, San Francisco. Animal experiments

373   were approved by the Institutional Animal Care and Use Committee of the University of

374   California, San Francisco. Cells were stimulated with immobilized biotinylated anti-CD3 (clone

375   2C11, 0.25 µg/mL, BioXcell) and anti-CD28 (clone 37.51, 1 µg/mL, BioXcell) bound to

376   Corning 10 cm cell culture dishes coated with Neutravidin (Thermo) at 10 µg/mL in PBS for 3 h

377   at 37 °C. Cells were left on stimulation for 3 days before being transferred to non-coated dishes

378   in T cell medium (Steiner et al., 2011) supplemented with recombinant human IL-2 (20 U/mL).

16

379   Th2 cell cultures were also supplemented with murine IL-4 (100 U/mL) and anti-mouse IFN-γ

380   (10 µg/mL). CD8 T cell cultures were also supplemented with 10 ng/mL recombinant murine IL-

381   12 (10 ng/mL). For re-stimulation, cells were treated with 20 nM phorbol 12-myristate 13-

382   acetate (PMA) and 1 µM ionomycin (Sigma) for 4 hours before harvest.

383

384   Measurement of mRNA Decay

385   Cells were stimulated with PMA and Ionomycin for 4 hours and then additionally treated with

386   Actinomycin-D (Sigma-Aldrich) at 5 µg/mL for an additional 0, 1, 2 or 4 hours. After treatment,

387   cells were lysed with Trizol LS (Life Technologies) and processed with Direct-zol ™ 96 well

388   RNA (Zymogen). RNA was quantified with an ND-1000 spectrophotometer (NanoDrop) and

389   reverse transcribed with SuperScript III First Strand Synthesis Kit (Invitrogen). Quantitative

390   PCR was performed in two separate experiments using SYBR Advantage qPCR Premix

391   (Clontech) on a Realplex 2S instrument (Eppendorf).

392

393   GCLiPP and RNAseq

394        $\sim100 \times 10^6$ mouse T cells cultured from 3 mice or $\sim100 \times 10^6$ Jurkat T cells were washed

395   and resuspended in ice cold PBS and UV irradiated with a 254 nanometer UV crosslinker

396   (Stratagene) in three doses of 4000 mJ, 2000 mJ and 2000 mJ, swirling on ice between doses.

397   Cells were pelleted and frozen at -80 °C. Thawed pellets were rapidly resuspended in 400 µL

398   PXL buffer without SDS (1X PBS with 0.5% deoxycholate, 0.5% NP-40, Protease inhibitor

399   cocktail) supplemented with 2000 U RNasin (Promega) and 10 U DNase (Invitrogen). Pellets

400   were incubated at 37 °C with shaking for 10 min, before pelleting of nuclei and cell debris

401   (17000 g for 5 min). Supernatants were biotinylated by mixing at room temperature for 30 min

17

402   with 500 µL of 10 mM EZ-Link NHS- SS-Biotin (Thermo) and 100 µL of 1 M sodium

403   bicarbonate. Supernatants were mixed with 1 mg of washed oligo-dT beads (New England

404   Biolabs) at room temperature for 30 min and washed 3 times with magnetic separation. Oligo-dT

405   selected RNA was eluted from beads by heating in poly-A elution buffer (NEB) at 65 °C with

406   vigorous shaking for 10 min. An aliquot of eluted RNA was treated with proteinase K and saved

407   for RNAseq analysis using Illumina TruSeq Stranded Total RNA Library Prep Kit according to

408   the manufacturer's instructions. Cells treated with Actinomycin-D as described above were also

409   collected for RNAseq to generate transcriptome wide measurements of transcript stability.

410   The remaining crosslinked, biotinylated mRNA-RBP complexes were captured on 250

411   µL of washed M-280 Streptavidin Dynabeads (Invitrogen) for 30 min at 4 °C with continuous

412   rotation to mix. Beads were washed 3 times with PBS and resuspended in 40 µL of PBS

413   containing 1000 U of RNase T1 (Thermo) for 1 min at room temperature. RNase activity was

414   stopped by addition of concentrated (10% w/v) SDS to a final concentration of 1% SDS. Beads

415   were washed successively in 1X PXL buffer, 5X PXL buffer and twice in PBS. 24 pmol of 3′

416   radiolabeled RNA linker was ligated to RBP bound RNA fragments by resuspending beads in 20

417   µL ligation buffer containing 10 U T4 RNA Ligase 1 (New England Biolabs) with 20% PEG

418   8000 at 37 degrees for 3 h. Beads were washed 3X with PBS and free 5′ RNA ends were

419   phosphorylated with polynucleotide kinase (New England Biolabs). Beads were washed 3X with

420   PBS and resuspended in ligation buffer containing 10 U T4 RNA Ligase 1, 50 pmol of 5′ RNA

421   linker and 20% PEG 8000 and incubated at 15 °C overnight with intermittent mixing. Beads

422   were again washed 3 times in PBS and linker ligated RBP binding fragments were eluted by

423   treatment with proteinase K in 20 µL PBS with high speed shaking at 55 °C. Beads and

424   supernatant were mixed 1:1 with bromophenol blue formamide RNA gel loading dye (Thermo)

18

425   and loaded onto a 15% TBE-Urea denaturing polyacrylamide gel (BioRad). Ligated products

426   with insert were visualized by autoradiography and compared to a control ligation (19 and 24 nt

427   markers). Gel slices were crushed and soaked in gel diffusion buffer (0.5 M ammonium acetate;

428   10 mM magnesium acetate; 1 mM EDTA, pH 8.0; 0.1% SDS) at 37 °C for 30 min with high

429   speed shaking, ethanol precipitated and resuspended in 20 μL of RNase free water. Ligated

430   RNAs were reverse transcribed with Superscript III reverse transcriptase (Invitrogen) and

431   amplified with Q5 polymerase (New England Biolabs). PCR was monitored using a real time

432   PCR thermal cycler and amplification was discontinued when it ceased to amplify linearly. PCR

433   products were run on a 10% TBE polyacrylamide gel, size selected for an amplicon with the

434   predicted 20-50 bp insert size to exclude linker dimers, and purified from the gel (Qiagen).

435   Cleaned up library DNA was quantified on an Agilent 2100 Bioanalyzer using the High

436   Sensitivity DNA Kit before being sequenced. All GCLiPP and RNAseq sequencing runs were

437   carried out on an Illumina HiSeq 2500 sequencer.

438

439   GCLiPP and RNAseq bioinformatics analysis pipeline

440   FastQ files were de-multiplexed and trimmed of adapters. Each experiment was

441   performed on three technical replicates per condition (resting and stimulated) per experiment.

442   Cloning replicates and experiments were pooled in subsequent analyses. Jurkat and mouse T cell

443   trimmed sequence reads were aligned to the hg38 human or mm10 mouse genome assembly

444   using bowtie2, respectively. After alignment, PCR amplification artifacts were removed by de-

445   duplication using the 2-nt random sequence at the 5′ end of the 3′ linker using a custom script

446   that counted only a single read containing a unique linker sequence and start and end position of

447   alignment per sequenced sample. Peaks of GCLiPP read density were called by convolving a

19

448   normal distribution against a sliding window of the observed read distribution. A 70 nucleotide

449   window was analyzed centered on every nucleotide within the 3' UTR. For each window, the

450   observed distribution of read density was compared to a normal distribution of the same

451   magnitude as the nucleotide in the center of the window. The Pearson's correlation coefficient

452   was computed for each nucleotide and peaks were defined as local maxima of goodness of fit

453   between observed GCLiPP read density and the normal distribution, requiring a read depth above

454   20% of the maximum read depth in the 3' UTR global minimum of 10 reads. RNAseq reads

455   were aligned using STAR Aligner (https://github.com/alexdobin/STAR) (Dobin et al., 2013) to

456   align against the mm10 genome, and gene expression data were calculated as fragments per

457   kilobase per million reads. All custom scripts are available as STAR Methods Key Resource.

458

459   Comparison of GCLiPP to individual eCLIP datasets

460   eCLIP data (Sundararaman et al., 2016) were downloaded via the ENCODE data portal

461   (http://www.encodeproject.org/). The first replicate set of bigwig files were downloaded for each

462   RBP deposited online at the time of analysis (December 2017) as well as CLIPper called peaks

463   for the same. To facilitate comparisons with GCLiPP we called GCLiPP peaks in the Jurkat data

464   using CLIPper (Lovci et al., 2013) after re-aligning Jurkat GCLiPP reads to hg19. Correlation

465   analysis was performed with a custom perl script that calculated the Spearman correlation for

466   read depth at each nucleotide in the 3' UTR of all genes that were expressed in each dataset (as

467   determined by CLIP read depth). ~5000-15000 expressed genes were included in the correlation

468   analysis for each RBP. For comparison to mRNP abundancy, log10 RBP mass spectrometry

469   spectra counts were utilized from (Baltz et al., 2012). To stratify RBPs by subcellular

470   localization, data were taken from the COMPARTMENTS database, with RBPs with a

471   localization score of 5 in the cytosol counted as cytosolic and lower counted as non-cytosolic

472   (Binder et al., 2014). All custom scripts are available as STAR Methods Key Resource.

473

474   <u>CRISPR editing</u>

475   Guide RNA sequences were selected using the Benchling online CRISPR design tool

476   (https://benchling.com/crispr) with guides selected to target genomic regions of GCLiPP read

477   density. Synthetic crRNAs and tracrRNA (Dharmacon) were resuspended in water at 160 µM at

478   1:1 ratio and allowed to hybridize at 37 c for 30 m. For CRISPR dissection experiments, all

479   crRNAs were mixed at an equimolar ratio before annealing to tracrRNA. This annealed gRNA

480   complex (80 µM) was then mixed 1:1 by volume with 40 µM *S. pyogenes* Cas9-NLS (University

481   of California Berkeley QB3 Macrolab) to a final concentration of 20 µM Cas9 ribonucleotide

482   complex (RNP). This complexed gRNA:Cas9 RNP was mixed with a carrier solution of salmon

483   sperm DNA (Invitrogen) and diluted to a final concentration between 5-20 µM. The diluted

484   gRNA:Cas9 RNPs were nucleofected into primary mouse T cells (24 hours after stimulation)

485   with the P3 Primary Cell 96-well Nucleofector ™ Kit and into Jurkat cells with the SE Cell Line

486   96-well Nucleofector ™ Kit using a 4-D Nucleofector following the manufacturer's

487   recommendations (Lonza). Cells were pipetted into pre-warmed media and then returned to

488   CD3/CD28 stimulation for another two days and expanded an additional 3 days (mouse primary

489   T cells) or cultured for 7-10 days (Jurkat).

490

491

492   <u>Quantification of CD69 protein and mRNA</u>

493   Jurkat cells were gene edited using the CRISPR-Cas9 system as described above. After 3

494   days in culture, cells were washed and stained with anti-human CD69 PE (BioLegend, clone

495   FN50). Samples were acquired on a FACSAria II (BD Biosciences) with CD69^{hi} cells single cell

496   sorted into a 96 well plate and incubated at 37°C at 5% CO_2 to allow generation of single cell

497   clones. To verify editing, gDNA was extracted from Jurkat clones using QuickExtract ™ DNA

498   Solution (Epicentre) according to the manufacturer's instructions. Samples were then amplified

499   by PCR using Q5® High Fidelity DNA Polymerase (New England Biolabs). Quantitative PCR

500   reactions were performed on an Eppendorf Realplex 2S thermocycler with the following

501   program: (95°C 60 s; 35 cycles of 95°C 30 s, 58°C 30 s, 72°C 60 s).

502

503   3' UTR dissection

504   3' UTR dissection was performed as described (Zhao et al., 2017). Gene edited cells were

505   harvested into Trizol reagent (Invitrogen) and total RNA was phase separated and purified from

506   the aqueous phase using the Direct-zol RNA miniprep kit with on-column DNase treatment

507   (Zymo). Genomic DNA was extracted from the remaining organic phase by vigorous mixing

508   with back extraction buffer (4 M guanidine thiocyanate, 50 mM sodium citrate, 1 M Tris base).

509   cDNA was prepared with oligo-dT using the SuperScript III reverse transcription kit

510   (Invitrogen). cDNA and genomic DNA were used as a template for PCR using MyTaq 2X Red

511   Mix (Bioline). To equilibrate the number of target molecules and number of PCR cycles between

512   samples, we performed semi-quantitative PCR followed by agarose gel electrophoresis to

513   determine a PCR cycle number where genomic DNA first showed visible bands. This cycle

514   number was then used with a titration of cDNA concentrations and a concentration that amplified

515   equivalently was selected for analysis by deep sequencing. To quantify relative RNA/DNA

516   ratios, cDNA and genomic DNA amplicons were purified using a QIAquick PCR purification up

517   kit (Qiagen) and quantified on an Agilent 2100 Bioanalyzer using the High Sensitivity DNA Kit

518   (Agilent).

519   Amplicons were tagmented with the Nextera XT kit (Illumina) and sequenced on an

520   Illumina 2500 HiSeq. Reads were aligned to a custom genome consisting of the targeted PCR

521   amplicon using STAR aligner and mutations were scored using an awk script

522   (https://github.com/alexdobin/STAR/blob/master/extras/scripts/sjFromSAMcollapseUandM.awk

523   ). RNA/DNA read ratios were calculated for all mutations over 20 nucleotides long and less than

524   250 nucleotides long, and relative expression was quantified as the median normalized

525   RNA/DNA ratio for this subset of mutations. Mutations had to have at least 10 reads in both the

526   RNA and gDNA amplicons and mutations with an RNA/DNA ratio of greater than 10 were

527   excluded as outliers. Effect sizes for each nucleotide of the amplicon in each experiment were

528   computed by comparing this median normalized RNA/DNA ratio for all mutations spanning a

529   given nucleotide to all other mutations. Combined p-values were calculated using a Welch's two

530   sample t-test comparing all mutations spanning a given nucleotide with all other mutations.

531

532   <u>Shared peak calling, motif analysis and icSHAPE and Phylogenetic analyses</u>

533   3′ UTR alignments of mouse and human were performed by downloading hg38 RefSeq

534   3' UTRs from UCSC genome browser, (http://genome.ucsc.edu), identifying syntenic regions of

535   the mouse genome in mm10 with the KentUtils liftOver program

536   (https://github.com/ucscGenomeBrowser/kent) and aligning UTRs with Clustal Omega

537   (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al., 2011). Biochemically shared peaks

538   were called by the following algorithm: Measure normalized GCLiPP read density (i.e. the

539　fraction of the maximal read depth within that 3' UTR) at each position. Calculate correlation

540　between mouse and human normalized signal, as well as observed data and a normal distribution

541　centered at the point being examined in both the mouse and human data tracks. These three

542　Spearman correlations were added together to calculate a numerical score, and shared peaks were

543　defined as local maxima of these scores. To identify high stringency peaks, peaks were only

544　accepted if they 1) had a correlation of >0.75 between mouse and human, 2) had a peak that had

545　a read density of >0.5 of the maximum read density within that 3' UTR in one data track (mouse

546　or human) and >0.2 in the other and 3) had >10 reads at that location in both mouse and human

547　datasets. Biological enrichment of genes with shared peaks was calculated using the Metascape

548　(Tripathi et al., 2015) online interface (http://metascape.org) using the default settings, with the

549　exception that a background set of genes was included in the analysis, specifically all genes that

550　contain a called GCLiPP peak in both human and mouse datasets that do not contain a

551　biochemically shared peak.

552　　　For motif calling, HOMER (Heinz et al., 2010) was used in RNA mode with the

553　"noweight" option to turn off GC correction to search for motifs of width 5, 6 or 7 nucleotides,

554　with otherwise default parameters. The positive sequence set was the mouse and human

555　sequences of the biochemically shared GCLiPP peaks, the negative sequence set was all other

556　GCLiPP called peaks from Jurkat and mouse T cells that were not shared across species. For

557　icSHAPE we used a published bigwig file of locally normalized icSHAPE signal intensity

558　generated in mouse ES cell (Spitale et al., 2015). Conservation of loci in the mouse and human

559　genomes were obtained from the UCSC genome browser as a bigwig of PhyloP scores of

560　conservation across 60 placental mammals (mouse) and 100 vertebrates (human)

24

561   (http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phyloP60way/,

562   http://hgdownload.cse.ucsc.edu/goldenpath/hg38/phyloP100way/).

563

564   <u>Oligonucleotide and primer sequences</u>

565   GCLiPP 3′ RNA linker: 5′-NNGUGUCUUUACACAGCUACGGCGUCG-3′

566   GCLiPP 5′ RNA linker: 5′-CGACCAGCAUCGACUCAGAAG-3′

567   GCLiPP Reverse transcription primer: 5′-

568   CAAGCAGAAGACGGCATACGAGATNNNNNNCGCTAGTGACTGGAGTTCAGACGTGT

569   GCTCTTCCGATCCGACGCCGTAGCTGTGTAAA-3′ (NNNNNN is barcode for

570   demultiplexing)

571   GCLiPP 3′ PCR primer: 5′-CAAGCAGAAGACGGCATACGAGAT-3′

572   GCLiPP 5′ PCR primer: 5′-

573   AATGATACGGCGACCACCGAGATCTACACTGGTACTCCGACCAGCATCGACTCAGA

574   AG-3′

575   Read1seq sequencing primer for GCLiPP: 5′-

576   ACACTGGTACTCCGACCAGCATCGACTCAGAAG-3′Index sequencer primer for GCLiPP:

577   5′-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3′

578   CD69 gRNA1: CTCAAGGAAATCTGTGTCAG

579   CD69 gRNA2: TCATTCTTGGGCATGGTTAT

580   CD69 gRNA3: CCTGTGATGCTTCTAGCTCA

581   CD69 gRNA4: AATAATGAAATAACTAGGCG

582   CD69 gRNA5: TAATTGAATCCCTTAAACTC

583   CD69 gRNA6: TGATGTGGCAAATCTCTATT

584   PIM3 (human) gRNA1: TGTGCAGGCATCGCAGATGG

585   PIM3 (human) gRNA2: GACTTTGTACAGTCTGCTTG

586   PIM3 (human) gRNA3: GTGGCTAACTTAAGGGGAGT

587   PIM3 (human) gRNA4: AAACAATAAATAGCCCCGGT

588   PIM3 (human) gRNA5: TTGAGAAAACCAAGTCCCGC

589   PIM3 (human) gRNA6: CAGGAGGAGACGGCCCACGC

590   PIM3 (human) gRNA7: TTTATGGTGTGACCCCCTGG

591   PIM3 (human) gRNA8: CCAAGCCCCAGGGGACAGTG

592   Pim3 (mouse) gRNA1: GTTCAATTCTGGGAGAGCGC

593   Pim3 (mouse) gRNA2 CTGGTTCAAGTATCCACCCA

594   Pim3 (mouse) gRNA3: CCATAAATAAGAGACCGTGG

595   Pim3 (mouse) gRNA4: GCTTCCTCCCGCAAACACGG

596   Pim3 (mouse) gRNA5: CTGGTGTGACTAAGCATCAG

597   Pim3 (mouse) gRNA6: TGGAGAAGGTGGTTGCTTGG

598   Primers

599   CD69 F: TGGAATGTGAGAAGAATTTATACTGG

600   CD69 R: GTAATAGAATTGATTTAGGAAAG

601   PIM3 F (human): TCCAGCAGCGAGAGCTTGTGAGGAG

602   PIM3 R(human): TGATCTCCAGACATCTCACTTTTGAACTG

603   PIM3 R2(human):

604   TGAGATAGGTGCCTCACTGATTAAGCATTGGTGATCTCCAGACATCTCACTTTTGAA

605   CTG

606   Pim3 F (mouse): GCGTTCCAGAGAACTGTGACCTTCG

26

607   Pim3 R (mouse): TATGATCTTCAGACATTTCACACTTTG

608

609   **References**

610   Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M.,
611   Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound
612   Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell *46*, 674–
613   690.

614   Bassell, G.J., and Kelic, S. (2004). Binding proteins for mRNA localization and local translation,
615   and their dysfunction in genetic neurological disease. Curr. Opin. Neurobiol. *14*, 574–581.

616   Beelman, C.A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. Cell *81*, 179–183.

617   Berry, M.J., Banu, L., Harney, J.W., and Larsen, P.R. (1993). Functional characterization of the
618   eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. EMBO J. *12*,
619   3315–3322.

620   Binder, J.X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S.I., Schneider, R., and
621   Jensen, L.J. (2014). COMPARTMENTS: unification and visualization of protein subcellular
622   localization evidence. Database *2014*.

623   Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013).
624   Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin,
625   DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

626   Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E.,
627   Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an
628   Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406.

629   Chao, J.A., Patskovsky, Y., Patel, V., Levy, M., Almo, S.C., and Singer, R.H. (2010). ZBP1
630   recognition of β-actin zipcode induces RNA looping. Genes Dev. *24*, 148–158.

631   Cibrián, D., and Sánchez-Madrid, F. (2017). CD69: from activation marker to metabolic
632   gatekeeper. Eur. J. Immunol. *47*, 946–953.

633   Cibrian, D., Saiz, M.L., de la Fuente, H., Sánchez-Díaz, R., Moreno-Gonzalo, O., Jorge, I.,
634   Ferrarini, A., Vázquez, J., Punzón, C., Fresno, M., et al. (2016). CD69 controls the uptake of L-
635   tryptophan through LAT1-CD98 and AhR-dependent secretion of IL-22 in psoriasis. Nat.
636   Immunol. *17*, 985–996.

637   Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder,
638   M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell
639   chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*,
640   1193–1203.

641   Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
642   M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf.
643   Engl. *29*, 15–21.

644   Fong, A.M., Premont, R.T., Richardson, R.M., Yu, Y.-R.A., Lefkowitz, R.J., and Patel, D.D.
645   (2002). Defective lymphocyte chemotaxis in beta-arrestin2- and GRK6-deficient mice. Proc.
646   Natl. Acad. Sci. U. S. A. *99*, 7478–7483.

647   Freeberg, M.A., Han, T., Moresco, J.J., Kong, A., Yang, Y.-C., Lu, Z.J., Yates, J.R., and Kim,
648   J.K. (2013). Pervasive and dynamic protein binding sites of the mRNA transcriptome in
649   Saccharomyces cerevisiae. Genome Biol. *14*, R13.

650   González-Amaro, R., Cortés, J.R., Sánchez-Madrid, F., and Martín, P. (2013). Is CD69 an
651   effective brake to control inflammatory diseases? Trends Mol. Med. *19*, 625–632.

652   Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A.,
653   Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification
654   of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

655   Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C.,
656   Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription
657   factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell
658   *38*, 576–589.

659   Johansson, L., Gafvelin, G., and Arnér, E.S.J. (2005). Selenocysteine in proteins-properties and
660   biotechnological use. Biochim. Biophys. Acta *1726*, 1–13.

661   Kafasla, P., Skliris, A., and Kontoyiannis, D.L. (2014). Post-transcriptional coordination of
662   immunological responses by RNA-binding proteins. Nat. Immunol. *15*, 492–502.

663   Kanitz, A., and Gerber, A.P. (2010). Circuitry of mRNA regulation. Wiley Interdiscip. Rev. Syst.
664   Biol. Med. *2*, 245–251.

665   Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet.
666   *8*, 533–543.

667   Kislauskis, E.H., Zhu, X., and Singer, R.H. (1994). Sequences responsible for intracellular
668   localization of beta-actin messenger RNA also affect cell phenotype. J. Cell Biol. *127*, 441–451.

669   Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA
670   Interactions. Mol. Cell *69*, 354–369.

671   Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H., and Fu, X.-D. (2017). GRID-seq reveals the global
672   RNA–chromatin interactome. Nat. Biotechnol. *35*, 940–950.

673   Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J.,
674   Gehman, L.T., Hoon, S., et al. (2013). Rbfox proteins regulate alternative mRNA splicing
675   through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol. *20*, 1434–1442.

28

676   Makeyev, A.V., and Liebhaber, S.A. (2002). The poly(C)-binding proteins: A multiplicity of
677   functions and a search for mechanisms. RNA *8*, 265–278.

678   Mariotti, M., Lobanov, A.V., Guigo, R., and Gladyshev, V.N. (2013). SECISearch3 and
679   Seblastian: new tools for prediction of SECIS elements and selenoproteins. Nucleic Acids Res.
680   *41*, e149.

681   Martin, K.C., and Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial
682   Dimension. Cell *136*, 719–730.

683   Martín, P., Gómez, M., Lamana, A., Cruz-Adalia, A., Ramírez-Huesca, M., Ursa, M.A., Yáñez-
684   Mo, M., and Sánchez-Madrid, F. (2010). CD69 association with Jak3/Stat5 proteins regulates
685   Th17 cell differentiation. Mol. Cell. Biol. *30*, 4877–4889.

686   Millevoi, S., and Vagner, S. (2009). Molecular mechanisms of eukaryotic pre-mRNA 3′ end
687   processing regulation. Nucleic Acids Res. gkp1176.

688   Narlik-Grassow, M., Blanco-Aparicio, C., and Carnero, A. (2014). The PIM family of
689   serine/threonine kinases in cancer. Med. Res. Rev. *34*, 136–159.

690   Nawijn, M.C., Alendar, A., and Berns, A. (2011). For better or for worse: the role of Pim
691   oncogenes in tumorigenesis. Nat. Rev. Cancer *11*, 23.

692   Papp, L.V., Lu, J., Holmgren, A., and Khanna, K.K. (2007). From Selenium to Selenoproteins:
693   Synthesis, Identity, and Their Role in Human Health. Antioxid. Redox Signal. *9*, 775–806.

694   Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev. *25*,
695   1770–1782.

696   Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl,
697   M., and Bohjanen, P.R. (2002). Genome-wide analysis of mRNA decay in resting and activated
698   primary human T lymphocytes. Nucleic Acids Res. *30*, 5529–5538.

699   Reed, R. (2003). Coupling transcription, splicing and mRNA export. Curr. Opin. Cell Biol. *15*,
700   326–331.

701   Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells
702   express mRNAs with shortened 3’ untranslated regions and fewer microRNA target sites.
703   Science *320*, 1643–1647.

704   Santis, A.G., López-Cabrera, M., Sánchez-Madrid, F., and Proudfoot, N. (1995). Expression of
705   the early lymphocyte activation antigen CD69, a C-type lectin, is regulated by mRNA
706   degradation associated with AU-rich sequence motifs. Eur. J. Immunol. *25*, 2142–2146.

707   Schwerk, J., and Savan, R. (2015). Translating the Untranslated Region. J. Immunol. *195*, 2963–
708   2971.

709   Seemann, S.E., Mirza, A.H., Hansen, C., Bang-Berthelsen, C.H., Garde, C., Christensen-
710   Dalsgaard, M., Torarinsson, E., Yao, Z., Workman, C.T., Pociot, F., et al. (2017). The
711   identification and functional annotation of RNA structures conserved in vertebrates. Genome
712   Res. *27*, 1371–1383.

713   Shiow, L.R., Rosen, D.B., Brdicková, N., Xu, Y., An, J., Lanier, L.L., Cyster, J.G., and
714   Matloubian, M. (2006). CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and
715   lymphocyte egress from lymphoid organs. Nature *440*, 540–544.

716   Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
717   Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple
718   sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

719   Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y.,
720   Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA
721   regulatory mechanisms. Nature *519*, 486–490.

722   Steiner, D.F., Thomas, M.F., Hu, J.K., Yang, Z., Babiarz, J.E., Allen, C.D.C., Matloubian, M.,
723   Blelloch, R., and Ansel, K.M. (2011). MicroRNA-29 regulates T-box transcription factors and
724   interferon-γ production in helper T cells. Immunity *35*, 169–181.

725   Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X.,
726   Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., et al. (2016). Resources for the Comprehensive
727   Discovery of Functional RNA Elements. Mol. Cell *61*, 903–913.

728   Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield,
729   N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape
730   of the human genome. Nature *489*, 75–82.

731   Timchenko, L.T., Miller, J.W., Timchenko, N.A., DeVore, D.R., Datar, K.V., Lin, L., Roberts,
732   R., Caskey, C.T., and Swanson, M.S. (1996). Identification of a (CUG)n triplet repeat RNA-
733   binding protein and its expression in myotonic dystrophy. Nucleic Acids Res. *24*, 4407–4414.

734   Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton,
735   H.M., DeJesus, P., Che, J., Mulder, L.C.F., et al. (2015). Meta- and Orthogonal Integration of
736   Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. Cell Host Microbe *18*,
737   723–735.

738   Tujebajeva, R.M., Copeland, P.R., Xu, X.-M., Carlson, B.A., Harney, J.W., Driscoll, D.M.,
739   Hatfield, D.L., and Berry, M.J. (2000). Decoding apparatus for eukaryotic selenocysteine
740   insertion. EMBO Rep. *1*, 158–163.

741   Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J.,
742   Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer Evolution across 20
743   Mammalian Species. Cell *160*, 554–566.

744   Weinreb, C., Riesselman, A.J., Ingraham, J.B., Gross, T., Sander, C., and Marks, D.S. (2016).
745   3D RNA and Functional Interactions from Evolutionary Couplings. Cell *165*, 963–975.

746 Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., Conboy, C.M., Vanes, L., Tybulewicz,
747 V.L.J., Fisher, E.M.C., Tavaré, S., and Odom, D.T. (2008). Species-specific transcription in mice
748 carrying human chromosome 21. Science *322*, 434–438.

749 Zhao, W., Siegel, D., Biton, A., Tonqueze, O.L., Zaitlen, N., Ahituv, N., and Erle, D.J. (2017).
750 CRISPR–Cas9-mediated functional dissection of 3′-UTRs. Nucleic Acids Res. *45*, 10800–10810.

751 Zubiaga, A.M., Belasco, J.G., and Greenberg, M.E. (1995). The nonamer UUAUUUAUU is the
752 key AU-rich sequence motif that mediates mRNA degradation. Mol. Cell. Biol. *15*, 2219–2230.

753

754
755
756 **Author Contributions**:

757
758 Author Contributions: A.L. and W.S.Z. performed the experiments and performed bioinformatic

759 analyses. R.K. established the bioinformatic pipeline for small RNA sequencing analysis. W.Z.

760 and D.E. helped design CRISPR dissection experiments. N.Z. consulted on data analysis and

761 interpretation. K.M.A., A.L. and W.S.Z. designed experiments, interpreted the data, and wrote

762 the manuscript. All authors discussed the results and approved the manuscript.

763

764 **Data availability:**

765 Datasets in this paper are available on Gene Expression Omnibus accessions GSE94554 and

766 GSE115886

767
768
769 **Acknowledgements**:
770

31

775   Asthma Basic Research Center, and a Scholar Award (K.M.A.) from The Leukemia &

776   Lymphoma Society. The authors declare no conflicts of interest.

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

A *In vitro* cultured T cells → Crosslinking, biotinylation oligo-dT selection → RNP capture & RNase T1 → Linker ligation & proteinase digestion → Library preparation and sequencing

B
ρ=0.743
Jurkat replicate 2 Normalized peak coverage
Jurkat replicate 1 Normalized peak coverage

C
ρ=0.904
Mouse primary Th2 cells Normalized peak coverage
Mouse primary CD8 T cells Normalized peak coverage

D
Jurkat human T cells   Primary mouse T cells

- 5' UTR
- Coding sequence
- 3' UTR
- Introns
- Misc. Non–coding RNA e.g. lncRNA
- snoRNA, tRNA and miRNA
- Ribosomal RNA
- Other/intergenic
- Transposons LINE,SINE,ERV,etc.

E
Jurkat human T cells   Primary mouse T cells

Relative coverage

5' UTR   CDS   introns   3' UTR   lncRNA   Transposons LINE/SINE etc.

798 **Figure 1. GCLiPP sequencing reveals RNA transcript protein occupancy. (A)** GCLiPP

799 method of global RBP profiling. T cell RNAs are crosslinked to RBPs and lysates are

800 biotinylated on primary amines. mRNAs are enriched with oligo-dT beads, and RBP protected

801 sites are digested, captured, sequenced and aligned to the genome. **(B)** Normalized GCLiPP read

802 depth (fraction of reads in called peak relative to all GCLiPP reads in annotated 3′ UTR) in two

803 replicates of Jurkat cells. ρ represents Pearson correlation. **(C)** Normalized GCLiPP read depth

804 in mouse primary Th2 and CD8 T cells. ρ represents Pearson correlation. **(D)** Proportion of

805 mapped GCLiPP reads derived from genomic features. **(E)** Relative coverage of genomic

806 features in GCliPP sequencing reads relative to total length of genomic features of indicated

807 class.

808

809

810

811

812

813

814

815

816

817

818

819

A Pairwise correlations of thousands of individual 3' UTRs

GCLiPP
TIA1 eCLIP
TIA1 eCLIP input

r=0.70
r= -0.36

CLIC1

GCLiPP
IGF2BP1 eCLIP
IGF2BP1 eCLIP input

r=0.93
r=0.43

OLA1

GCLiPP
PUM2 eCLIP
PUM2 eCLIP input
PUM2 UGUA motif

r=-0.23
r=0.09

OSTC

B Paired t-test of eCLIP correlations vs. eCLIP input correlations

t = 15.4
Input correlation
TIA1 correlation

t = 9.8
Input correlation
IGF2BP1 correlation

t = -5.8
Input correlation
PUM2 correlation

C Overlap of called peaks

Shuffled peak overlap | Observed TIA1 overlap
69σ
Fraction of TIA1 peaks overlapping GCLIPP peak

Shuffled | Observed
71σ
Fraction of IGF2BP1 peaks overlapping GCLIPP peak

Shuffled | Observed
62σ
Fraction of PUM2 peaks overlapping GCLIPP peak

D

Non-cytoplasmic RBPs
r=0.14
p=0.41
log10 RBP abundance in mRNP
GCLiPP-eCLIP / GCLiPP-eCLIP input t statistic

Cytoplasmic RBPs
r=0.41
p=0.03
log10 RBP abundance in mRNP
GCLiPP-eCLIP / GCLiPP-eCLIP input t statistic

820 **Figure 2. Comparisons with eCLIP reveal abundant cytosolic RBPs drive GCLiPP signal**

821 (**A**) Genomic snapshots of individual 3' UTRs showing exemplary correlation between eCLIP
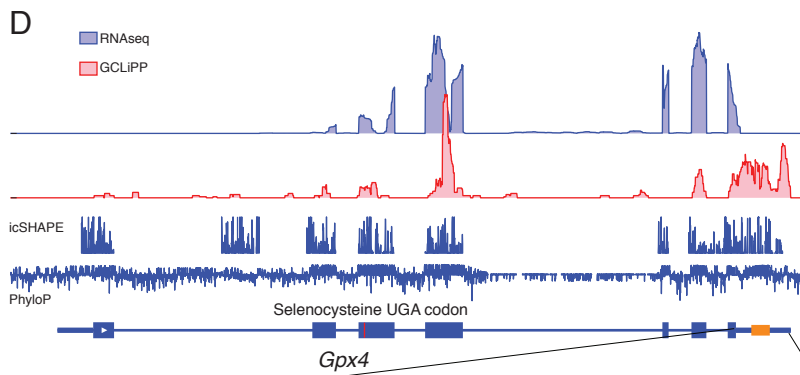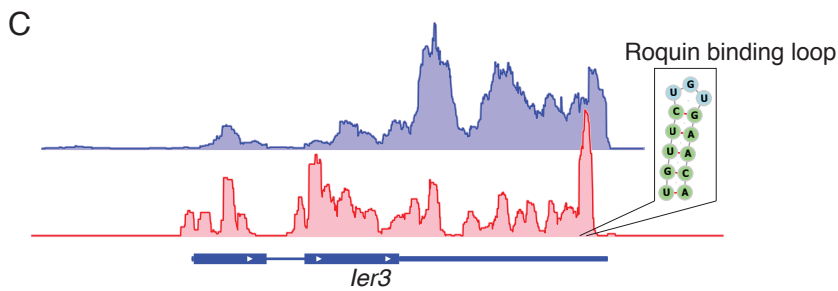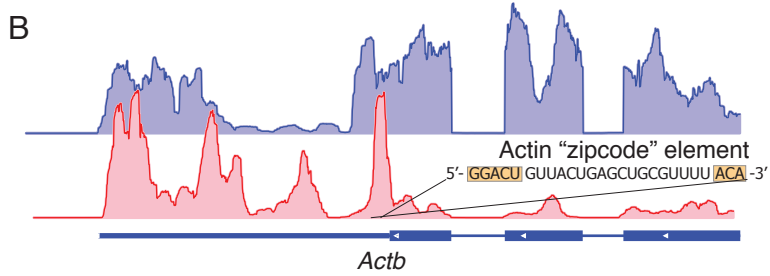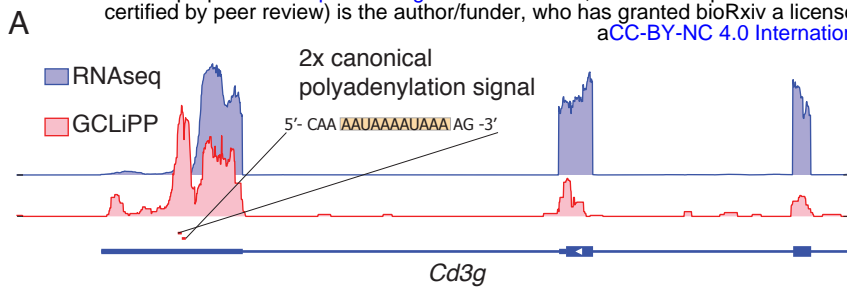
822 datasets and GCLiPP. GCLiPP is shown in red, while the indicated RBP eCLIP data is shown in

823 blue, and matched control input samples are shown in gray, shown for the 3' UTRs of the

824 indicated gene. r indicates Pearson correlation between pairs of normalized read density at a

825 given nucleotide for the indicated comparisons. (**B**) 2D density plots showing matched

826 correlations between GCLiPP and eCLIP for the indicated RBP (X-axis) and GCLiPP and the

827 matched control input sample (Y-axis) for individual 3' UTR for all expressed genes in eCLIP

828 and GCLiPP datasets. The t-statistic shown is for a paired t-test of the correlations. (**C**) Overlap

829 of CLIPper called peaks in 3' UTRs in GCLiPP and eCLIP. Red lines indicate observed overlap

830 of GCLiPP peaks and eCLIP peaks. Green distribution represents bootstrapped expected overlap,

831 computed by shuffling called eCLIP peaks within the same 3' UTR, computing overlap of

832 shuffled set with GCLiPP called peaks, and repeating this analysis 500 times. The indicated

833 distance represents the number of standard deviations above the mean shuffled overlap of the

834 observed overlap. (**D**) Correlation of eCLIP-GCLiPP paired t-tests (from (**B**) and RBP

835 abundance in mRNPs). RBPs shown in gray score are not cytosolic localized (<5 cytosolic

836 according to COMPARTMENTS) whereas RBPs in red are cytosolic localized (5 cytosolic

837 according to COMPARTMENTS).

838

839

840

841

34

A

RNAseq
GCLiPP

2x canonical
polyadenylation signal

5'- CAA AAUAAAAUAAA AG -3'

Cd3g

B

Actin "zipcode" element
5'- GGACU GUUACUGAGCUGCGUUUU ACA -3'

Actb

C

Roquin binding loop

Ier3

D

RNAseq
GCLiPP

icSHAPE

PhyloP

Selenocysteine UGA codon

Gpx4

E

icSHAPE

PhyloP

AAUAAA

Selenocysteine insertion
(SECIS) element predicted by
SECISearch 3

F

Exposed
adenines
in apical loop

SECIS
inferred structure

G

icSHAPE
0   1

Vienna rnafold
predicted structure
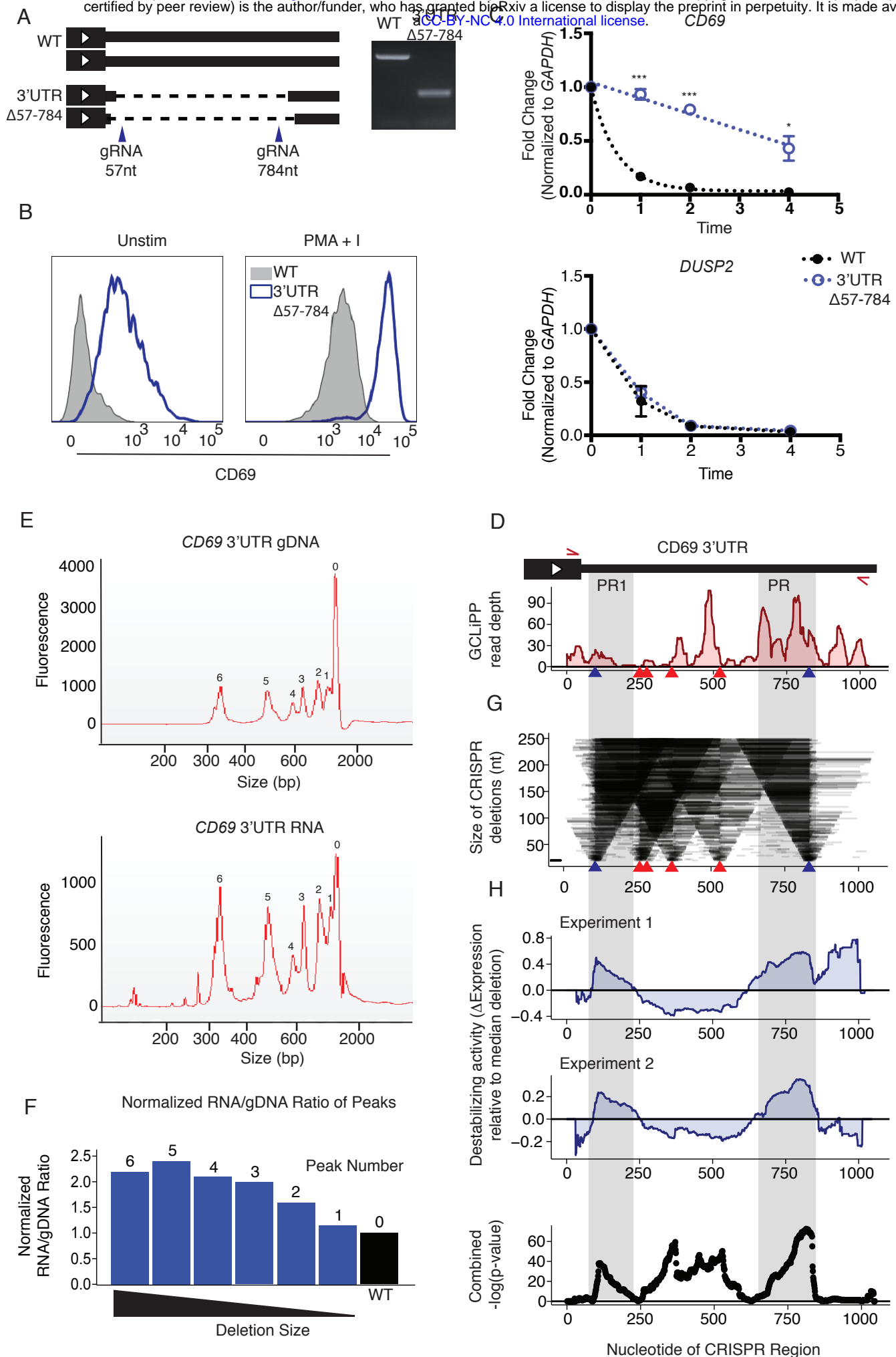
842 **Figure 3. GCLiPP recapitulates previously described mRNA-RBP interactions in primary**

843 **T cells.** RNAseq and GCLiPP tracks for **(A)** *Cd3g* (**B**) *Actb* **(C)** *Ier3* and (**D-G**) *Gpx4*. RNAseq

844 track is from resting Th2 cells. GCLiPP is sum of five experiments, three in Th2 and two in CD8

845 T cells. Location of known RBP binding determinants are shown as insets.

867 **Figure 4. GCLiPP guides identification of destabilizing regions in CD69 3'UTR.** (A)

868 Schematic illustration and gel image of Jurkat WT and CD69 3'UTR edited clone (3'UTR Δ57-

869 784). Editing was performed using CRISPR-Cas9 and verified through PCR. Blue arrows

870 indicate gRNA placement with their positions in the 3'UTR indicated by nucleotide (nt) number.

871 (B) CD69 protein expression of WT and 3'UTR Δ57-784 clone as measured by flow cytometry.

872 Cells were either untreated or stimulated with PMA and Ionomycin (PMA + I) for 4 hours. (C)

873 mRNA decay of *CD69* (top) and *DUSP2 (bottom)* transcript in WT and 3'UTR Δ57-784 clones.

874 Cells were stimulated for 4 h with PMA+I and then treated with Actinomycin-D (act-D).

875 Transcript expression was measured 0, 1, 2 or 4 hours post addition of act-D by qPCR and

876 normalized to *GAPDH* expression. Data was generated from two separate experiments each with

877 N=2 and significance was calculated using multiple t-test corrected with Holm-Sidak method

878 *p<0.05, **p<0.01, ***p<0.001 (D-H) CRISPR-Cas9 dissection of *CD69* 3'UTR. (D) CD69

879 3'UTR GCLiPP peaks aligned to schematic illustration of 3'UTR. Red arrows indicate primer

880 position during PCR amplification of gDNA and RNA from pooled crRNP transfected Jurkats.

881 Arrow heads represent gRNA placement. Blue arrow heads were gRNAs also used in experiment

882 in Figure 4A. (E) Microfluidic capillary electrophoresis of CD69 3'UTR gDNA (top) and reverse

883 transcribed cDNA (bottom). Individual fragments identified by the Agilent Bioanalyzer software

884 are indicated by the numbers on the graph. (F) Ratio of RNA/gDNA for each labeled peak in (D)

885 normalized to predicted wild-type allele (black bar). Estimated molarity of each peak from the

886 electrophoresis was used to calculate RNA/gDNA ratio. The labeled peaks are plotted in

887 descending order based on deletion size. (G) Size of deletions generated using CRISPR-Cas9.

888 Arrow heads represent gRNA placement as mentioned for Figure 4D. (H) Change in expression

889 along the 3'UTR relative to median expression of all possible deletions. Per-nucleotide effect

37

890 score was calculated by comparing median normalized RNA/gDNA ratio for all shown deletions

891 spanning a given nucleotide with all shown deletions. Experiment 1 and 2 are duplicate samples

892 which were transfected with 80μM or 120μM of gRNAs respectively. Grey shaded area PR1 and

893 PR2 indicate regions of significant destabilizing activity. Unadjusted -log10 p-values from

894 Welch's two sample t-test comparing all deletions spanning a nucleotide with all other deletions

895 across both experiments (bottom).

896

897

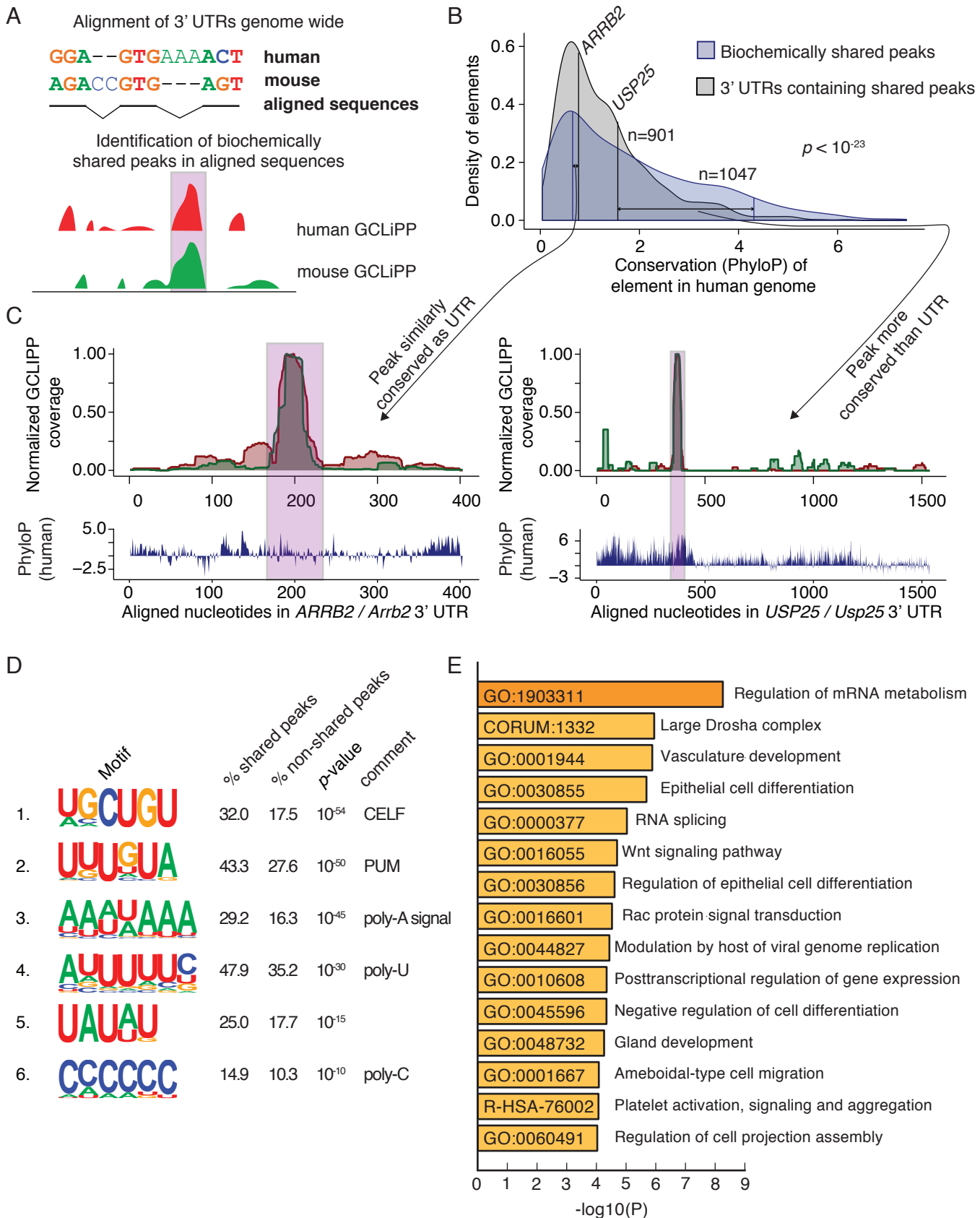898

899

900

901

902

903

904

905

906

907

908

909

910

911

A Alignment of 3' UTRs genome wide

```
GGA--GTGAAAACT   human
AGACCGTG---AGT   mouse
                 aligned sequences
```

Identification of biochemically
shared peaks in aligned sequences

human GCLiPP

mouse GCLiPP

B

Density of elements vs Conservation (PhyloP) of element in human genome

- Biochemically shared peaks
- 3' UTRs containing shared peaks

ARRB2
USP25
n=901
n=1047
$p < 10^{-23}$

C

Peak similarly conserved as UTR

Normalized GCLiPP coverage / PhyloP (human) — Aligned nucleotides in *ARRB2 / Arrb2* 3' UTR

Peak more conserved than UTR

Normalized GCLiPP coverage / PhyloP (human) — Aligned nucleotides in *USP25 / Usp25* 3' UTR

D

	Motif	% shared peaks	% non-shared peaks	*p*-value	comment
1.	UGCUGU	32.0	17.5	10^{-54}	CELF
2.	UUUGUA	43.3	27.6	10^{-50}	PUM
3.	AAAUAAA	29.2	16.3	10^{-45}	poly-A signal
4.	AUUUUUC	47.9	35.2	10^{-30}	poly-U
5.	UAUAU	25.0	17.7	10^{-15}	
6.	CCCCCC	14.9	10.3	10^{-10}	poly-C

E

GO term	Description
GO:1903311	Regulation of mRNA metabolism
CORUM:1332	Large Drosha complex
GO:0001944	Vasculature development
GO:0030855	Epithelial cell differentiation
GO:0000377	RNA splicing
GO:0016055	Wnt signaling pathway
GO:0030856	Regulation of epithelial cell differentiation
GO:0016601	Rac protein signal transduction
GO:0044827	Modulation by host of viral genome replication
GO:0010608	Posttranscriptional regulation of gene expression
GO:0045596	Negative regulation of cell differentiation
GO:0048732	Gland development
GO:0001667	Ameboidal-type cell migration
R-HSA-76002	Platelet activation, signaling and aggregation
GO:0060491	Regulation of cell projection assembly

-log10(P)

912 **Figure 5. Comparison between mouse and human GCLiPP reveals principles of shared**

913 **post-transcriptional regulation.** (**A**) Schematic illustration of 3' UTR alignment and

914 biochemically shared GCLiPP peak calling. (**B**) Distribution of conservation across 100

915 vertebrates (PhyloP score) of regions in the human genome. Blue indicates biochemically shared

916 peaks, gray indicates the 3' UTRs of the transcripts that those peaks are contained within. For

917 both peaks within *ARRB2* and *USP25,* their matched conservation of peak and UTR are indicated

918 by connected vertical lines. (**C**) Human and mouse normalized GCLiPP density and conservation

919 (PhyloP) across aligned nucleotides of the indicated 3' UTRs. Biochemically shared peaks of

920 GCLiPP read density are indicated in pink. (**D**) HOMER called motifs enriched in biochemically

921 shared peaks. Percentages indicate the frequency of occurrence of the indicated motif in

922 biochemically shared peaks versus non-shared background peaks. P-value indicates HOMER

923 calculated p-value of enrichment. (**E**) Metascape called biological enrichment categories of

924 genes containing biochemically shared peaks. The background set was all genes that contained

925 peaks in both mouse and human GCLiPP datasets that did not contain a shared peak.

926

927

928

929

930

931

932

933
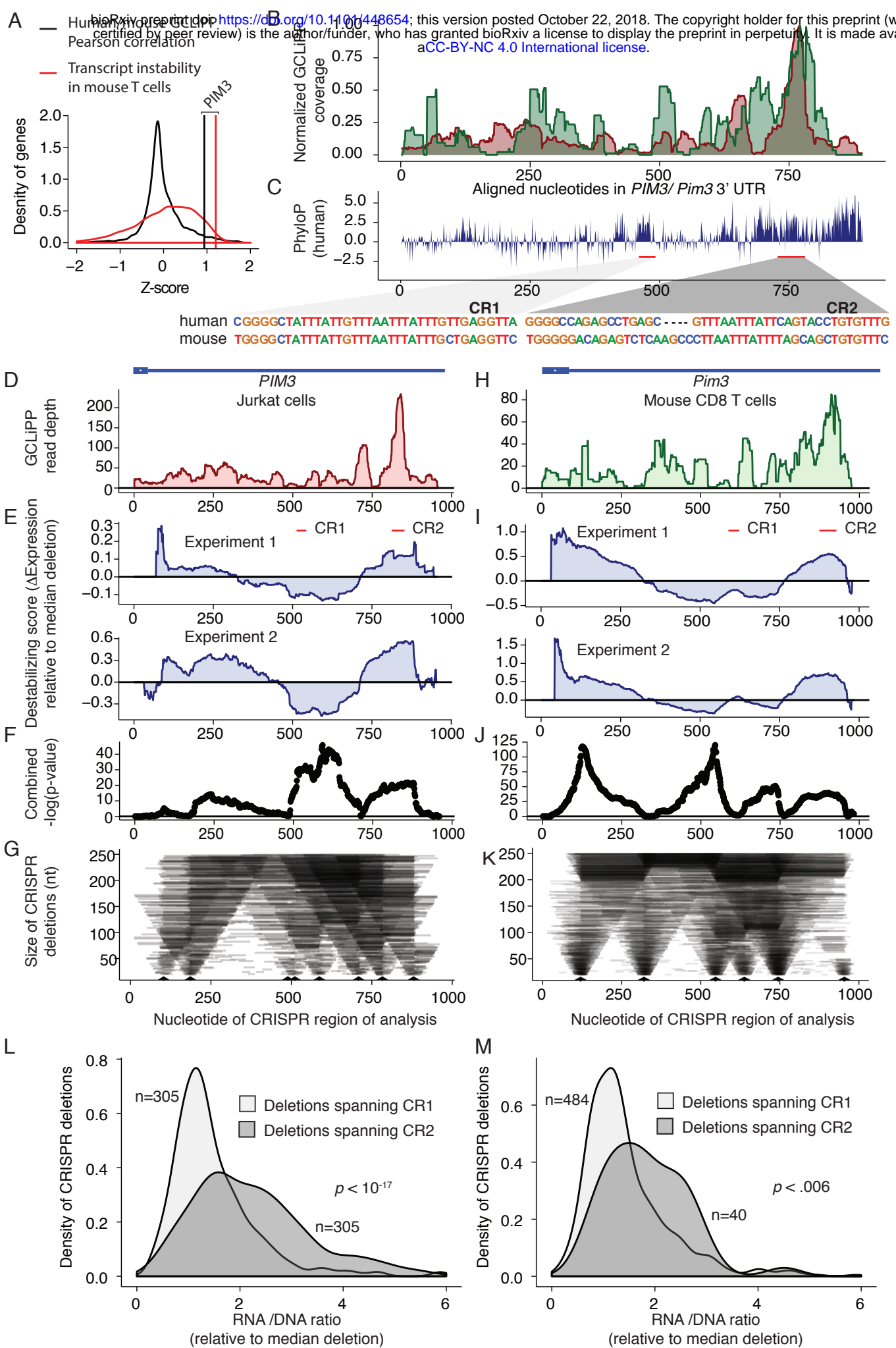
934   **Figure 6. Biochemically and functionally shared post-transcriptional regulation of PIM3 in**

935   **human and mouse cells. (A)** Z-scores of Pearson correlation between mouse and human

936   GCLiPP (black distribution) and transcript instability as measured by comparing transcript read

937   abundance in untreated versus actinomycin treated mouse T cells (red distribution) for 7541

938   genes with matched data. Vertical lines indicate observations for PIM3. (**B**) Normalized human

939   and mouse GCLiPP read density and (**C**) PhyloP across aligned nucleotides of PIM3 3' UTR (as

940   depicted in Figure 5). Insets show sequences of putative regulatory elements. (**D-L**) Dissection

941   of human PIM3 3'UTR in Jurkat T cells (**D**) GCLiPP peaks aligned to schematic illustration of

942   3'UTR. (**E**) Change in expression along the 3'UTR relative to median expression of all possible

943   deletions. Per-nucleotide effect score was calculated by comparing median normalized

944   RNA/gDNA ratio for all shown deletions spanning a given nucleotide with all shown deletions.

945   Experiment 1 and 2 are biological duplicates which were transfected with 80μM or 120μM of

946   gRNAs respectively. Red bars indicate putative ARE-containing cis regulatory elements. (**F**)

947   Unadjusted -log10 p-values from Welch's two sample t-test comparing all deletions spanning a

948   nucleotide with all other deletions across both experiments. (**G**) Size of deletions generated using

949   CRISPR-Cas9. Arrow heads represent gRNA placement. (**H-K**) Dissection of mouse PIM3

950   3'UTR. Data are represented identically to human data, except that mouse primary CD8 T cells

951   were used, and both mouse experiments 1 and 2 used a gRNA concentration of 80μM. (**L**) Effect

952   of deletions spanning putative ARE containing cis-regulatory elements. The RNA/DNA ratio for

953   mutants deleting ARE1 and ARE2 are shown in human Jurkat T cells. (**M**) Same as in (**L**) but

954   using data from mouse primary T cells.

955