RIF1 regulates replication origin activity and early replication timing in B cells

Daniel Malzl[1,6], Mihaela Peycheva[1,6], Ali Rahjouei[2], Stefano Gnan[3], Kyle N. Klein[4], Mariia Nazarova[1], Ursula E. Schoeberl[1], David M. Gilbert[4], Sara C. B. Buonomo[3], Michela Di Virgilio[2], Tobias Neumann[1,5,7] and Rushad Pavri[1,7]

1. Research Institute of Molecular Pathology (IMP), Vienna Biocenter, 1030 Vienna, Austria
2. Max-Delbruck Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany
3. Institute of Cell Biology, University of Edinburgh, Edinburgh EH9 3FF, UK
4. Dept. of Biological Science, Florida State University, Tallahassee FL 32306, USA
5. Present address: Quantro Therapeutics, Vienna Biocenter, 1030 Vienna, Austria.
6. Equal contribution
7. Corresponding authors (tobias.neumann@imp.ac.at, rushad.pavri@imp.ac.at)

ABSTRACT

The mammalian DNA replication timing (RT) program is crucial for the proper functioning and integrity of the genome. The best-known mechanism for controlling RT is the suppression of late origins of replication in heterochromatin by RIF1. Here, we report that in antigen-activated B lymphocytes, RIF1 binds predominantly to early-replicating active chromatin, regulates early origin firing and promotes early replication. RIF1 has a minor role in gene expression and genome organization in B cells. Furthermore, we find that RIF1 functions in a complementary and non-epistatic manner with minichromosome maintenance (MCM) proteins to establish early RT signatures genome-wide and, specifically, to ensure the early replication of highly transcribed genes. These findings reveal new layers of regulation within the B cell RT program, driven by the coordinated activity of RIF1 and MCM proteins.

INTRODUCTION

The faithful and timely replication of the genome is essential for inheriting genetic information and avoiding chromosomal abnormalities. To ensure this, large metazoan genomes initiate replication from several discrete loci, termed origins of replication. Origins are not activated simultaneously across the genome, but rather, in an asynchronous manner referred to as the DNA replication timing (RT) program[1]. A hallmark of the RT program is that genomic A compartments enriched in transcriptionally active genes generally replicate earlier in S phase whereas B compartments harboring silent heterochromatin typically replicate later in S phase[2,3]. Deregulation of RT has been correlated with defects in chromosome condensation, sister chromatid cohesion, gene expression and genome instability [4-6]. Altered RT can disrupt the distribution of active and repressive epigenetic marks causing major alterations in genome architecture[7]. In genetic diseases and cancer, defects in RT have been correlated with deleterious chromosomal translocations[8-12]. Recently, RT has been directly implicated in the biogenesis of oncogenic translocations found in B cell lymphomas and other leukemias[13]. In addition, late replication has been consistently associated with higher rates of mutation across species [11,14-17] and stress-induced delays in replication are a hallmark of common fragile sites in long, transcribed genes[18-20]. Yet, despite its key role in maintaining cellular physiology and genome integrity, our understanding of the mechanisms regulating RT remains incomplete.

Origins are specified in the late G2/M and early G1 phases of the cell cycle via the loading of the origin recognition complex (ORC) and associated factors[21-25]. The subsequent recruitment of the hexameric ring-shaped MCM complex helicase (MCM2-7) licenses these sites for activation[22], resulting in the formation of the pre-replication complex (pre-RC)[21-25]. In S phase, a subset of licensed origins is activated by a set of proteins collectively called replication firing factors, resulting in the initiation of replication[21-24]. A subset of licensed, MCM-bound origins can be activated by firing factors at any given point in S phase[26]. This has led to the idea that firing factors must be recycled to activate subsequent sets of origins later in S phase, thus ensuring timely completion of replication. In support, overexpression of firing factors has been shown to advance the RT of late-replicating chromatin[27-29]. These studies imply that the time of recruitment of the firing factors to a licensed origin determines its RT [30]. In essence, the probability of origin activation in S phase determines the RT of a region, meaning that early RT domains harbor origins that typically fire earlier in S phase whereas late RT domains contain origins that tend to fire later in S phase.

81 The best-studied mechanism of RT control is the suppression of late origin firing by the
82 multifunctional protein, RIF1[31-33]. RIF1 associates with late-replicating chromatin in large, mega-base
83 domains called RIF1-associated domains (RADs) and represses origin firing via protein phosphatase
84 1 (PP1)-mediated dephosphorylation of MCM4[34-36]. Ablation of RIF1 (*Rif1*$^{-/-}$) in human and mouse
85 embryonic stem cells (hESCs and mESCs) and in primary mouse embryonic fibroblasts [31] resulted in
86 a genome-wide loss of early and late RT domain distinction associated with alterations in epigenetic
87 marks and genome compartmentalization[7,35,36]. However, in some *Rif1*$^{-/-}$ cell lines, this RT phenotype
88 was considerably weaker [7], suggesting the existence of additional modes of RT control.

89 We recently showed that in MCM-depleted CH12 cells, a murine B cell line, the RT program was
90 globally deregulated without major changes in transcription or genome architecture[13]. Since RIF1 is
91 the only other factor whose loss could lead to such a major phenotype[7,31,35], we investigated the role
92 of RIF1 in in the B cell RT program. We report the surprising findings that RIF1 is predominantly
93 bound to active chromatin in B cells, regulates early origin activity in these regions and promotes their
94 early replication. In addition, we find that RIF1 acts in a complementary and non-epistatic manner with
95 MCM complexes to drive early replication, especially of highly transcribed regions. In sum, our study
96 reveals an additional regulatory layer within the global RT program and a new role for RIF1 in
97 promoting early replication.

98

99 **RESULTS**

100 **RIF1 regulates early replication in B cells**

101 To measure RT, we performed Repli-seq from early (E) and late (L) S phase fractions in normal and
102 *Rif1*$^{-/-}$ CH12 B cells[37,38] (Fig. S1A). MCM complexes were depleted by infection with lentiviruses
103 expressing short hairpin RNAs (shRNAs) targeting Mcm6 (shMcm6), and an shRNA against LacZ
104 (shLacZ) served as a control for infection and shRNA expression, as described previously[13,39] (Fig.
105 S1B). RT was calculated as the (log2) E/L ratio[40] (Fig. 1A). RT values showed the expected bimodal
106 distribution in shLacZ cells reflecting distinct early (E; positive RT) and late (L; negative RT) domains
107 (Fig. 1A), which was also reflected in the genomic RT profiles (Fig. 1B). In shMcm6 cells, this
108 distinction was globally weakened with all early and late RT values approaching zero (Fig. 1A-B). In

109 *Rif1*$^{-/-}$ shLacZ cells, many early-replicating domains underwent a delay in replication (Fig. 1B). In
110 comparison, late RT domains showed a mixed phenotype in *Rif1*$^{-/-}$ shLacZ cells with some, typically
111 smaller, domains showing advanced replication signatures, and other, typically larger, domains
112 undergoing delayed replication relative to shLacZ cells (Fig. 1A-B). Direct comparison of RT in 20 kb
113 genomic bins between the different conditions showed that most early- and late-replicating bins were
114 strongly shifted towards zero in shMcm6 cells indicative of a deregulation of the RT program (Fig.
115 1C). In *Rif1*$^{-/-}$ shLacZ cells, most early-replicating regions underwent delayed replication to varying
116 degrees (note the red stripe in Fig. 1C), albeit to a lesser extent than in shMcm6 cells. Although late-
117 replicating bins were mildly affected in *Rif1*$^{-/-}$ shLacZ cells, many of these bins underwent slightly
118 delayed RT (Fig. 1C). The latter observation is reminiscent of reports in other *Rif1*$^{-/-}$ cell lines where
119 the RT of many late-replicating domains that were not associated with RIF1 was unaffected by the
120 loss of RIF1 [7].

121 We also performed Repli-seq in a more physiological system, namely, primary activated splenic B
122 cells from *Rif1*$^{+/+}$, RIF1 heterozygous (*Rif1*$^{-/+}$) and *Rif1*$^{-/-}$ mice[41] (Fig. S1C). WT primary B cells had
123 relatively fewer early replicating domains and more mid-replicating domains than shLacZ CH12 cells,
124 resulting in different RT density profiles (Fig. 1D, compare with Fig. 1A). Strikingly, we observed a
125 nearly exclusive effect on early-replicating genomic bins in heterozygous *Rif1*$^{-/+}$ cells indicating that
126 reduced levels of RIF1 protein was sufficient to delay replication of early-replicating regions without
127 majorly affecting late replicating ones (Fig. 1 D-E). This phenotype was exacerbated in homozygous
128 *Rif1*$^{-/-}$ cells as seen by the further delay in replication of early RT domains, which was accompanied
129 by the earlier replication of several late RT domains albeit to varying degrees (Fig. 1 B and D-E).
130 These results, especially from heterozygous cells, highlight a new role for RIF1 in regulating early
131 replication in B cells.

132 To systematically measure changes in RT upon loss of RIF1, we generated three RT states, early
133 (E), middle (M) and late (L), in shLacZ CH12 cells using a Hidden Markov Model (HMM) to segment
134 the genome into 20 kb bins based on their RT value (Fig. 1F). We split the M state at the zero RT
135 value into early-like (E-like) having positive RT values and late-like (L-like) having negative RT values,
136 which resulted in four RT classes (Fig. 1F). To allow direct comparison of RT changes between
137 conditions, we created density plots for each of the four RT states in shLacZ cells displaying the RT of

138 all other experimental conditions in those genomic bins. The results showed that E and E-like bins in
139 shLacZ replicated later in *Rif1*$^{-/-}$ cells, but, importantly, that this was not accompanied by
140 commensurate advances in the RT of L bins suggesting that loss of RIF1 does not globally deregulate
141 RT (Fig. 1G). Indeed, most L bins retained their RT values with only some bins showing advanced RT
142 and some others showing further delays in RT relative to shLacZ cells (Fig. 1G). In contrast, in
143 shMcm6 cells, all shLacZ RT bins shifted towards zero RT values implying that these cells have
144 undergone a global deregulation of the RT program (Fig. 1G). The same HMM-based analysis in
145 primary, splenic B cells revealed that in *Rif1*$^{-/+}$ heterozygotes, there was a strong delay in the RT of E
146 bins associated with a much weaker advance in the RT of L bins and virtually no changes in L-like
147 bins, which further demonstrates that RIF1 primarily regulates early replication in B cells (Fig. 1H-I).
148 These changes were exacerbated in *Rif1*$^{-/-}$ B cells as seen by the further delay in the RT of E bins
149 which as accompanied by the advanced RT of L bins relative to *Rif1*$^{-/+}$ cells (Fig. 1H-I).

150 We conclude that RIF1 functions as modulator of early replication in B cells rather than as a global
151 regulator of RT.

152 **RIF1 plays a minor role in gene expression and genome compartmentalization in B cells**

153 To determine the role of RIF1 in gene expression, we performed RNA-seq from two independent
154 clones (clones 1 and 2) of *Rif1*$^{-/-}$ CH12 cells and identified differentially regulated genes with the
155 DESeq2 software [42] (Fig. S2A-B). Only 22 downregulated and 3 upregulated genes were common
156 between the two clones. A gene ontology (GO) analysis of the common downregulated genes did not
157 reveal enrichments in pathways related to DNA replication, cell division, cell proliferation or genome
158 organization (Fig. S2C), and there were no pathways enriched in the common upregulated genes.

159 RIF1 has been implicated in regulating genome compartmentalization, in part, through
160 deregulation of RT [7]. However, in *Rif1*$^{-/-}$ shLacZ CH12 cells, Hi-C analysis revealed no gross changes
161 in compartment profiles based on evaluation of Hi-C heatmaps where compartments manifest as the
162 checkerboard patterns off the diagonal (Fig. 2A). We quantified the compartment signals based on the
163 first principal component (PC1) of the Hi-C contact matrix where positive PC1 values denote active,
164 early-replicating A compartments and negative PC1 values denote silent, late-replicating B
165 compartments (Fig. 2B-C)[43]. Comparative PC1 analysis showed that although the majority of the bins
166 in *Rif1*$^{-/-}$ shLacZ cells underwent minor changes in their PC1 values relative to shLacZ cells, a few

167 bins did shift substantially, including PC1 sign flips in both directions, which gave a more dispersed
168 appearance to the distribution relative to control cells (Fig. 2B, right). Visualization of PC1 profiles in
169 various genomic regions confirmed that small changes in PC1 occurred in many locations in *Rif1*$^{-/-}$
170 shLacZ cells, but importantly, that these did not always correlate with the changes in RT (Fig. 2C).
171 However, the magnitude of PC1 changes and the frequency of compartment switching (PC1 sign
172 flips) is considerably milder than that reported in other cell types where loss of RIF1 led to substantial
173 compartment switching[7,35]. In comparison, PC1 values in shMcm6 cells were comparable to those in
174 shLacZ cells (Fig. 2B, left and Fig. 2C)[13].

175 We also generated PC1 versus RT density contour plots which allowed us to simultaneously
176 compare the changes in compartmentalization and RT within the same set of genomic 20 kb bins.
177 The density contour distribution profile in *Rif1*$^{-/-}$shLacZ cells showed a marked shift of early RT bins
178 towards delayed RT but without major changes in their PC1 values (Fig. 2D). This implies that the
179 roles of RIF1 in RT and genome architecture are separable in B cells, as they are in other cells[7,35].
180 However, *Rif1*$^{-/-}$ shLacZ cells showed a gain of A-B inter-compartmental interactions with a
181 corresponding decrease in A-A and B-B contacts (Fig. 2E-F). This effect was clearly manifested in
182 fold-change saddle plots where contacts between regions of highest PC1 (early-replicating) and
183 lowest PC1 (late-replicating) were increased in *Rif1*$^{-/-}$ shLacZ cells (Fig. 2F). Thus, RIF1 is important
184 for maintaining normal compartmentalization in B cells by preventing the mixing of A and B
185 compartments. Importantly, this compartmentalization phenotype was observed in mESCs and
186 hESCs where, in contrast to B cells, loss of RIF1 caused a severe deregulation of the RT program[7,35].
187 This suggests that the changes in compartmentalization in *Rif1*$^{-/-}$ cells are conserved between
188 diverse cell lineages, but that these are unrelated to the changes in RT. Moreover, in shMcm6 cells,
189 where RT is globally disrupted to a similar degree as in *Rif1*$^{-/-}$ mESCs and hESCs [7,13,35], we observed
190 relatively minor changes in compartmentalization, as seen by the mild increase in A-A contacts and a
191 similar decrease in B-B contacts (Fig. 2E-F).

192 In sum, the uncoupling of RT from genome organization in B cells and other cells [7,13,35] leads us to
193 conclude that although RIF1 contributes to the normal spatial separation of A and B compartments in
194 B cells, it's role in regulating early RT is unlikely to be directly linked to these structural changes.

195

196 **RIF1 is predominantly located in early-replicating transcribed chromatin in B cells**

197 We next investigated the genomic occupancy of RIF1 in B cells. We performed chromatin
198 immunoprecipitation (ChIP) from primary, mature splenic B cells derived from the *Rif1*$^{FH/FH}$ mouse line
199 wherein RIF1 is endogenously tagged at its C terminus with the Flag and Hemagglutinin (HA)
200 epitopes[31,36]. RIF1 chromatin occupancy was determined via ChIP sequencing (ChIP-seq) with an
201 anti-HA antibody in *Rif1*$^{FH/FH}$ and *Rif1*$^{+/+}$ cells (Fig. S3A).

202 Visual analysis on the genome browser revealed RIF1 binding predominantly in early replicating
203 domains in B cells (Fig. 3A). Comparison with ChIP-seq profiles from *Rif1*$^{FH/FH}$ mouse embryonic
204 fibroblasts (MEFs)[36] showed that while RIF1 localized to broad domains in late-replicating regions in
205 MEFs, in B cells, enrichments were mostly in early-replicating domains (Fig. 3B). We next identified
206 peaks of RIF1 occupancy [44] as well as RADs[45] in both B cells and MEFs. Peaks in both cell types
207 were enriched in early-replicating regions (Fig. 3C-D). However, although RADs in MEFs were mostly
208 late-replicating, as described[36], in B cells, they were largely early-replicating (Fig. 3C-D).

209 We next determined the overlap of RIF1 peaks with nascent transcription measured by precision
210 run-on sequencing (PRO-seq)[46], chromatin accessibility, measured by assay for transposase-
211 accessible chromatin (ATAC-seq)[47], and ChIP-seq of histone H3 acetylated at lysine 27 (H3K27Ac), a
212 mark of active transcription start sites (TSSs) and enhancers. The resulting heatmaps, centered on
213 the RIF1 peak summits, revealed an enrichment of RIF1 in transcribed, H3K27ac-rich accessible
214 chromatin (Fig. 3E). In agreement, ~70% of RIF1 peaks and ~50% of RADs overlapped TSSs or gene
215 bodies (Fig. 3F), and a metagene analysis showed that RIF1 was relatively enriched at TSSs (Fig.
216 3G).

217 We conclude that RIF1 predominantly occupies early-replicating, transcribed chromatin in B cells,
218 suggesting that RIF1 promotes their early replication via direct association.

219 **RIF1 regulates the activity of a subset of origins of replication in active chromatin**

220 The localization of RIF1 in active chromatin coupled with the delayed replication of these domains in
221 *Rif1*$^{-/-}$ cells led us to investigate whether RIF1 played a role in promoting origin firing in activated B
222 cells. To address this, we measured origin activity with short nascent strand sequencing (SNS-seq) in
223 WT and *Rif1*$^{-/-}$ CH12 cells (Fig. S3B). SNS-seq maps the location and relative usage of origins

224 (termed replication initiation sites; ISs) in a population of cells by quantifying the levels of nascent
225 leading strands[48,49]. As described in our previous work [13], ISs and initiation zones were defined based
226 on peak-calling and peak-clustering, respectively, and only ISs within initiation zones were used for
227 analysis.

228 We generated heatmaps of IS read densities in a 4 kb window centered at the IS peak summit
229 and ordered by the IS density in WT cells (Fig. 4B). To visualize how IS density correlated with
230 various chromatin features in the genomic neighborhood, we also generated heatmaps showing the
231 densities of DNase hypersensitive sites (DHSs) which mark accessible chromatin at active promoters
232 and enhancers, H3K36me3, which is enriched in the bodies of transcriptionally active genes, and the
233 repressive mark, H3K9me3 (Fig. 4B). In WT cells, the most active ISs were embedded in active
234 chromatin and gene bodies whereas the least active ISs were associated with H3K9me3 (Fig. 4B). In
235 *Rif1*$^{-/-}$ cells, origin activity was repressed in active chromatin but increased in H3K9me3 domains (Fig.
236 4B). This phenotype is reminiscent of what we previously reported in shMcm6 B cells [13]. Importantly,
237 however, only ~5% of ISs are deregulated (1.5-fold) upon loss of RIF1 compared to ~38% in shMcm6
238 cells (Fig. S3C), which correlates with the magnitude of their respective RT phenotypes (Fig. 1).

239 To quantify the changes in origin efficiency, we calculated fold-changes of ISs (log2 WT/ *Rif1*$^{-/-}$),
240 ranked them from highest to lowest and created five equal classes (quintiles) such that class 1
241 contained the most downregulated ISs in *Rif1*$^{-/-}$ cells and class 5 contained the most upregulated ISs
242 in *Rif1*$^{-/-}$cells. We next generated violin plots displaying the density of ISs (Fig. 4C) or RT values (Fig.
243 4D) in WT and *Rif1*$^{-/-}$ cells within each IS class. The results showed that in WT cells, class 1 ISs were
244 the most active (Fig. 4C) and were mostly early-replicating (Fig. 4D) whereas class 5 ISs were the
245 least active with the majority being late-replicating (Fig. 4C-D). Importantly, class 5 IS densities in
246 *Rif1*$^{-/-}$ cells were comparable to class 1 IS densities in WT cells, suggesting that the upregulated
247 origins in *Rif1*$^{-/-}$ cells (class 5), which are normally the weakest, fire at similar efficiencies as the most
248 active origins in WT cells (class 1) (Fig. 4C).

249 We conclude that RIF1 is required for the optimal activity of a subset of early replicating origins in
250 B cells.

251

252 **RIF1 and MCM complexes act in a complementary manner to regulate the B cell RT program**

253 The differing RT phenotypes in *Rif1*$^{-/-}$ and shMcm6 cells led us to hypothesize that they may function
254 in a non-epistatic manner to drive early replication in B cells. To address this, we infected *Rif1*$^{-/-}$ cells
255 with lentiviruses expressing shMcm6 (*Rif1*$^{-/-}$ shMcm6) or shLacZ (*Rif1*$^{-/-}$ shLacZ) (Fig. S5A) and
256 performed Repli-seq (Fig. S4A). Of note, the converse experiment, that is, depletion of RIF1 in
257 shMcm6 cells was precluded by the fact that the viability of shMcm6 cells was severely compromised
258 upon viral infection with *Rif1*-specific sgRNAs or shRNAs, in line with the observation that cells with
259 limiting MCM proteins are sensitive to stress[50-52].

260 Repli-seq revealed that there was a further loss of early and late RT domain distinction in *Rif1*$^{-/-}$
261 shMcm6 cells compared to *Rif1*$^{-/-}$ shLacZ cells or shMcm6 cells, with a shift of both E and L RT
262 values towards zero (Fig. 5A). An HMM-based classification of RT values, as in Fig. 1D, showed that
263 E bins replicated later, and L bins replicated earlier in *Rif1*$^{-/-}$ shMcm6 compared to shMcm6 cells (Fig.
264 5B). Direct comparison of RT values in 20 kb genomic bins showed that this exacerbation of the RT
265 phenotype in *Rif1*$^{-/-}$ shMcm6 relative to *Rif1*$^{-/-}$ shLacZ and shMcm6 cells was observed in most of the
266 RT bins (Fig. 5C). Additionally, in *Rif1*$^{-/-}$ shMcm6 cells, many large early-replicating domains showed
267 considerable fluctuation in the RT values resulting in a highly fragmented RT profile (Fig. 5D).
268 However, this fragmentation was considerably lower in large late-replicating domains (Fig. 5D). The
269 RT profiles in *Rif1*$^{-/-}$ shMcm6 cells were also marked by switching of RT signatures (both E to L and L
270 to E) and loss of clear domain boundaries (Fig. 5D). These changes, and especially the global
271 fluctuations of early RT values, result in *Rif1*$^{-/-}$ shMcm6 cells acquiring an RT signature distinct from
272 shMcm6 cells (Fig. 5A, C). In sum, these findings reveal an additive effect of MCM depletion and loss
273 of RIF1 on RT in B cells.

274 Despite the changes in RT, gene expression (Fig. S4B-C and Table S2), compartment identity
275 (Fig. 5D and Fig. S4D) and compartment strength (Fig. S4E-F) were not majorly altered in *Rif1*$^{-/-}$
276 shMcm6 cells relative to *Rif1*$^{-/-}$ shLacZ cells, although slight gains in A-B, B-A and A-A interactions
277 were observed, akin to the changes seen in shMcm6 cells (Fig. S4F, compare with Fig. 2F). However,
278 the major changes in RT led to a distinct pattern of PC1 versus RT profiles in in *Rif1*$^{-/-}$ shMcm6 cells

279 relative to in *Rif1*$^{-/-}$ shLacZ cells (Fig. S4G). Thus, the deregulation of the RT program by MCM
280 depletion does not majorly impact genome architecture in normal (Fig. 2) or in *Rif1*$^{-/-}$ cells.

281 To determine whether the increased deregulation of RT in *Rif1*$^{-/-}$ shMcm6 cells was associated
282 with changes in underlying replication origin activity, we performed SNS-seq in *Rif1*$^{-/-}$ shLacZ and
283 *Rif1*$^{-/-}$ shMcm6 cells (Fig. 6A). We further classified ISs based on fold-changes
284 (*Rif1*$^{-/-}$ shLacZ/*Rif1*$^{-/-}$ shMcm6) into quintiles and analyzed the distribution of IS densities (Fig. 6B)
285 and RT values (Fig. 6C) within the IS classes. The most downregulated ISs (class 1) were the most
286 active in *Rif1*$^{-/-}$ shLacZ and were predominantly early-replicating, whereas the most upregulated ISs
287 (class 5) were the least active in *Rif1*$^{-/-}$ shMcm6 cells and were mostly late-replicating (Fig. 6B-C).
288 This is consistent with the phenotype we reported previously in shMcm6 cells[13]. Indeed, SNS-seq
289 profiles showed a characteristic downregulation of origin activity in A compartments accompanied by
290 upregulation in B compartments in both shMcm6 and *Rif1*$^{-/-}$ shMcm6 cells (Fig. 6D). We infer that the
291 additional delay in early replication in *Rif1*$^{-/-}$ shMcm6 cells relative to *Rif1*$^{-/-}$ shLacZ cells is likely due
292 to the further decrease in the efficiency of early-firing origins.

293 We conclude that MCM proteins and RIF1 regulate RT in a complementary and non-epistatic
294 manner in B cells with the contribution of each factor correlating with their impact on origin firing
295 efficiency.

296 **Within early RT domains, the RT of highly transcribed regions is most sensitive to the**
297 **depletion of MCM complexes and, to a weaker extent, the loss of RIF1**

298 A closer examination of the RT profiles revealed that that the extensive fragmentation seen in *Rif1*$^{-/-}$
299 shMcm6 cells was also visible to a lesser extent in shMcm6 and to the weakest extent in *Rif1*$^{-/-}$
300 shLacZ cells (Fig. 5C-D). Furthermore, we observed a high degree of similarity in the fragmentation
301 patterns between the different conditions. This suggested that these were not random fluctuations of
302 the RT signals but reflected an underlying mechanism supporting early replication that is reliant on
303 MCM proteins and RIF1. We therefore investigated whether the levels of nascent transcription within
304 genomic RT bins could explain these altered RT profiles. This reasoning was based on the
305 observations that fragmentation was typically seen in early RT domains where most transcribed

306 genes are located, and secondly, that in *Rif1*$^{-/-}$, shMcm6 and *Rif1*$^{-/-}$ shMcm6 cells, early origins in
307 active chromatin were downregulated despite there being no major changes in transcription.

308 To address this, we divided the genome into 20 kb bins and extracted all early-replicating (E)
309 bins in shLacZ cells using the HMM-based approach described in Fig. 1C. Within this group, we
310 defined transcribed bins as those having a PRO-seq density (RPM) ≥ 10 and these were divided into
311 three equal sub-groups (tertiles) based on their RPM values, termed High (199-3358 RPM), Medium
312 (78-198.9 RPM) and Low (10-77.9 RPM) (Fig. S4H). The remaining early-replicating bins were termed
313 Untranscribed (0-9.9 RPM) (Fig. S4H). To compare changes in RT values, we generated RT
314 heatmaps for the four transcription-based groups and ranked each of them by decreasing WT PRO-
315 seq read density such that, effectively, the entire set of bins were ranked from highest to lowest PRO-
316 seq density in shLacZ cells (Fig. 7A). In addition, we generated density plots to visualize the
317 differences in the distribution of RT values between the four groups in all experimental conditions (Fig.
318 7B).

319 In shLacZ cells, all four transcription-based groups showed nearly identical distributions of RT
320 values, indicating that transcribed and non-transcribed regions normally have similar probabilities of
321 origin activation and early replication (Fig. A-B). The High and Medium groups appeared very similar
322 in terms of the RT values between bins (Fig. 7A) and the overall distribution profiles (Fig. 7B).
323 However, in shMcm6 and *Rif1*$^{-/-}$ shMcm6 cells, the High and Medium transcribed bins showed the
324 strongest delays in RT whereas the Untranscribed group showed the least delays (Fig. 7A-B). These
325 differences were observed across all bins (Fig. 7A). The shift in RT between the High and
326 Untranscribed groups was ~2-fold (log2 RT ~1), as measured by the distance between the modes of
327 the distributions (marked by dotted lines in Fig. 7B). The lowly transcribed regions had an
328 intermediate distribution of RT values between the High/Medium and Untranscribed groups in both
329 shMcm6 and *Rif1*$^{-/-}$ shMcm6 cells (Fig. 7A-B). Thus, the strongest delays in RT are associated with
330 regions harboring higher levels of transcription. In *Rif1*$^{-/-}$ shLacZ cells, all groups had a broad spread
331 of RT values with similar distribution profiles (Fig. 7A-B). However, a closer inspection of the heatmap
332 and the modes of the distributions revealed a slightly greater delay in RT for the high and medium
333 transcribed regions compared to the lowly transcribed and non-transcribed bins (Fig. 7B).

334 Taken together, these results suggest that the RT of transcribed regions, and hence the
335 underlying activity of early origins, is more sensitive to the reduction of MCM proteins than to the loss
336 of RIF1, which correlates with the magnitude of the fragmentation observed in the early RT domains
337 (Fig. 5C-D). We conclude that the fragmented appearance of early RT profiles is due to underlying
338 transcription, with stronger delays correlating with higher levels of transcription. These results suggest
339 that transcriptional strength negatively regulates the probability of early origin firing at active genes,
340 and that this is normally overcome by the presence of a normal complement of MCM proteins and, to
341 a lesser extent, RIF1.

342

343 **DISCUSSION**

344 Our study reveals two new layers of regulation within the RT program. First, a mechanism that
345 specifically promotes early RT via the functional repurposing of RIF1. Second, the importance of
346 normal levels of MCM proteins, with a minor role for RIF1, in ensuring early RT of transcribed regions.
347 We propose that since mutation rates are known to be higher in late-replicating regions [11,14-17] and
348 delays in RT are the major cause of common fragile sites in long transcribed genes[18-20], ensuring
349 early RT may be a safeguarding mechanism to ensure genome integrity and normal cellular functions
350 post mitosis. As explained below, this may be particularly relevant for antigen-activated B cells.

351 How might transcription negatively impact the efficiency of replication origins? There is evidence
352 from *in vitro* studies that RNA polymerase (RNAP) complexes can push MCM complexes along DNA
353 [53,54]. Indeed, MCM and ORC proteins were found to be relatively depleted in gene bodies[55] and
354 Okazaki fragment sequencing (OK-seq) studies have reported that replication initiation zones occur
355 largely in the intergenic regions flanking transcribed genes[56]. A recent single molecule imaging study,
356 where replication-transcription encounters were reconstituted using purified proteins, may be
357 instructive in this regard [54]. This study found that T7 RNAP could efficiently push DNA-bound ORC,
358 OCCM (an intermediate of the pre-RC consisting of ORC, Cdc6, Cdt1 and a single MCM hexamer)
359 and MCM double hexamers (the configuration within the fully assembled pre-RC). However, whereas
360 OCCM and double hexamers were rarely ejected by RNAP, ORC alone was frequently evicted by
361 RNAP. Moreover, most of the ORC molecules repositioned by RNAP were unstable [54]. Since ORC
362 loading is the first step in origin licensing [25], the inference is that the labile binding of ORC makes

363 licensing in transcribed regions inherently less efficient than in non-transcribed regions, but that this is
364 overcome by the association of MCM complexes, which minimizes the loss of single ORC complexes
365 by RNAP.

366 Importantly, our data show that, normally, the distribution of early RT values is similar between
367 transcribed and non-transcribed regions. Hence, we propose that, in WT cells, the large pool of MCM
368 complexes ensures that the loading of ORC is rapidly followed by the assembly of the OCCM and
369 pre-RC such that the loss of single ORC complexes by RNAP is minimized. This ensures that both
370 transcribed and non-transcribed regions have a similar probability of early origin activation and early
371 replication in WT cells. When MCM complexes are limiting, licensing in both transcribed and non-
372 transcribed regions is reduced. However, given the inhibitory effect of transcription on ORC stability,
373 the formation of OCCM and pre-RC formation in highly transcribed regions will be inefficient, allowing
374 for more eviction of ORC by RNAP. This will reduce early origin licensing efficiencies in highly
375 transcribed chromatin to a greater degree than in non-transcribed chromatin, leading to our
376 observation of a higher probability of early replication in non-transcribed regions relative to transcribed
377 regions.

378 Although the mechanism by which RIF1 regulates early origin firing in B cells remains to be
379 deciphered, RIF1 can interact with MCM complexes in activated B cells [38]. Intriguingly, a previous
380 study in human cells reported a role for RIF1 in promoting origin licensing by protecting ORC1 from
381 phosphorylation-mediated degradation [57] . Thus, it is plausible that RIF1 prevents premature
382 disassembly of OCCM and pre-RC complexes in B cells. In addition, although RIF1 binds broadly
383 across early-replicating chromatin domains in B cells, ~70% of RIF1 peaks and ~50% of RADs in B
384 cells are in genes (Fig. 3). This suggests that the protective function of RIF1 in origin licensing may be
385 more pronounced in highly transcribed regions. According to this model, the loss of RIF1 would lead
386 to fewer licensed origins in highly transcribed regions compared to neighboring non-transcribed
387 regions within the same early-replicating domains, resulting in non-transcribed regions replicating
388 slightly later than highly transcribed regions, as we observe.

389 Given that RIF1 has not been previously implicated in regulating early replication, our study
390 raises the question of why antigen-activated B cells have functionally repurposed RIF1 in this manner.
391 Antigen-activated B cells are amongst the fastest proliferating cells in the body with a cell cycle

392 duration of 6-8 h *in vivo* within germinal centers and 8-12 h in culture[58-60]. A unique feature of these
393 cells is that they undergo extremely high rates of somatic hypermutation at the highly transcribed
394 immunoglobulin genes, a necessary event in antibody maturation upon infection or vaccination.
395 Genome instability is further elevated by the fact that somatic hypermutation also occurs at many
396 other transcribed loci, including proto-oncogenes like *BCL6* and *MYC* [61-64], which can result in
397 oncogenic translocations typical of most mature B cell cancers [65-67]. Moreover, the DNA repair
398 pathways activated by somatic hypermutation result in single- and double-strand breaks as well as
399 singe-strand patches, all of which are impediments for replication[68,69]. Therefore, these B cells
400 proliferate in a highly genotoxic environment where replication stress is elevated. Under such
401 conditions, it is plausible that B cells have functionally repurposed RIF1 as an additional mechanism
402 to safeguard its genome by enforcing the early replication of active genes, which would help to
403 decrease mutations and common fragile sites associated with late replication or delayed replication of
404 long genes[11,14-20]. It is also possible that this function of RIF1 is masked in other cell types where the
405 dominant mode of RT regulation is that of RIF1-mediated suppression of origins in late-replicating
406 chromatin[7,34,35].

407

408 **METHODS**

409 Cell culture

410 CH12 cells were maintained in RPMI medium with 10% fetal bovine serum (FBS), glutamine, sodium
411 pyruvate, Hepes and antibiotic/antimycotic mix. LentiX packaging cells were maintained in DMEM
412 medium with 10% FBS and antibiotics.

413 Transfections and infections

414 Transfection of LentiX cells with shRNAs against LacZ and Mcm6 followed by infection of CH12 cells
415 with lentiviral supernatants was done exactly as described in our previous study[39].

416 Sample preparation for replication timing (RT) analysis

417 Repli-seq was performed as previously described[40]. In brief, two million asynchronously dividing cells
418 were seeded and incubated with 100μM BrdU (Sigma) for 2h in a light protected environment to

419 maintain BrdU stability. Cells were fixed and incubated with a mix of RNase A (Invitrogen) and
420 propidium iodide (Sigma) for 30 min (light protected). For each sample, three fractions were sorted:
421 G1 phase, early S phase and late S phase cells, and for each fraction, two independent samples of
422 50,000 cells (technical replicates) were sorted on a Sony SH800S Cell Sorter. Sorted cells were lysed
423 with Proteinase K buffer overnight. Extracted DNA was sonicated for 9 min in a Diagenode Bioruptor
424 resulting in 100-500 bp DNA fragments as determined on an agarose gel. Sonicated DNA was
425 subjected to end-repair and adapter ligation using the NEBNext® Ultra™ II DNA Library Prep Kit
426 (NEB) following the NEB protocol. Adapter-ligated DNA was incubated with 25 μg/ml of anti-BrdU
427 antibody (BD Pharmingen) for 4h with rotation followed by incubation with 40 µg of anti-mouse IgG
428 antibody (Sigma) for 1h with rotation (light protected). DNA was precipitated via
429 Centrifugation at 16,000*g* for 5 min at 4°C. Pellet was resuspended in 200 µl of digestion buffer(for 50
430 ml of digestion buffer, combine 44 ml of autoclaved double-distilled water, 2.5 ml of 1M Tris-HCl, pH
431 8.0, 1 ml of 0.5M EDTA and 2.5 ml of 10% SDS) with freshly added 0.25 mg/ml proteinase K and
432 incubated overnight at 37°C. The immunoprecipitated DNA was used for Repli-qPCR or next
433 generation sequencing (Repli-seq). Libraries that were successfully validated by Repli-qPCR were
434 sequenced on an Illumina HiSeq 2500 machine (50bp, single-end). Up to 12 barcoded samples were
435 pooled per lane.

436 Chromatin immunoprecipitation for RIF1 occupancy in primary B cells

437 Chromatin immunoprecipitation was performed as described [36]. In summary, primary B cells were
438 isolated from Rif1$^{FH/FH}$ and C57BL/6 (Rif1$^{WT/WT}$) mice spleens using anti-CD43 MicroBeads (Miltenyi
439 Biotec) and expanded in complete RPMI containing 5 µg/ml LPS (Sigma-Aldrich) and 5 ng/ml mouse
440 recombinant IL-4 (Sigma-Aldrich) to allow B cell activation and class switch recombination to IgG1. B
441 cells were harvested 72h after activation and 4 x 10^7 cells were cross-linked by using 2 mM
442 disuccinimidyl glutarate (ThermoFisher 20593) in PBS for 45 min and 1% formaldehyde for 5 min
443 (Thermo Scientific 28908). The reaction was quenched with 0.125 M glycine. Cells were washed
444 thrice with ice-cold PBS and lysed in SDS lysis buffer. Chromatin fragmentation was performed using
445 a Covaris E220 sonicator to obtain fragments between 200 and 600 bp. Chromatin was quantified
446 with a ND-1000 NanoDrop spectrophotometer. Immunoprecipitation was performed with 0.5 µg of
447 anti-HA antibody (Santa Cruz sc-7392) and 50 µl of Dynabeads protein G (Thermofisher 10003D) and

448 25 µg chromatin. ChIP libraries were prepared by NEBNext ultra II DNA library preparation kit (NEB
449 E7645L) and sequenced on one lane of a NovaSeq 6000 (Illumina) machine.

450 RNA-seq from CH12 cells

451 RNA-seq was performed on CH12 and two independent RIF1-deficient (RIF1$^{-/-}$) CH12 clonal
452 derivatives [37] with three replicates. Cells were expanded in complete RPMI and 1 million cells were
453 collected by centrifugation. RNA was extracted with TRIzol (Invitrogen) according to manufacturer's
454 instructions. TruSeq RNA Library Prep Kit v2 (Illumina) was used to prepare a whole-transcriptome
455 sequencing library and sequenced on one lane of a NovaSeq 6000 SP (Illumina) machine.

456 Isolation of short nascent strands (SNSs)

457 SNSs were isolated following an established protocol[48] kindly provided by Dr. Maria Gomez (CBMSO,
458 Madrid) and also described in our previous study (Wiedemann et al, 2016). In brief, 200 million cells
459 per replicate from asynchronous cell cultures were harvested and genomic DNA was extracted. DNA
460 was denatured at 95°C for 10 minutes and then subjected to size fractionation via 5-20% neutral
461 sucrose gradient centrifugation (24,000g for 20h). Fractions were analyzed by alkaline agarose gel
462 electrophoresis and those in the 500-2000 nt range were pooled. Prior to all following enzymatic
463 treatments, ssDNA was heat denatured for 5 min at 95°C. DNA was phosphorylated for 1h at 37°C
464 with T4 Polynucleotide kinase (NEB or made in-house by the Molecular Biology Service, IMP). To
465 enrich for nascent DNA strands, the phosphorylated ssDNA was digested overnight at 37°C with
466 Lambda exonuclease (NEB or made in-house by the Molecular Biology Service, IMP). Both T4 PNK
467 and Lambda exonuclease steps were repeated twice for a total of three rounds of phosphorylation
468 and digestion. After the final round of digestion, the DNA was treated with RNaseA/T1 mix (Thermo
469 Scientific) to remove 5' RNA primers and genomic RNA contamination. DNA was purified via phenol-
470 chloroform extraction and ethanol precipitation This material was either used directly for qPCR or
471 further processed for library preparation.

472 SNS library preparation for SNS-seq

473 SNSs prepared as described above were converted to double stranded DNA (dsDNA) via random
474 priming with random hexamer primer phosphate (Roche) and ligation with Taq DNA ligase (NEB)[48].
475 DNA was checked on a fragment analyzer and 50ng was used for library preparation with the

476 NEBNext® Ultra™ II DNA Library Prep Kit (NEB) following the manufacturer's protocol. Libraries were
477 barcoded using the NEBNext® Singleplex Oligos (NEB) as per the NEB protocol which allowed 4-8
478 libraries to be pooled per run. Sequencing was performed on an Illumina HiSeq 2500 machine (50 bp,
479 single-end).

480 In situ Hi-C

481 Hi-C was performed as previously described[43] with minor modifications. In brief, 5 million cells were
482 crosslinked with 1% formaldehyde (Sigma) for 10 min and quenched with 0.6M glycine for 5 min.
483 Cells were lysed with Hi-C lysis buffer for 1h on ice and nuclei were collected by centrifugation. Nuclei
484 were digested with 375U of MboI (NEB) overnight at 37°C with rotation. Biotin-14-dATP (Life
485 Technologies) was incorporated for 1h at 37°C with rotation. Ligation of overhangs was performed
486 with 20,000U of T4 DNA ligase (NEB) for 4h at room temperature with rotation. Nuclei were pelleted
487 and reverse-crosslinked overnight. Purified DNA was sonicated for 14 min in a Diagenode Bioruptor
488 to obtain a size range of 200-700bp. This material was purified using Agencourt AMPure XP beads
489 (Beckman Coulter). Between 8-15 μg of DNA was incubated with 100μl (~10 mg) of Dynabeads
490 MyOne Streptavidin C1 (Invitrogen) for 15 min with rotation. End repair and adapter ligation using the
491 NEBNext® Ultra™ II DNA Library Prep Kit was performed on-beads following the kit manual. The
492 adapter-ligated DNA was washed, eluted and PCR-amplified with KAPA 2X HiFi HotStart ReadyMix
493 (Kapa Biosystems) and the NEBNext Multiplex Oligos for Illumina® (Dual Index Primers Set 1). Four
494 pooled, barcoded samples were sequenced on an Illumina NovaSeq 6000 machine (50 bp, paired-
495 end).

496 PRO-seq

497 PRO-seq was performed as described previously[70] with minor modifications. To isolate nuclei, CH12
498 cells and *Drosphila* S2 cells were resuspended in cold Buffer IA (160mM sucrose, 10mM Tris-Cl pH 8,
499 3mM CaCl2, 2mM MgAc2, 0.5% NP-40, 1mM DTT added fresh), incubated on ice for 3 min and
500 centrifuged at 700g for 5 min. The pellet was resuspended in nuclei resuspension buffer NRB (50mM
501 Tris-Cl pH 8, 40% glycerol, 5mM MgCl2, 0.1mM EDTA). For each run-on, 10 million CH12 nuclei were
502 spiked with 10% *Drosophila* S2 nuclei in a total of 100μL NRB and incubated at 30°C for 3 min with
503 100 μL 2x NRO buffer including 5μl of each 1mM NTP (biotinylated ATP and GTP, and unlabelled
504 UTP and CTP). The following steps were performed as described[70] with the following changes: (1) we

505 used different adapters, namely, 3` RNA adapter
506 5Phos/NNNNNNNGAUCGUCGGACUGUAGAACUCUGAAC/3InvdT-3' and 5'RNA adapter: 5'-
507 CCUUGGCACCCGAGAAUUCCANNNN-3'. (2) 3' and 5' ligations which were done at 16°C overnight,
508 and (3) CapClip pyrophosphatase (Cellscript) used for 5' decapping. RNA was reverse transcribed by
509 SuperScript III RT (Invitrogen) with RP1 Illumina primer to generate cDNA libraries. Libraries were
510 amplified using barcoding Illumina RPI-x primers and the universal RP1 and KAPA HiFi Real-Time
511 PCR Library Amplification Kit. Amplified libraries were subjected to electrophoresis on 2.5% low
512 melting agarose gel and amplicons from 150-350 bp were extracted from the gel, multiplexed and
513 sequenced on Illumina platform NextSeq 550 SR75. Bioinformatics analyses were performed as
514 described[70] but additionally the random 8-mer was used to exclude PCR duplicates and only
515 deduplicated reads were aligned.

516 Biological and technical replicates

517 For biological replicates, different frozen vials of CH12 cells were thawed and kept separate
518 throughout the course of the experiment. Replicates for all next-generation sequencing experiments
519 were derived in this manner. Technical replicates, where used (such as in RT-qPCR or ChIP-qPCR),
520 were subsets of the biological replicate.

521 Statistical analyses

522 For correlation scatter plots, the Pearson correlation coefficient was calculated to determine the
523 degree of correlation. In all other cases, the two-tailed Student's t test was used for statistical
524 significance.

525

526 **Bioinformatics**

527 Main next generation sequencing data analysis workflows for SNS-seq, Repli-seq and Hi-C have
528 been wrapped with *Nextflow* workflow language [71] and are centrally available at
529 https://github.com/pavrilab[13]. The usage of Docker and Singlularity containers ensures portability,
530 reproducibility and reliability for all workflows. Along with integration of GitHub repositories for self-
531 contained pipelines all workflows are easily ported to all major HPC computing platforms such as
532 SGE, SLURM, AWS. Furthermore, continuous checkpoints for pipeline execution allow for resuming

533 and automatic retrial of failed steps. All workflows produce elaborate QC-reports and out-of-the box,
534 resources consumption reports to allow tailoring resources requirements to your datasets which
535 especially for Hi-C datasets can vary by several orders of magnitudes.

536 Processing of raw Repli-seq, Hi-C, SNS-seq and RNA-seq data was done as described
537 previously.[13]

538 Generating heatmaps for IS peak summit neighborhoods

539 Sequencing coverage normalized bigWig tracks were generated from raw aligned reads of NGS data
540 sets with deeptools bamCoverage v3.3.0 [72] using command-line parameters --normalizeUsing CPM --
541 exactScaling and --ignoreDuplicates. Next, we computed the position of the replication start site
542 (RSS) as the genomic coordinate of the maximum pile-up of the merged tracks of compared
543 conditions. The generated summit positions were then grouped by their associated peak class and
544 sorted in decending order on the log2-ratio of CPM values of treatment versus control. The sorted
545 summits were then used as reference points for deeptools computeMatrix v3.3.0 [72] to compute signal
546 distributions for previously generated CPM-normalized bigWig tracks within a 4kb region centered on
547 the summit using a binsize of 50 bp and setting missing values to zero. The results were then plotted
548 using the associated plotHeatmap command.

549 PRO-seq analysis

550 The 3' end sequence of the reads (NNNNTGGAATTCTCGGTGCC) were removed using cutadapt
551 v1.4.2 [73] and 9 nucleotides from their 5′ ends, containing the random 8mer and *in vitro* run-on
552 nucleotide, were removed. The trimmed reads were reverse complemented using
553 fastx_reverse_complement (http://hannonlab.cshl.edu/fastx_toolkit; version 0.0.13) followed by
554 deduplication based on the 8mer sequence. Trimmed reads longer than 18 bp were aligned to a
555 hybrid mouse drosophila genome (mouse genome build NCBI m37, GCA_000001635.18, drosophila
556 FlyBase release 5) using bowtie v1.0.0 [74] with -v 2 --best --strata --tryhard -m 1 --chunkmbs 256.
557 Unique mappers from the resulting BAM file were used to create bigwigs with deeptools
558 bamCoverage 3.3.0.

559

560

561 RT domain analysis

562 20kb binned RT tracks for all conditions were segmented into three states using hmm_bigwigs
563 (https://github.com/gspracklin/hmm_bigwigs). Fitted states were then mapped to either early (E), mid
564 (M) or late (L) replication timing based on the RT value distribution of each state (E for RT > 0, L for
565 RT < 0 and M for RT ~ 0). M state bins were split at 0 into L-like (values < 0) and E-like (values > 0)
566 state bins.

567 Generating saddle plots from Hi-C data

568 Saddle plots were computed using cooltools v0.3.2. In brief, we subdivided all bins of a 20Kb KR-
569 normalized contact matrix into 50 equal-sized groups based on the bins compartment signal as
570 derived from the eigenvector of the WT data, where group 1 has the lowest signal (i.e. most B) and
571 group 50 has the highest signal (i.e. most A). Subsequently,we compute the mean observed/expected
572 value for each pair of groups and plot it as a 50 x 50 matrix. Similarly, replication timing and H3K9me3
573 data can be used for bin group assignment.

574 RIF1 peak and RIF1 associated domains (RADs) analyses

575 Peaks and domains (RADs) were called as described previously [75] additionally calling narrow peaks
576 with MACS2. In brief, raw ChIP-seq and input reads in MEFs and primary B cells were trimmed with
577 trim_galore v0.6.4 (https://github.com/FelixKrueger/TrimGalore) setting minimum length to 18 bases
578 and default arguments otherwise and aligned to the mm9 reference genome with bowtie v1.0.0 [74] with
579 -S --trim5 0 --trim3 0 -v 2 --best --strata --tryhard -m 1 --phred33-quals --chunkmbs 256.
580 Subsequently, RIF1 narrow and broad peaks were called for each replicate with MACS v2.2.6 [76], RIF1
581 associated domains (RADs) were called using EDD v1.1.19 [77]. RADs, narrow and broad peaks were
582 merged using bedtools v2.27.1 [78] and only those peaks called in both replicates were retained for
583 downstream analysis. Genome coverage tracks were computed with deeptools bamCoverage v3.3.0
584 [72] using command-line parameters --normalizeUsing CPM --exactScaling –binSize 50 and –
585 ignoreDuplicates. Log2 ratio tracks were computed with deeptools bamCompare v3.3.0 [72] using
586 command-line parameters --scaleFactorsMethod readCount --operation log2 --pseudocount 1 –
587 binSize 50 –ignoreDuplicates.

588

589 Summary of all NGS data used

590 *Table 1: Summary of used data sets in this study including public Accession numbers*

Sample Name	Data type	Source	Accession
CH12 shLacZ	SNS-seq	GEO	GSE161822
CH12 shMcm6	SNS-seq	GEO	GSE161822
CH12 WT	SNS-seq	This study	-
CH12 Rif1$^{-/-}$	SNS-seq	This study	-
CH12 Rif1$^{-/-}$ shLacZ	SNS-seq	This study	
CH12 Rif1$^{-/-}$ shMcm6	SNS-seq	This study	
CH12 H3K4me1	ChIP-seq	GEO/UCSC	GSM946546/wgEncodeEM002370
CH12 H3K4me3	ChIP-seq	GEO/UCSC	GSM946528/wgEncodeEM002366
CH12 H3K9me3	ChIP-seq	GEO/UCSC	GSM946548/wgEncodeEM002372
CH12 H3K27ac	ChIP-seq	GEO/UCSC	GSM1000117/wgEncodeEM003167
CH12 H3K36me3	ChIP-seq	GEO/UCSC	GSM946530/wgEncodeEM002364
CH12 DHS-seq	DHS-seq	GEO/UCSC	GSM1014153/wgEncodeEM003416
CH12 shLacZ	Hi-C	GEO	GSE161822
CH12 shMcm6	Hi-C	GEO	GSE161822
CH12 Rif1$^{-/-}$ shLacZ	Hi-C	This study	-
CH12 Rif1$^{-/-}$ shMcm6	Hi-C	This study	-
CH12 shLacZ	Repli-seq	This study	-
CH12 shMcm6	Repli-seq	This study	-
CH12 Rif1$^{-/-}$ shLacZ	Repli-seq	This study	-
CH12 Rif1$^{-/-}$ shMcm6	Repli-seq	This study	-
CH12 shLacZ	RNA-seq	GEO	GSE161822
CH12 shMcm6	RNA-seq	GEO	GSE161822
CH12 Rif1$^{-/-}$ shLacZ	RNA-seq	This study	-
CH12 Rif1$^{-/-}$ shMcm6	RNA-seq	This study	-
CH12 shLacZ	PRO-seq	This study	-
primary B/CH12 input track	ChIP-seq	GEO	GSM2058441

primary B Rif1$^{FH/FH}$	ChIP-seq	This study	-
primary B Rif1$^{WT/WT}$	ChIP-seq	This study	-
primary B WT Repli-seq	Repli-seq	This study	-
primary B Rif1$^{-/+}$ Repli-seq	Repli-seq	This study	-
primary B Rif1$^{-/-}$ Repli-seq	Repli-seq	This study	-
primary B WT ATAC-seq	ATAC-seq	GEO	GSE132029
primary B WT PRO-seq	PRO-seq	This study	-
primary B WT H3K9me3	ChIP-seq	This study	-

591 Summary of software used in the study

592 All custom Python scripts and tools, except for scripts using ClusterScan, were developed, tested and
593 executed using Python v3.7.3. ClusterScan and related scripts were executed using Python v2.7.13.
594 The used software and packages are summarized in Table , Table 3 and Table 4.

595 *Table 2: Summary of standalone software*

Software	**Version**	**Usage**
bedtools	2.27.1	general operations on sets of genomic regions
bowtie	1.0.0/1.2.3	Alignment of reads
bowtie2	2.2.9	Alignment of reads
bwa	0.7.17	Alignment of Repli-seq data
ClusterScan	0.2.1	Clustering of initiation sites
cutadapt	2.6	Adapter trimming
deeptools	3.3.0	Read counting in regions and heatmap generation
DESeq2	1.22.2	Differential expression analysis
EDD	1.1.19	Calling of Rif1 associated domains
epic2	0.0.41	Peak calling for broad peak type data
FastQC	0.11.5	Quality control of sequencing reads
featureCounts	2.0.0	Computing RNA-seq data counts

HICUP	0.7.3	Processing of Hi-C data
hmm_bigwig	1.3	Segmentation of genome into RT states
HOMER	4.10	Hi-C Eigenvector and TAD annotation
UCSCkentUtils	356	Conversion between file formats
MACS2	2.2.6	Peak calling for narrow peak type data
Python2	2.7.13	Execution of ClusterScan and associated scripts
Python3	3.7.3	General script execution
R	3.5.1	Execution of R code
samtools	1.9	General operations on aligned reads
STAR	2.4.2a	Alignment of RNA-seq data
Trim_galore	0.6.4	Quality control of sequencing data

Table 3: Summary of R packages

Package	***Version***	***Usage***
ashr	2.2-47	Differential expression analysis

Table 4: Summary of Python packages

Package	**Version**	**Usage**
argparse	1.1	Passing command-line arguments to scripts
BioPython	1.77	Mapping of EntrezID to gene symbol
cooler	0.8.3	Hi-C matrix generation and storage
cooltools	0.3.2	Generic Hi-C downstream analysis
genomepy	0.8.3	Genome masking in Python
krbalancing	0.0.5	HiCExplorer implementation of KR for matrix correction
logging	0.5.1.2	Printing information to stdout
matplotlib	3.1.0	Generation of general plots
matplotlib_venn	0.11.5	Generation of Venn diagrams

numpy	1.16.4	General computations on data in matrices
pandas	1.1.0	General computations on tabulated data
pybedtools	0.8.0	Using bedtools within Python
pyBigWig	0.3.16	Generation of bigWig files
pysam	0.15.2	Using samtools in Python
re	2.2.1	Regular expressions for efficient text parsing
rpy2	3.0.5	Using R packages in Python
scipy	1.3.0	General statistical computation in Python
seaborn	0.11.1	Kernel density estimation of scatter plots
tables	3.5.2	Reading of HDF5 files

Data availability

All next generation sequencing data (Repli-seq, RNA-seq, Hi-C, ChIP-seq and SNS-seq) has been deposited in GEO under accession number GSE228880.

ACKNOWLEDGEMENTS

We are grateful to the Vienna Biocenter Core Facilities (VBCF) for next-generation sequencing and the IMP/IMBA core facilities especially the animal house, bio-optics, molecular biology service and proteomics service. We are grateful to Roman Stocsits and Dr. Anton Goloborodko for advice on Hi-C data analysis. This work was funded by Boehringer Ingelheim, The Austrian Industrial Research Promotion Agency (Headquarter Grant FFG-834223), grants from the Austrian Science Fund to RP (FWF P 29163-B26) and UES (FWF T 795-B30), and the Helmholtz-Gemeinschaft Zukunftsthema "Immunology and Inflammation" ZT810 0027 to MDV and AR.

AUTHOR CONTRIBUTIONS

DM performed most of the bioinformatics analyses, established pipelines and analyzed data. MP performed Repli-seq, Hi-C, RNA-seq and SNS-seq, and analyzed data. AR performed RIF1 ChIP-seq from primary B cells and RNA-Seq in CH12 cells. SG and KNK performed Repli-seq from primary B cells. US performed PRO-seq. MN collaborated on Repli-seq from CH12 cells. DMG, SCBB and MDV

619 contributed personnel, resources, extensive discussions and critical reading of the manuscript. TN
620 and RP co-supervised the project and analyzed data. RP conceived the project and wrote the
621 manuscript with inputs from all authors.

622

623 **Supplementary Table Legends**

624 **Table S1: RNA-seq analysis from two independent *Rif1*$^{-/-}$ CH12 clones relative to WT cells**

625 The first two tabs contain the analysis of all genes from *Rif1*$^{-/-}$ CH12 clones 1 and 2 analyzed using
626 the DESeq2 software[42]. The next two tabs list the common two-fold downregulated (n = 22) and
627 upregulated (n = 3) genes.

628

629 **Table S2: RNA-seq analysis from *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 CH12 cells**

630 The table contains the analysis of all genes analyzed using DESeq2[42].

631

632

633

634

635

636

637

638

639

640

641

642

643 **REFERENCES**

644 1. Marchal, C., Sima, J. & Gilbert, D.M. Control of DNA replication timing in the 3D genome. *Nat*
645 *Rev Mol Cell Biol* **20**, 721-737 (2019).
646 2. Pope, B.D. *et al.* Topologically associating domains are stable units of replication-timing
647 regulation. *Nature* **515**, 402-5 (2014).
648 3. Sima, J. *et al.* Identifying cis Elements for Spatiotemporal Control of Mammalian DNA
649 Replication. *Cell* **176**, 816-830 e18 (2019).
650 4. Rivera-Mulia, J.C. *et al.* DNA replication timing alterations identify common markers
651 between distinct progeroid diseases. *Proc Natl Acad Sci U S A* **114**, E10972-e10980 (2017).
652 5. Platt, E.J., Smith, L. & Thayer, M.J. L1 retrotransposon antisense RNA within ASAR lncRNAs
653 controls chromosome-wide replication timing. *J Cell Biol* **217**, 541-553 (2018).
654 6. Sasaki, T. *et al.* Stability of patient-specific features of altered DNA replication timing in
655 xenografts of primary human acute lymphoblastic leukemia. *Exp Hematol* **51**, 71-82.e3
656 (2017).
657 7. Klein, K.N. *et al.* Replication timing maintains the global epigenetic state in human cells.
658 *Science* **372**, 371-378 (2021).
659 8. Du, Q. *et al.* Replication timing and epigenome remodelling are associated with the nature of
660 chromosomal rearrangements in cancer. *Nat Commun* **10**, 416 (2019).
661 9. Janoueix-Lerosey, I. *et al.* Preferential occurrence of chromosome breakpoints within early
662 replicating regions in neuroblastoma. *Cell Cycle* **4**, 1842-6 (2005).
663 10. Shugay, M., Ortiz de Mendibil, I., Vizmanos, J.L. & Novo, F.J. Genomic hallmarks of genes
664 involved in chromosomal translocations in hematological cancer. *PLoS Comput Biol* **8**,
665 e1002797 (2012).
666 11. Sima, J. & Gilbert, D.M. Complex correlations: replication timing and mutational landscapes
667 during cancer and genome evolution. *Curr Opin Genet Dev* **25**, 93-100 (2014).
668 12. Blumenfeld, B., Ben-Zimra, M. & Simon, I. Perturbations in the Replication Program
669 Contribute to Genomic Instability in Cancer. *International Journal of Molecular Sciences* **18**,
670 1138 (2017).
671 13. Peycheva, M. *et al.* DNA replication timing directly regulates the frequency of oncogenic
672 chromosomal translocations. *Science* **377**, eabj5502 (2022).
673 14. Stamatoyannopoulos, J.A. *et al.* Human mutation rate associated with DNA replication
674 timing. *Nature Genetics* **41**, 393-395 (2009).
675 15. Agier, N. & Fischer, G. The Mutational Profile of the Yeast Genome Is Shaped by Replication.
676 *Molecular Biology and Evolution* **29**, 905-913 (2011).
677 16. Weber, C.C., Pink, C.J. & Hurst, L.D. Late-Replicating Domains Have Higher Divergence and
678 Diversity in Drosophila melanogaster. *Molecular Biology and Evolution* **29**, 873-882 (2011).
679 17. Lang, G.I. & Murray, A.W. Mutation Rates across Budding Yeast Chromosome VI Are
680 Correlated with Replication Timing. *Genome Biology and Evolution* **3**, 799-811 (2011).
681 18. Blin, M. *et al.* Transcription-dependent regulation of replication dynamics modulates
682 genome stability. *Nat Struct Mol Biol* **26**, 58-66 (2019).
683 19. Brison, O. *et al.* Transcription-mediated organization of the replication initiation program
684 across large genes sets common fragile sites genome-wide. *Nature Communications* **10**,
685 5693 (2019).
686 20. Letessier, A. *et al.* Cell-type-specific replication initiation programs set fragility of the FRA3B
687 fragile site. *Nature* **470**, 120-123 (2011).
688 21. Bleichert, F. Mechanisms of replication origin licensing: a structural perspective. *Curr Opin*
689 *Struct Biol* **59**, 195-204 (2019).
690 22. Bochman, M.L. & Schwacha, A. The Mcm complex: unwinding the mechanism of a replicative
691 helicase. *Microbiol Mol Biol Rev* **73**, 652-83 (2009).

692 23. Burgers, P.M.J. & Kunkel, T.A. Eukaryotic DNA Replication Fork. *Annu Rev Biochem* **86**, 417-
693 438 (2017).
694 24. Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. DNA replication origin activation in
695 space and time. *Nat Rev Mol Cell Biol* **16**, 360-74 (2015).
696 25. Hu, Y. & Stillman, B. Origins of DNA replication in eukaryotes. *Molecular Cell* **83**, 352-372
697 (2023).
698 26. Hyrien, O., Marheineke, K. & Goldar, A. Paradoxes of eukaryotic DNA replication: MCM
699 proteins and the random completion problem. *Bioessays* **25**, 116-25 (2003).
700 27. Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C. & Zegerman, P. Titration of four replication
701 factors is essential for the Xenopus laevis midblastula transition. *Science* **341**, 893-6 (2013).
702 28. Tanaka, S., Nakato, R., Katou, Y., Shirahige, K. & Araki, H. Origin association of Sld3, Sld7, and
703 Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol* **21**, 2055-63
704 (2011).
705 29. Mantiero, D., Mackenzie, A., Donaldson, A. & Zegerman, P. Limiting replication initiation
706 factors execute the temporal programme of origin firing in budding yeast. *Embo j* **30**, 4805-
707 14 (2011).
708 30. Douglas, Max E. & Diffley, John F.X. Replication Timing: The Early Bird Catches the Worm.
709 *Current Biology* **22**, R81-R82 (2012).
710 31. Cornacchia, D. *et al.* Mouse Rif1 is a key regulator of the replication-timing programme in
711 mammalian cells. *EMBO J* **31**, 3678-90 (2012).
712 32. Yamazaki, S. *et al.* Rif1 regulates the replication timing domains on the human genome.
713 *EMBO J* **31**, 3667-77 (2012).
714 33. Hayano, M. *et al.* Rif1 is a global regulator of timing of replication origin firing in fission
715 yeast. *Genes Dev* **26**, 137-50 (2012).
716 34. Hiraga, S. *et al.* Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse
717 Cdc7-mediated phosphorylation of the MCM complex. *Genes Dev* **28**, 372-83 (2014).
718 35. Gnan, S. *et al.* Nuclear organisation and replication timing are coupled through RIF1-PP1
719 interaction. *Nat Commun* **12**, 2910 (2021).
720 36. Foti, R. *et al.* Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing
721 Program. *Mol Cell* **61**, 260-73 (2016).
722 37. Delgado-Benito, V. *et al.* The Chromatin Reader ZMYND8 Regulates Igh Enhancers to
723 Promote Immunoglobulin Class Switch Recombination. *Mol Cell* **72**, 636-649.e8 (2018).
724 38. Balasubramanian, S. *et al.* Protection of nascent DNA at stalled replication forks is mediated
725 by phosphorylation of RIF1 intrinsically disordered region. *Elife* **11**(2022).
726 39. Wiedemann, E.M., Peycheva, M. & Pavri, R. DNA Replication Origins in Immunoglobulin
727 Switch Regions Regulate Class Switch Recombination in an R-Loop-Dependent Manner. *Cell*
728 *Rep* **17**, 2927-2942 (2016).
729 40. Marchal, C. *et al.* Genome-wide analysis of replication timing by next-generation sequencing
730 with E/L Repli-seq. *Nat Protoc* **13**, 819-839 (2018).
731 41. Enervald, E. *et al.* RIF1 and KAP1 differentially regulate the choice of inactive versus active X
732 chromosomes. *The EMBO Journal* **40**, e105862 (2021).
733 42. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
734 RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
735 43. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of
736 chromatin looping. *Cell* **159**, 1665-80 (2014).
737 44. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
738 45. Lund, E., Oldenburg, A.R. & Collas, P. Enriched domain detector: a program for detection of
739 wide genomic enrichment domains robust against local variations. *Nucleic Acids Res* **42**, e92
740 (2014).
741 46. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how
742 promoters direct initiation and pausing. *Science* **339**, 950-3 (2013).

743 47. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of
744 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding
745 proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
746 48. Lombrana, R. *et al.* High-resolution analysis of DNA synthesis start sites and nucleosome
747 architecture at efficient mammalian replication origins. *EMBO J* **32**, 2631-44 (2013).
748 49. Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their
749 organization in specific but flexible sites defined by conserved features. *Genome Res* **21**,
750 1438-49 (2011).
751 50. Ge, X.Q., Jackson, D.A. & Blow, J.J. Dormant origins licensed by excess Mcm2-7 are required
752 for human cells to survive replicative stress. *Genes Dev* **21**, 3331-41 (2007).
753 51. Woodward, A.M. *et al.* Excess Mcm2-7 license dormant origins of replication that can be
754 used under conditions of replicative stress. *J Cell Biol* **173**, 673-83 (2006).
755 52. Ibarra, A., Schwob, E. & Mendez, J. Excess MCM proteins protect human cells from
756 replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A* **105**,
757 8956-61 (2008).
758 53. Gros, J. *et al.* Post-licensing Specification of Eukaryotic Replication Origins by Facilitated
759 Mcm2-7 Sliding along DNA. *Mol Cell* **60**, 797-807 (2015).
760 54. Scherr, M.J., Wahab, S.A., Remus, D. & Duderstadt, K.E. Mobile origin-licensing factors
761 confer resistance to conflicts with RNA polymerase. *Cell Reports* **38**, 110531 (2022).
762 55. Kirstein, N. *et al.* Human ORC/MCM density is low in active genes and correlates with
763 replication time but does not delimit initiation zones. *eLife* **10**, e62161 (2021).
764 56. Petryk, N. *et al.* Replication landscape of the human genome. *Nat Commun* **7**, 10208 (2016).
765 57. Hiraga, S.I. *et al.* Human RIF1 and protein phosphatase 1 stimulate DNA replication origin
766 licensing but suppress origin activation. *EMBO Rep* **18**, 403-419 (2017).
767 58. Dowling, M.R. *et al.* Stretched cell cycle model for proliferating lymphocytes. *Proceedings of*
768 *the National Academy of Sciences* **111**, 6377-6382 (2014).
769 59. Zhang, J., MacLennan, I.C.M., Liu, Y.-J. & Lane, P.J.L. Is rapid proliferation in B centroblasts
770 linked to somatic mutation in memory B cell clones? *Immunology Letters* **18**, 297-299 (1988).
771 60. Liu, Y.-J., Zhang, J., Lane, P.J.L., Chan, E.Y.-T. & Maclennan, I.C.M. Sites of specific B cell
772 activation in primary and secondary responses to T cell-dependent and T cell-independent
773 antigens. *European Journal of Immunology* **21**, 2951-2962 (1991).
774 61. Liu, M. *et al.* Two levels of protection for the B cell genome during somatic hypermutation.
775 *Nature* **451**, 841-5 (2008).
776 62. Alvarez-Prado, A.F. *et al.* A broad atlas of somatic hypermutation allows prediction of
777 activation-induced deaminase targets. *J Exp Med* **215**, 761-771 (2018).
778 63. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell
779 lymphomas. *Nature* **412**, 341-6 (2001).
780 64. Shen, H.M., Peters, A., Baron, B., Zhu, X. & Storb, U. Mutation of BCL-6 gene in normal B cells
781 by the process of somatic hypermutation of Ig genes. *Science* **280**, 1750-2 (1998).
782 65. Nussenzweig, A. & Nussenzweig, M.C. Origin of chromosomal translocations in lymphoid
783 cancer. *Cell* **141**, 27-38 (2010).
784 66. Casellas, R. *et al.* Mutations, kataegis and translocations in B cells: understanding AID
785 promiscuous activity. *Nat Rev Immunol* **16**, 164-76 (2016).
786 67. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat Rev Immunol*
787 **15**, 172-84 (2015).
788 68. Methot, S.P. & Di Noia, J.M. Molecular Mechanisms of Somatic Hypermutation and Class
789 Switch Recombination. *Adv Immunol* **133**, 37-87 (2017).
790 69. Zeman, M.K. & Cimprich, K.A. Causes and consequences of replication stress. *Nature Cell*
791 *Biology* **16**, 2-9 (2014).
792 70. Mahat, D.B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases
793 using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-76 (2016).

71. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature Biotechnology* **35**, 316-319 (2017).
72. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160-W165 (2016).
73. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
74. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
75. Foti, R. *et al.* Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Molecular Cell* **61**, 260-273 (2016).
76. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
77. Lund, E., Oldenburg, A.R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research* **42**, e92-e92 (2014).
78. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

Figure 1

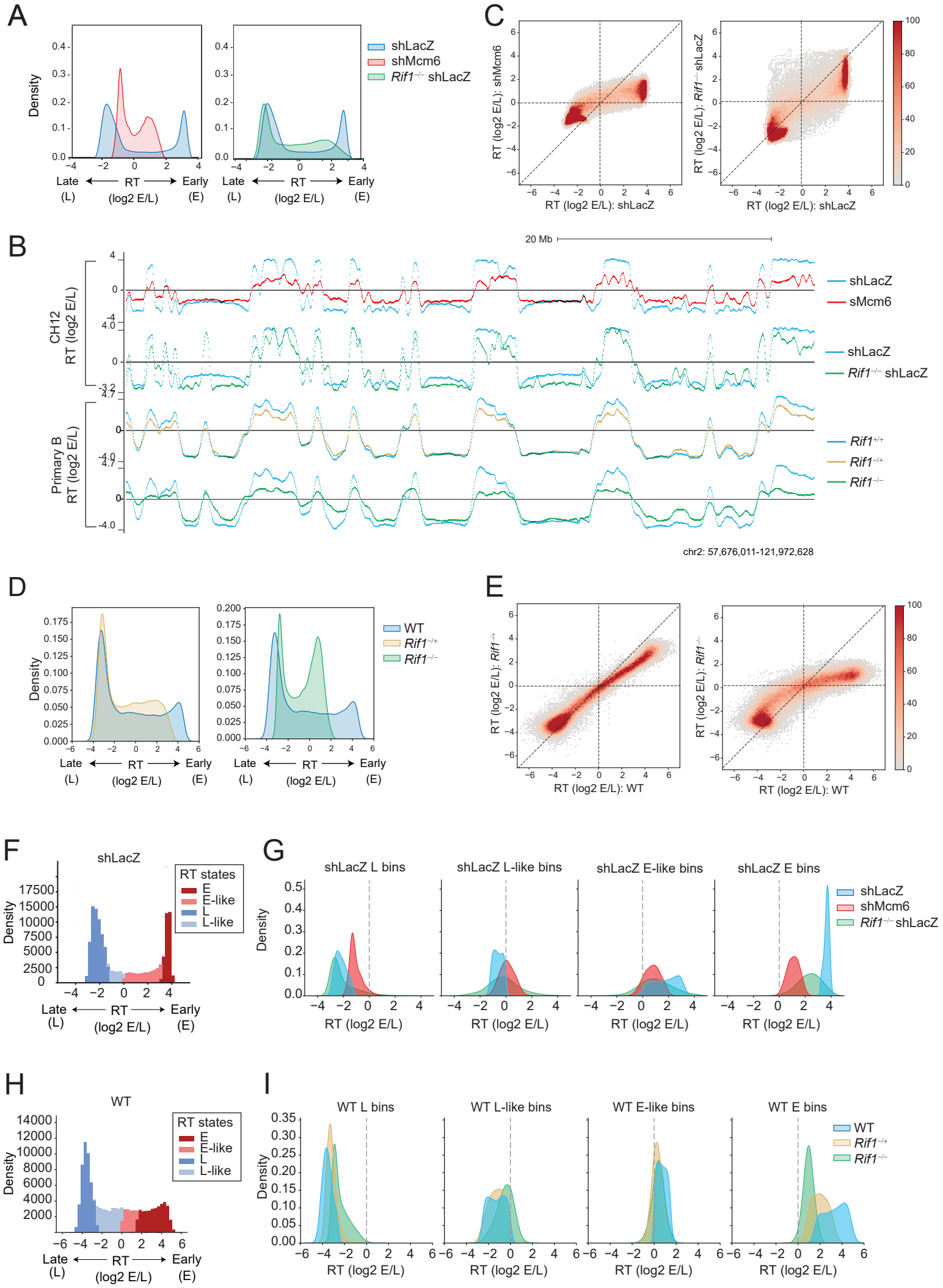

A
Density
shLacZ
shMcm6
Rif1−/− shLacZ
Late (L) ← RT (log2 E/L) → Early (E)
C
RT (log2 E/L): shMcm6
RT (log2 E/L): Rif1−/− shLacZ
RT (log2 E/L): shLacZ
B
20 Mb
CH12 RT (log2 E/L)
Primary B RT (log2 E/L)
shLacZ
sMcm6
Rif1−/− shLacZ
Rif1+/+
Rif1−/+
Rif1−/−
chr2: 57,676,011-121,972,628
D
WT
Rif1−/+
Rif1−/−
E
RT (log2 E/L): Rif1−/+
RT (log2 E/L): Rif1−/−
RT (log2 E/L): WT
F
shLacZ
RT states
E
E-like
L
L-like
G
shLacZ L bins
shLacZ L-like bins
shLacZ E-like bins
shLacZ E bins
RT (log2 E/L)
H
WT
I
WT L bins
WT L-like bins
WT E-like bins
WT E bins

Figure 1: RIF1 regulates early RT in activated B cells

A. RT histograms in 5 kb genomic bins from shLacZ (blue), shMcm6 (red) and *Rif1*$^{-/-}$ shLacZ (green) CH12 cells. RT is calculated as the ratio of the read densities of early (E) and late (L) Repli-seq fractions (log2 E/L).

B. UCSC genome browser view of RT for the indicated conditions in CH12 cells (top panel) and primary, activated splenic B cells (bottom panel). Positive and negative values correspond to early and late-replicating regions.

C. Comparison of RT values in 20 kb genomic bins between shLacZ and shMcm6 CH12 cells (left) and shLacZ and *Rif1*$^{-/-}$ shLacZ cells (right).

D. RT histograms as in (A) from WT (blue), *Rif1*$^{-/+}$ (ochre) and *Rif1*$^{-/-}$ (green) primary, activated splenic B cells.

E. Comparison of RT values in 20 kb genomic bins between WT, *Rif1*$^{-/+}$ and *Rif1*$^{-/-}$ in primary B cells.

F. Classification of shLacZ RT bins into four states using a Hidden Markov Model (HMM). Three states E, M (mid) and L were called, and the middle state was split into into E-like and L-like at the log2 (E/L) = 0 boundary.

G. RT histograms showing how the bins in each shLacZ RT state (blue) called in (D) change in shMcm6 (red) and *Rif1*$^{-/-}$ shLacZ (green) CH12 cells.

H. HMM analysis as in F but using WT RT values from primary B cells to call RT states.

I. RT histograms showing how bins in each WT RT state (blue) change in *Rif1*$^{-/+}$ (ochre) and *Rif1*$^{-/-}$ (green) primary B cells.

Figure 2

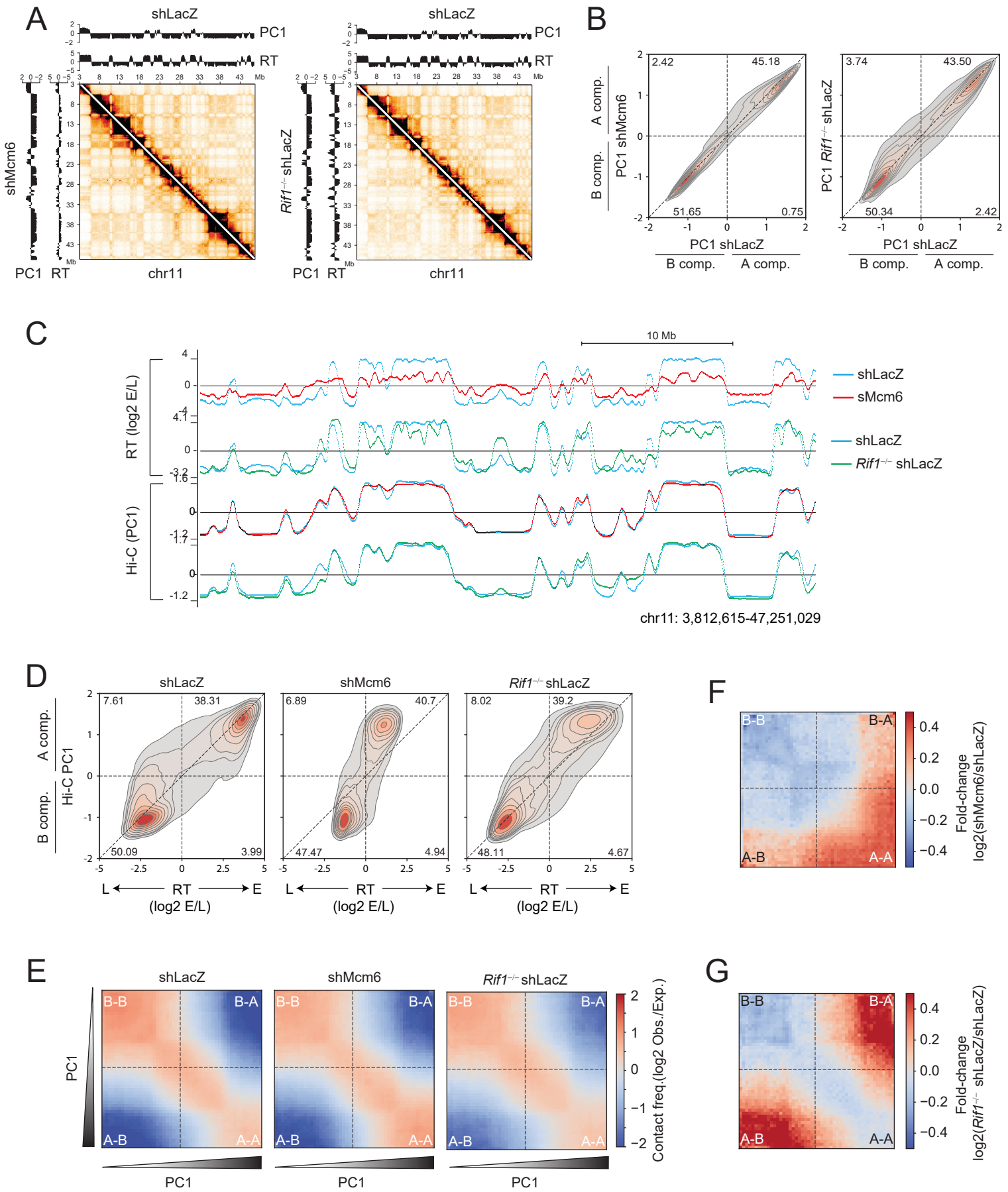

A
shLacZ
PC1
RT
shMcm6
Rif1−/− shLacZ
chr11
B
A comp.
B comp.
PC1 shMcm6
PC1 shLacZ
PC1 Rif1−/− shLacZ
2.42
45.18
51.65
0.75
3.74
43.50
50.34
2.42
C
10 Mb
RT (log2 E/L)
Hi-C (PC1)
shLacZ
sMcm6
Rif1−/− shLacZ
chr11: 3,812,615-47,251,029
D
shLacZ
shMcm6
Rif1−/− shLacZ
Hi-C PC1
7.61
38.31
50.09
3.99
6.89
40.7
47.47
4.94
8.02
39.2
48.11
4.67
L ← RT → E
(log2 E/L)
E
B-B
B-A
A-B
A-A
PC1
Contact freq.(log2 Obs./Exp.)
F
Fold-change
log2(shMcm6/shLacZ)
G
Fold-change
log2(Rif1−/− shLacZ/shLacZ)

Figure 2: RIF1 has a minor role in B cell genome architecture

A. Left: Hi-C contact matrix of chromosome 11 showing shLacZ contacts (above the diagonal) and shMcm6 contacts (below the diagonal). The PC1 compartment signals (Hi-C eigenvector eigenvalues) and RT (log2 E/L) tracks are shown above (for shLacZ) and on the left (for shMcm6) for each matrix. Right: Same as before but comparing shLacZ and *Rif1*$^{-/-}$ shLacZ cells.

B. Density-contour plot of PC1 compartment signals per 20 kb genomic bin. The left plot compares the PC1 signals in shLacZ and shMcm6 cells, and the right plot compares PC1 values between shLacZ and *Rif1*$^{-/-}$ shLacZ cells. The numbers within the plots are the percentage of bins in that quadrant.

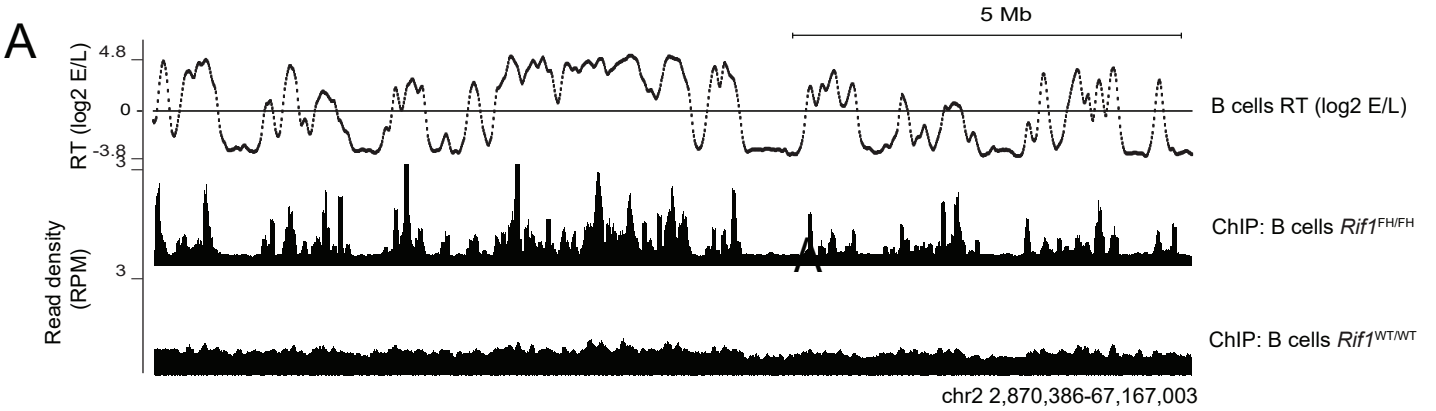
C. A representative UCSC genome browser view comparing Hi-C PC1 profiles with RT (log2 E/L) profiles in shLacZ, shMcm6 and *Rif1*$^{-/-}$ shLacZ cells.

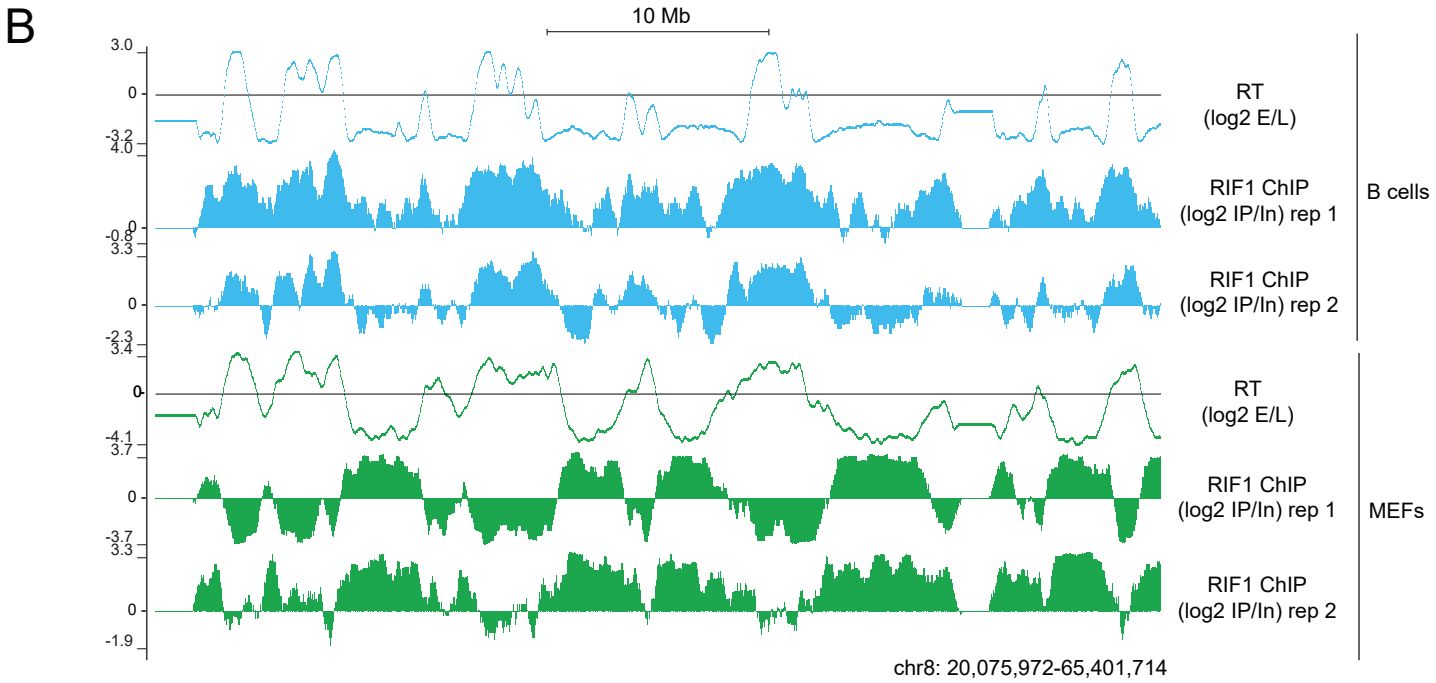
D. PC1 versus RT density-contour plots in 20 kb genomic bins to compare changes in compartmental identities (PC1) with RT in shLacZ, shMcm6 and *Rif1*$^{-/-}$ shLacZ CH12 cells.

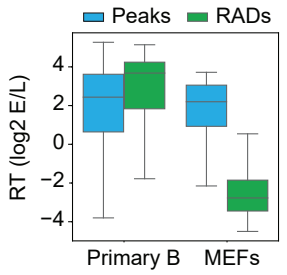
E. Saddle plot from shLacZ, shMcm6 and *Rif1*$^{-/-}$ shLacZ CH12 cells showing long-range (> 2Mb) intra-chromosomal contact enrichments between bins of varying compartment signal strength (PC1). The values were computed from 20 kb KR-normalized contact matrices.

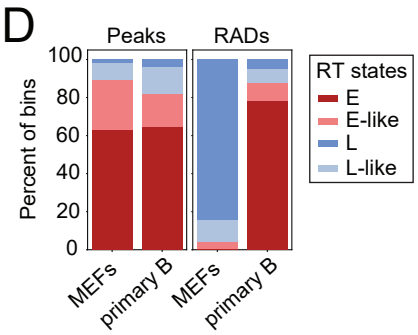
F-G Fold-change saddle plots highlighting the changes in compartmentalization in shMcm6 and *Rif1*$^{-/-}$ shLacZ cells relative to shLacZ cells.

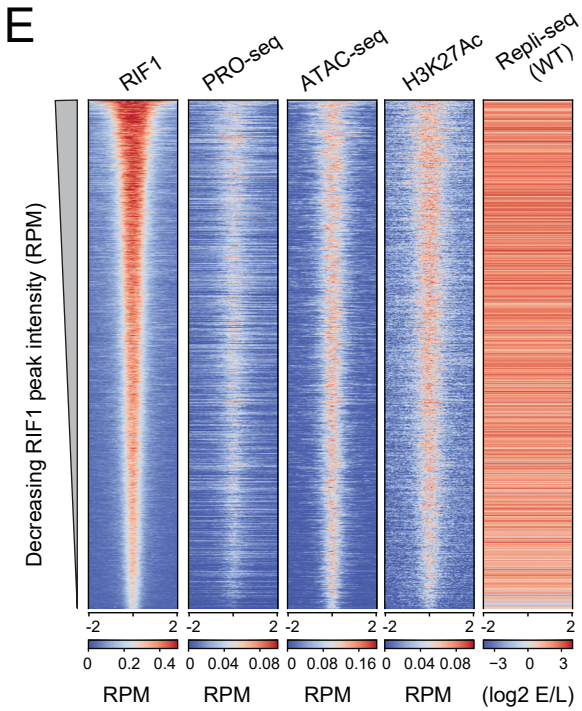
Figure 3

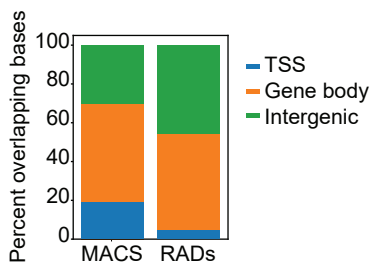

A
5 Mb
RT (log2 E/L)
4.8
0
-3.8
B cells RT (log2 E/L)
Read density (RPM)
3
ChIP: B cells Rif1FH/FH
3
ChIP: B cells Rif1WT/WT
chr2 2,870,386-67,167,003


B
10 Mb
RT (log2 E/L)
RIF1 ChIP (log2 IP/In) rep 1
RIF1 ChIP (log2 IP/In) rep 2
B cells
RT (log2 E/L)
RIF1 ChIP (log2 IP/In) rep 1
RIF1 ChIP (log2 IP/In) rep 2
MEFs
chr8: 20,075,972-65,401,714


C
Peaks
RADs
RT (log2 E/L)
Primary B
MEFs


D
Peaks
RADs
Percent of bins
RT states
E
E-like
L
L-like
MEFs
primary B
MEFs
primary B


E
RIF1
PRO-seq
ATAC-seq
H3K27Ac
Repli-seq (WT)
Decreasing RIF1 peak intensity (RPM)
RPM
RPM
RPM
RPM
(log2 E/L)


F
Percent overlapping bases
TSS
Gene body
Intergenic
MACS
RADs

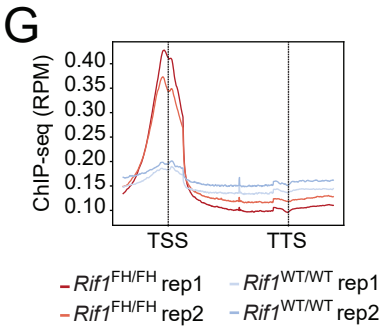

G
ChIP-seq (RPM)
TSS
TTS
Rif1FH/FH rep1
Rif1FH/FH rep2
Rif1WT/WT rep1
Rif1WT/WT rep2

Figure 3: RIF1 localizes predominantly to active chromatin in activated B cells

A. A representative UCSC genome browser view of RIF1 occupancy in *Rif1*$^{FH/FH}$ primary, activated splenic B cells. ChIP-seq was performed with an anti-HA antibody and *Rif1*$^{WT/WT}$ cells were used as a negative control. The RT track from WT, primary B cells provides a reference for early and late-replicating domains.

B. UCSC browser snapshot comparing ChIP enrichments of RIF1 in primary B and MEFs from *Rif1*$^{FH/FH}$ mice. To allow direct comparison, ChIP signal was normalized to input signal and the ratio (ChIP/Input) tracks are shown. RT tracks are from WT cells.

C. Box plots comparing RT values of RIF1 peaks and RADs in MEFs and primary B cells. 16,043 peaks and 289 RADs were identified in primary B cells and 862 peaks and 332 RADs were called in MEFs.

D. Peaks and RADs were called as in C above and classified into the four RT states described in Fig. 1D.

E. Heatmap analysis of the chromatin locale in a 4 kb window surrounding RIF1 peaks called in *Rif1*$^{FH/FH}$ cells. The heatmap is centered on the RIF1 peak summit and ordered by decreasing RIF1 read density (reads per million, RPM). All other genomic tracks shown (PRO-seq, ATAC-seq, H3K27Ac and Repli-seq) are from WT primary, activated splenic B cells.

F. Distribution of RIF1 peaks and RADs at TSSs, gene bodies and intergenic regions.

G. Metagene analysis showing average RIF1 occupancy patterns at genes in *Rif1*$^{FH/FH}$ and *Rif1*$^{WT/WT}$ cells from two replicates each (rep1 and rep2). The TSS and TTS are the transcription start site and transcription termination site.

Figure 4

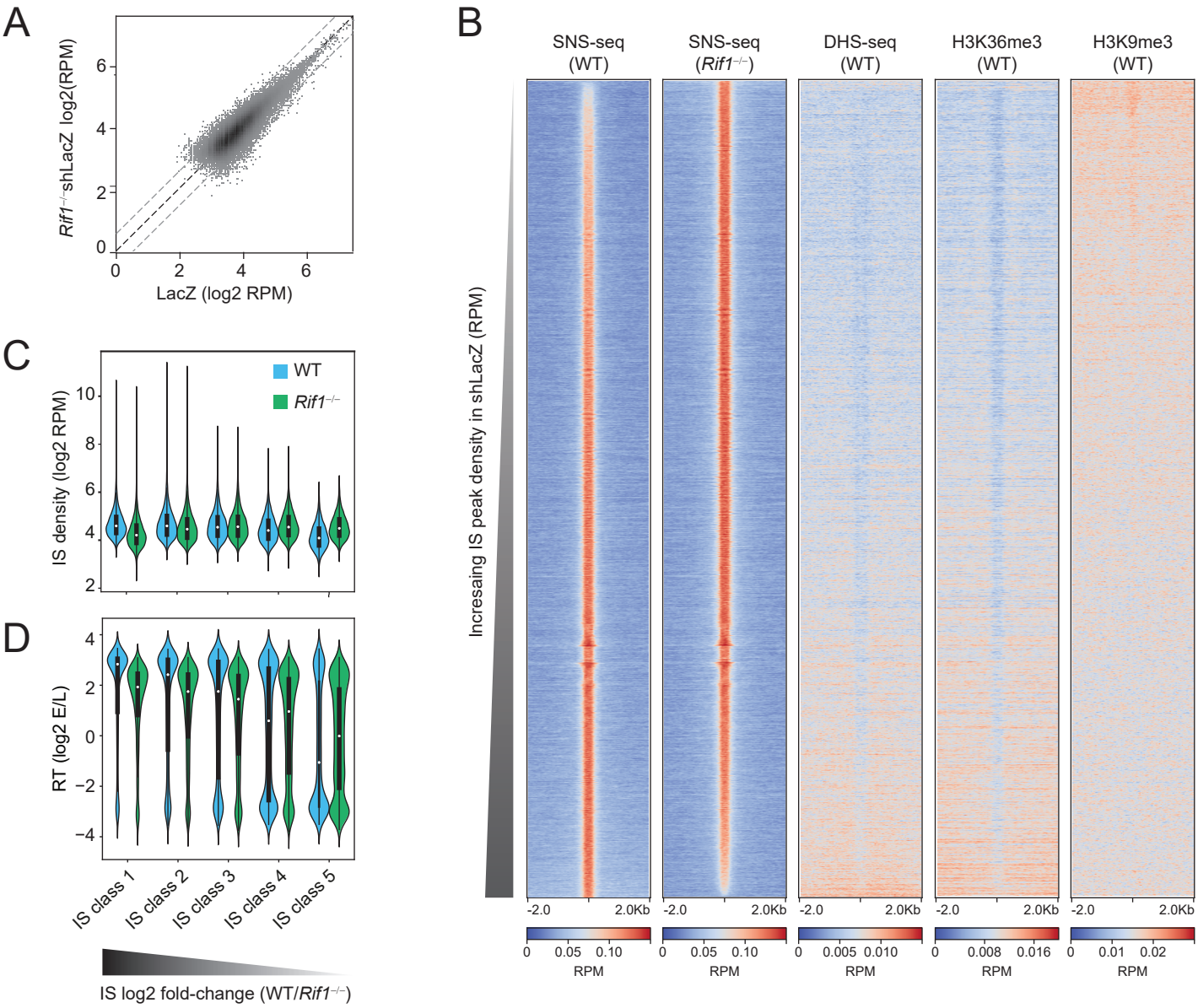


Figure 4: RIF1 regulates the activity of early origins of replication in B cells

A. Scatter plot of IS density in log2 RPM (reads per million) in WT and *Rif1*$^{-/-}$ cells identified from SNS-seq.

B. Heatmaps spanning a 4 kb region centered on the IS peak summit were generated to visualize the profiles of SNS-seq, DNase hypersensitivity (DHS)-seq and histone modifications (H3K36me3 and H3K9me3) around ISs. All heatmaps display read densities as RPM. Within each heatmap, the ordering is based on decreasing SNS-seq fold-change (*Rif1*$^{-/-}$/WT).

C. ISs were split into five equal classes (quintiles) based on their fold-change (log2 *Rif1*$^{-/-}$/WT) such that class 1 contained the most downregulated ISs and class 5 contained the most upregulated ISs, respectively, in *Rif1*$^{-/-}$ cells. The violin plots show the IS density within each class in WT and *Rif1*$^{-/-}$ cells.

D. ISs were classified as in C above. The violin plots show the RT values within each IS class.

Figure 5

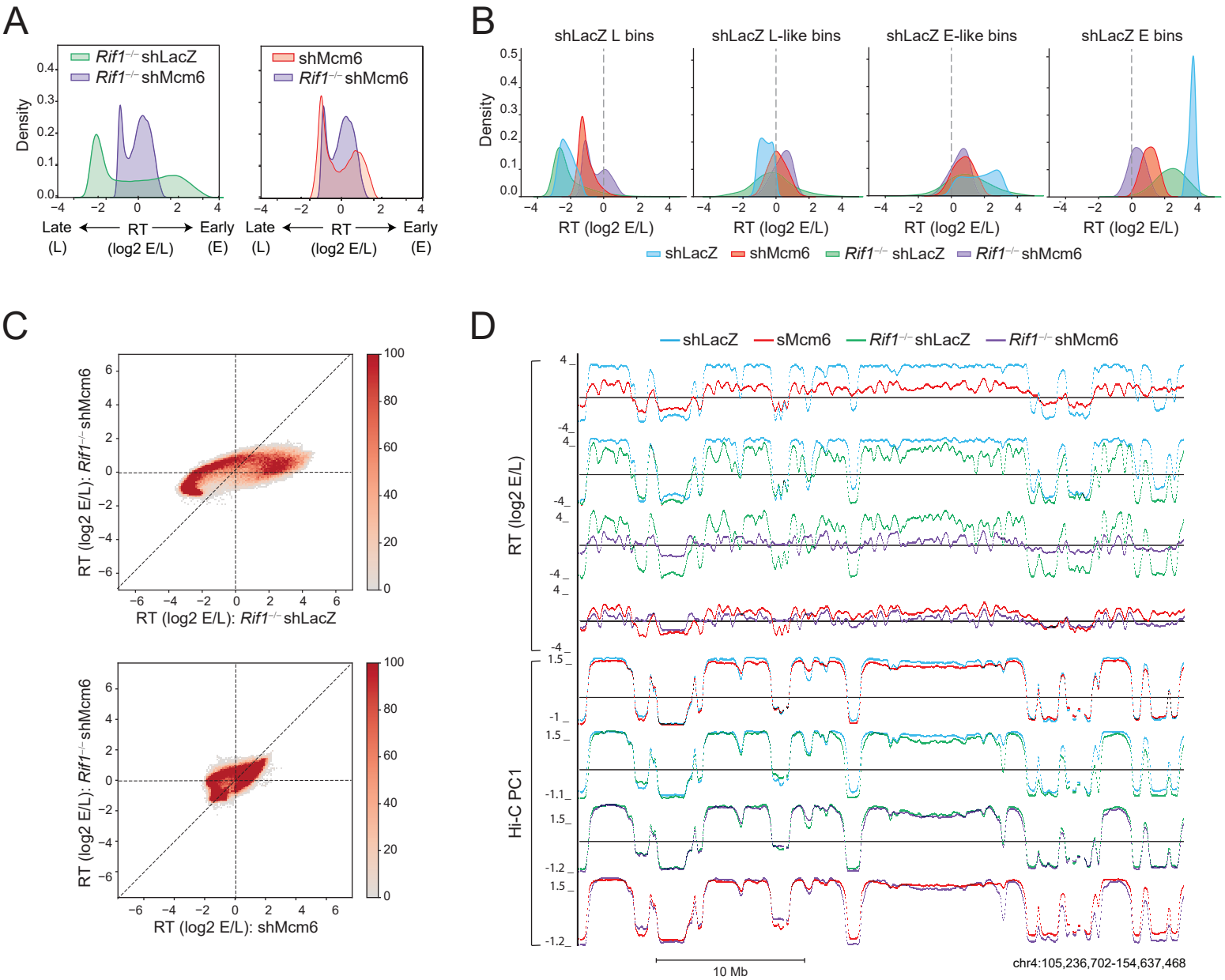


Figure 5: RIF1 and MCM complexes act in an additive manner to regulate early RT in B cells

A. RT histogram comparing *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 CH12 cells.

B. Identical to Fig. 1G showing how HMM-based RT states called in shLacZ cells change in the other three conditions.

C. Comparison of RT values in 20 kb genomic bins between *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells (top) and between shMcm6 and *Rif1*$^{-/-}$ shMcm6 cells (bottom).

D. A representative UCSC genome browser view comparing Hi-C PC1 profiles with RT (log2 E/L) profiles in shLacZ, shMcm6, *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 CH12 cells.

Figure 6

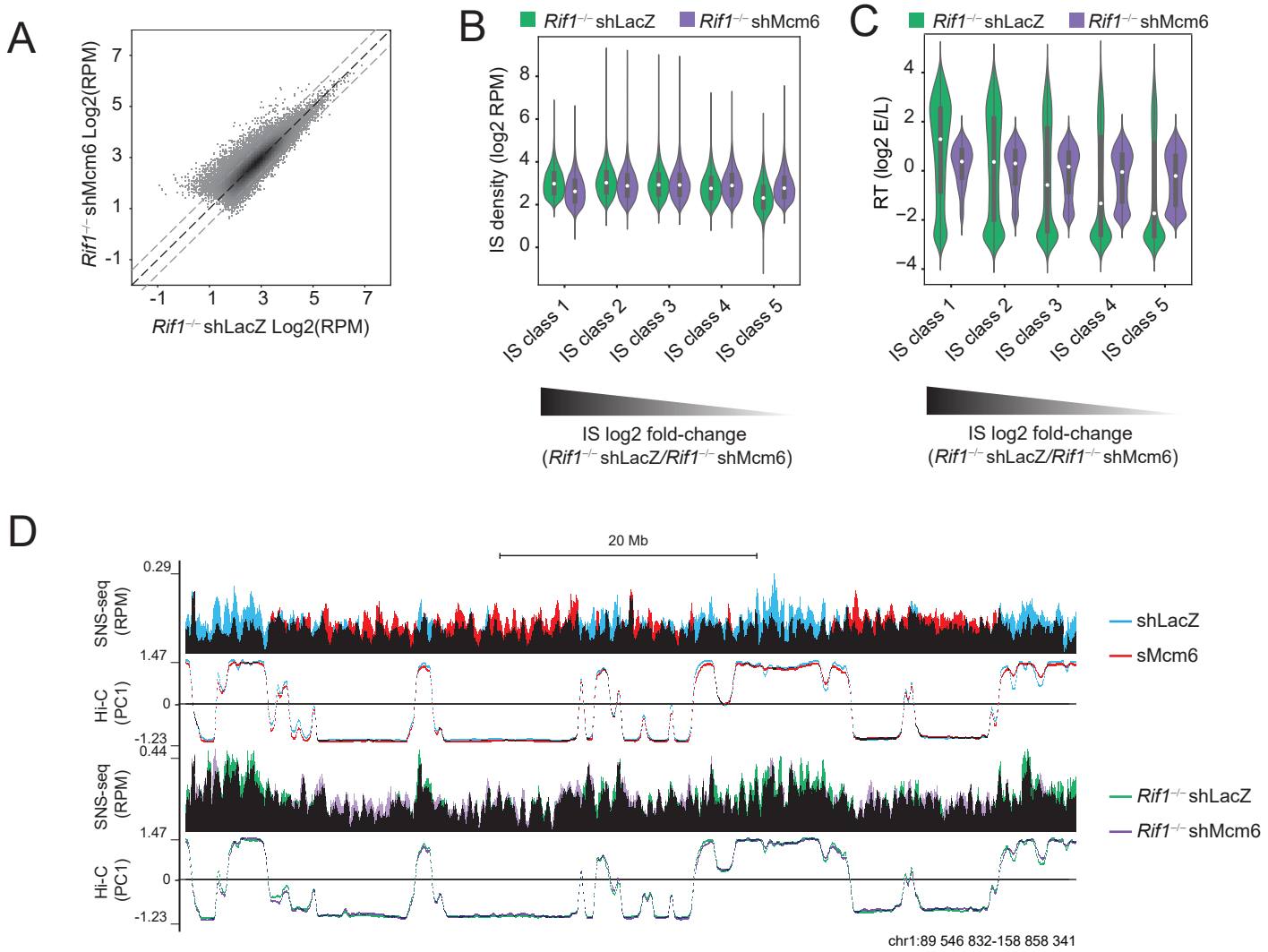


Figure 6: MCM depletion in *Rif1*$^{-/-}$ cells leads to a further deregulation of early origin activity

A. Scatter plot of IS density in log2 RPM in *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells identified from SNS-seq.

B. ISs were split into five equal classes (quintiles) based on their fold-change (log2 *Rif1*$^{-/-}$ shMcm6/ *Rif1*$^{-/-}$ shLacZ) such that class 1 contained the most downregulated ISs and class 5 contained the most upregulated ISs, respectively, in *Rif1*$^{-/-}$ shMcm6 cells. The violin plots show the IS density within each class in *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells.

C. ISs were classified into five classes as in C above. The violin plots show the RT values within each IS class.

D. Representative genomic snapshot of SNS-seq and HiC PC1 profiles. Data from shLacZ (blue) and shMcm6 (red) are overlaid with black being the overlap between them. *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 tracks are in green and purple, respectively, with black indicating the overlap between them.

Figure 7

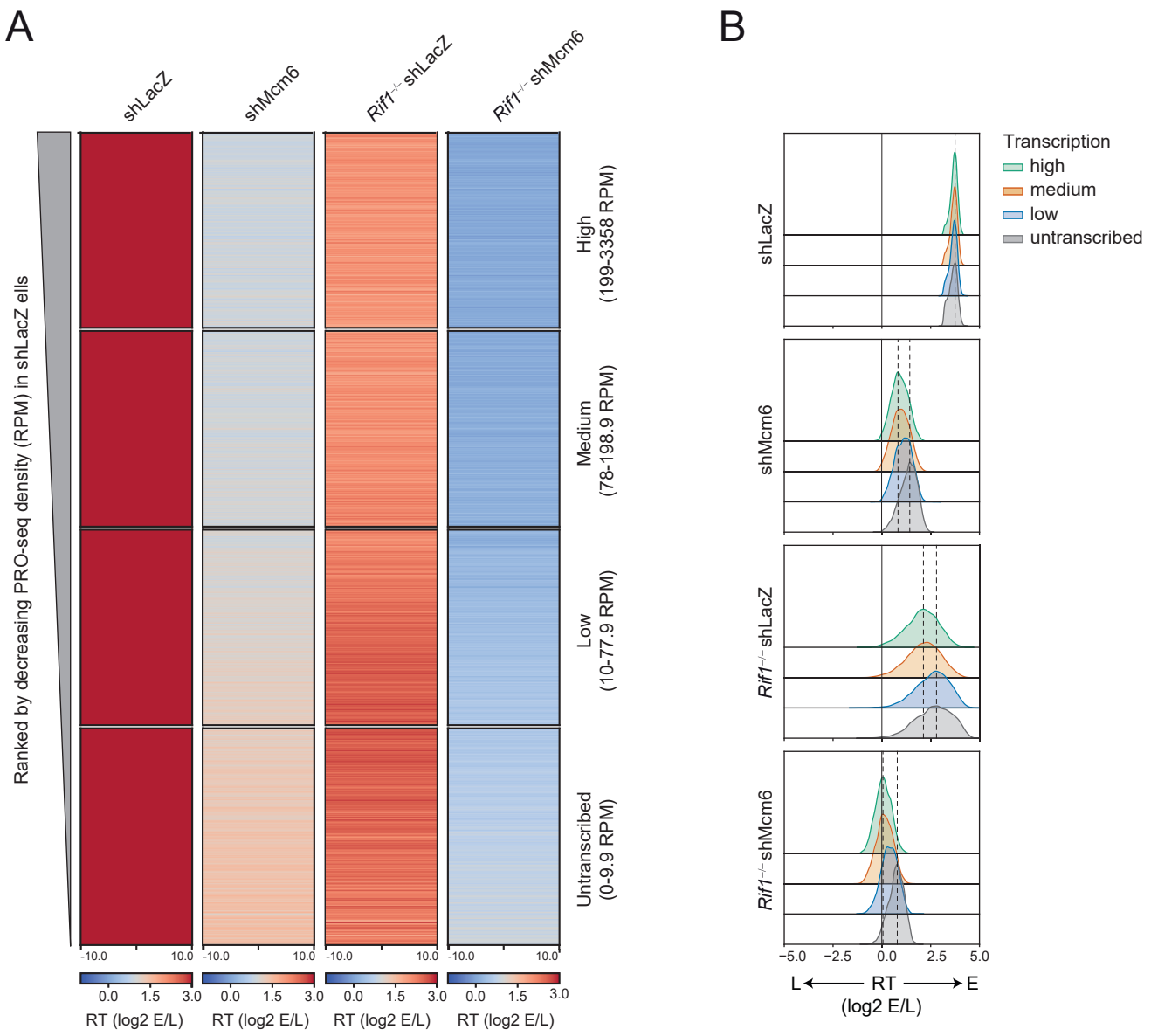


Figure 7: Within early-replicating domains, the RT of highly transcribed regions is more sensitive to reduction in MCM proteins and loss of RIF1 than poorly transcribed regions

A. Heatmap of RT values in 20 kb E bins from shLacZ cells (from Fig. 1G) that were split into four groups based on PRO-seq RPM. The range of PRO-seq RPMs in each group is indicated on the left. All heatmaps are ranked from highest to lowest PRO-seq density such that the entire set of heatmaps is effectively ranked from highest to lowest PRO-seq density for all shLacZ E bins.

B. Density distribution of the RT values in the four transcription-based groups from A above. The dashed black lines indicate the modes of the distributions in the High (green) and Untranscribed (grey) classes.

Figure S1

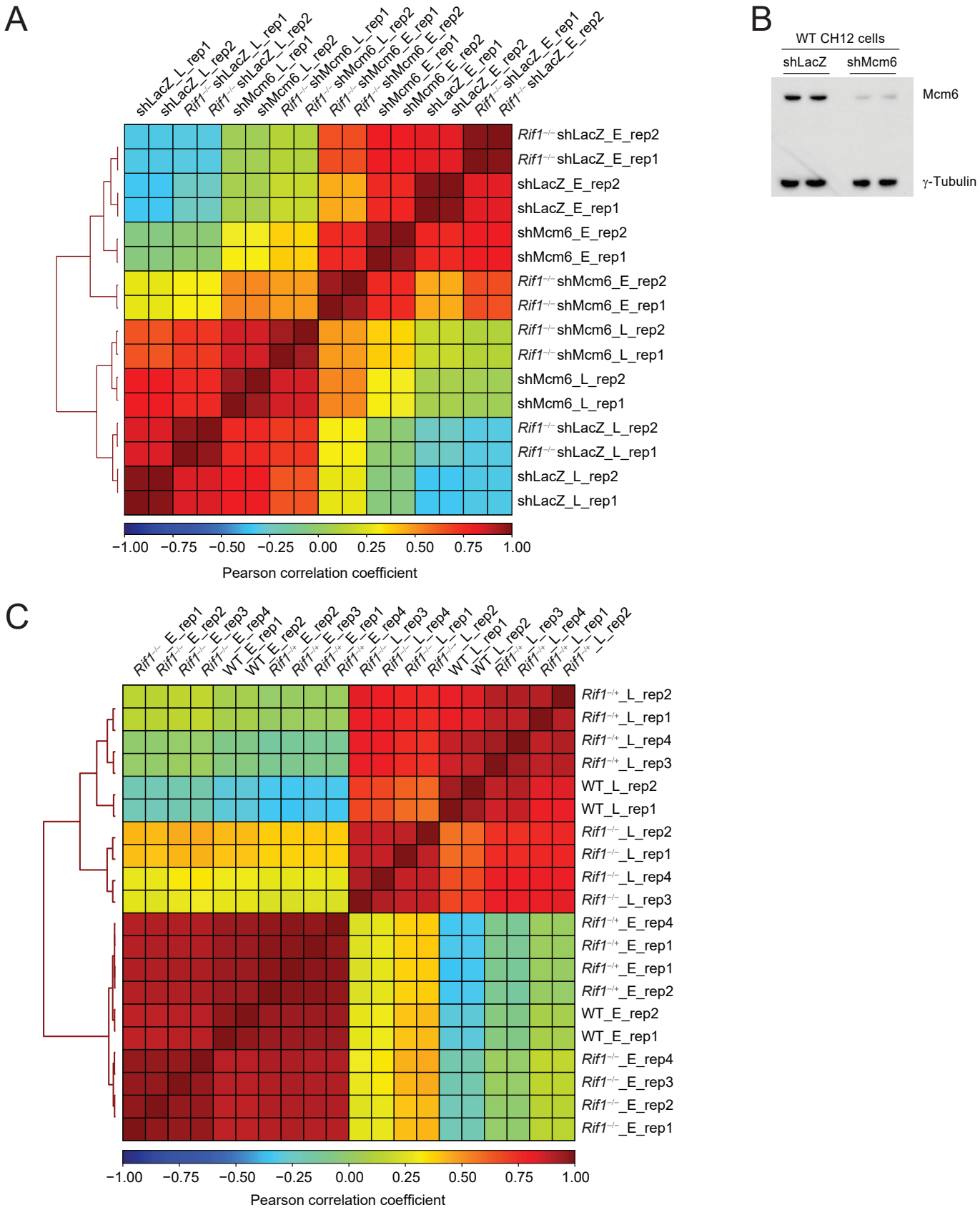


Figure S1: Analysis of Repli-seq replicates

A. Pearson correlation matrix of Repli-seq data from two replicates (rep1 and rep2) of shLacZ, shMcm6, *Rif1*$^{-/-}$shLacZ and *Rif1*$^{-/-}$ shMcm6 CH12 cells. E and L indicate the early and late fractions, respectively. The color gradient (bottom bar) corresponds to the Pearson correlation coefficient

B. Western blot for Mcm6 protein levels from nuclear extracts obtained from shLacZ and shMcm6 cells. Data from two replicates are shown with γ-Tubulin serves as the loading control. In all cases, 20 μg of extract was loaded.

C. Repli-seq correlation analysis as in A from WT (two replicates), *Rif1*$^{-/-}$(four replicates) and *Rif1*$^{-/-}$ (four replicates) primary, activated splenic B cells.

Figure S2

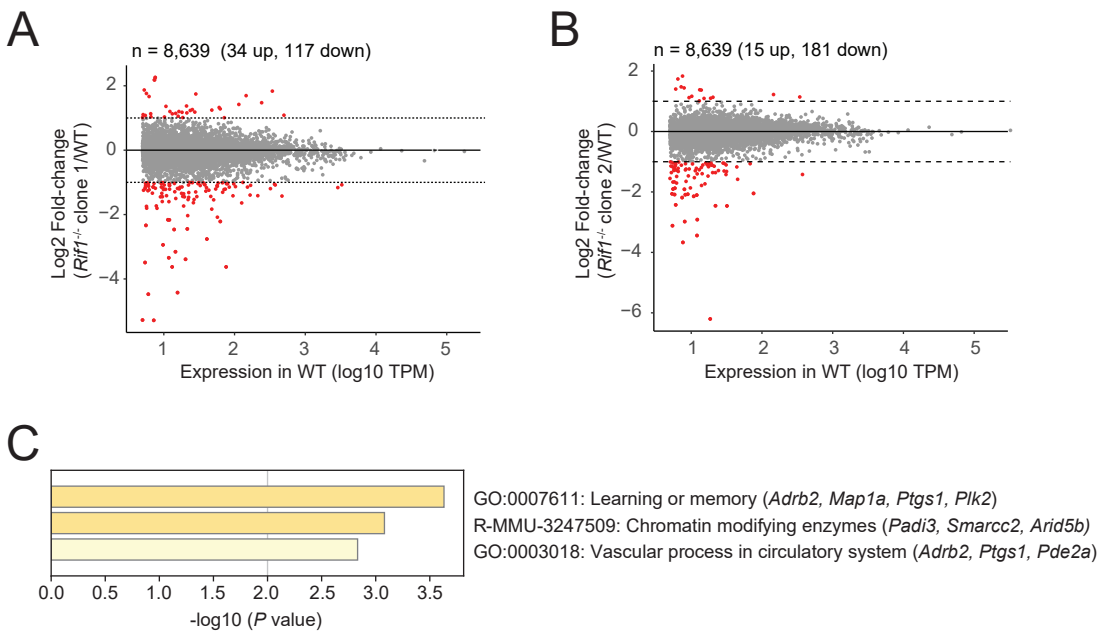


Figure S2: RNA-seq analysis from Rif1$^{-/-}$ CH12 cells

A. RNA-seq analysis of *Rif1*$^{-/-}$ clone 1 compared to WT cells. The data are shown as a scatter plot of RNA-seq read densities in transcripts per million (TPM) versus fold-change (FC; *Rif1*$^{-/-}$ /WT) for 8,639 expressed genes (defined as TPM > 5 in WT cells). The dotted lines mark log2 FC 1 or -1 corresponding to 2-fold upregulated and 2-fold downregulated genes, respectively. See also Table S1.

B. As in A but for *Rif1*$^{-/-}$ clone 2 compared to WT cells.

C. GO term analysis of the downregulated genes common to *Rif1*$^{-/-}$ clones 1 and 2 (n = 22). There were no enrichments for the common upregulated genes (n = 3). See also Table S1.

Figure S3

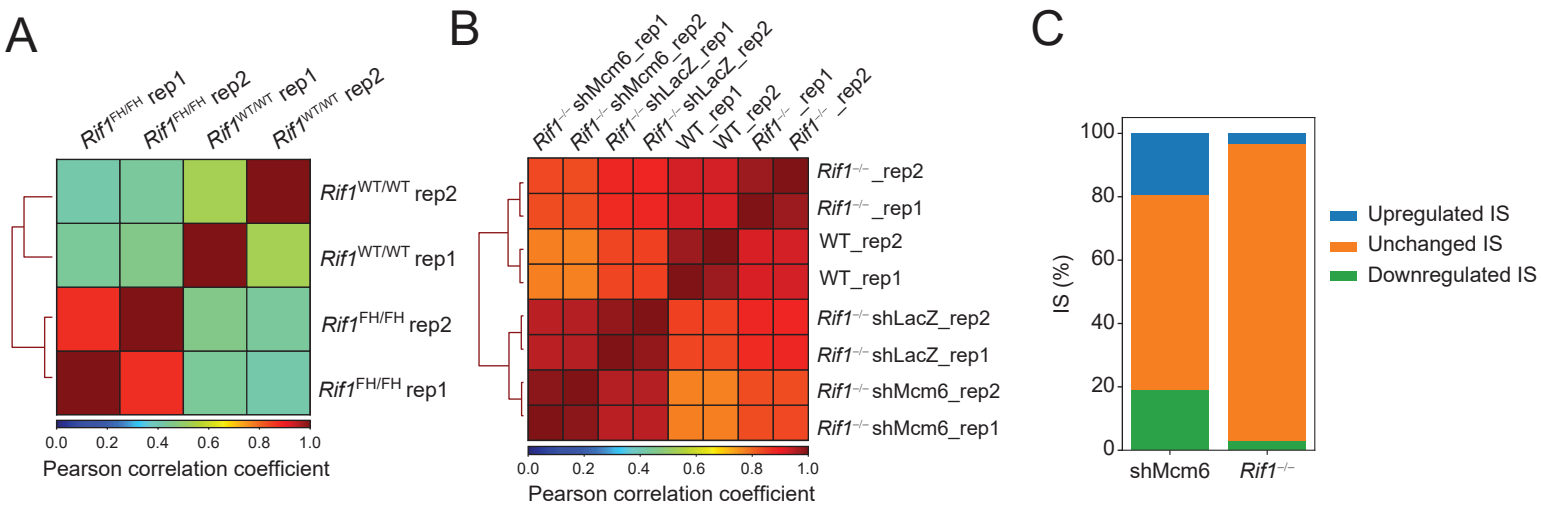


Figure S3: Analysis of ChIP-seq and SNS-seq replicates

A. Pearson correlation analysis of RIF1 ChIP-seq datasets. Two replicates from *Rif1*$^{FH/FH}$ and *Rif1*$^{WT/WT}$ primary B cells were analyzed.

A. Pearson correlation analysis of SNS-seq data. Two replicates from each indicated condition were analyzed.

B. Bar plots showing the percent of ISs upregulated (1.5-fold), downregulated (1.5-fold) or unchanged in shMcm6 (relative to shLacZ) or *Rif1*$^{-/-}$ (relative to WT) cells. The shMcm6 data is from our previous study (ref. 13).

Figure S4

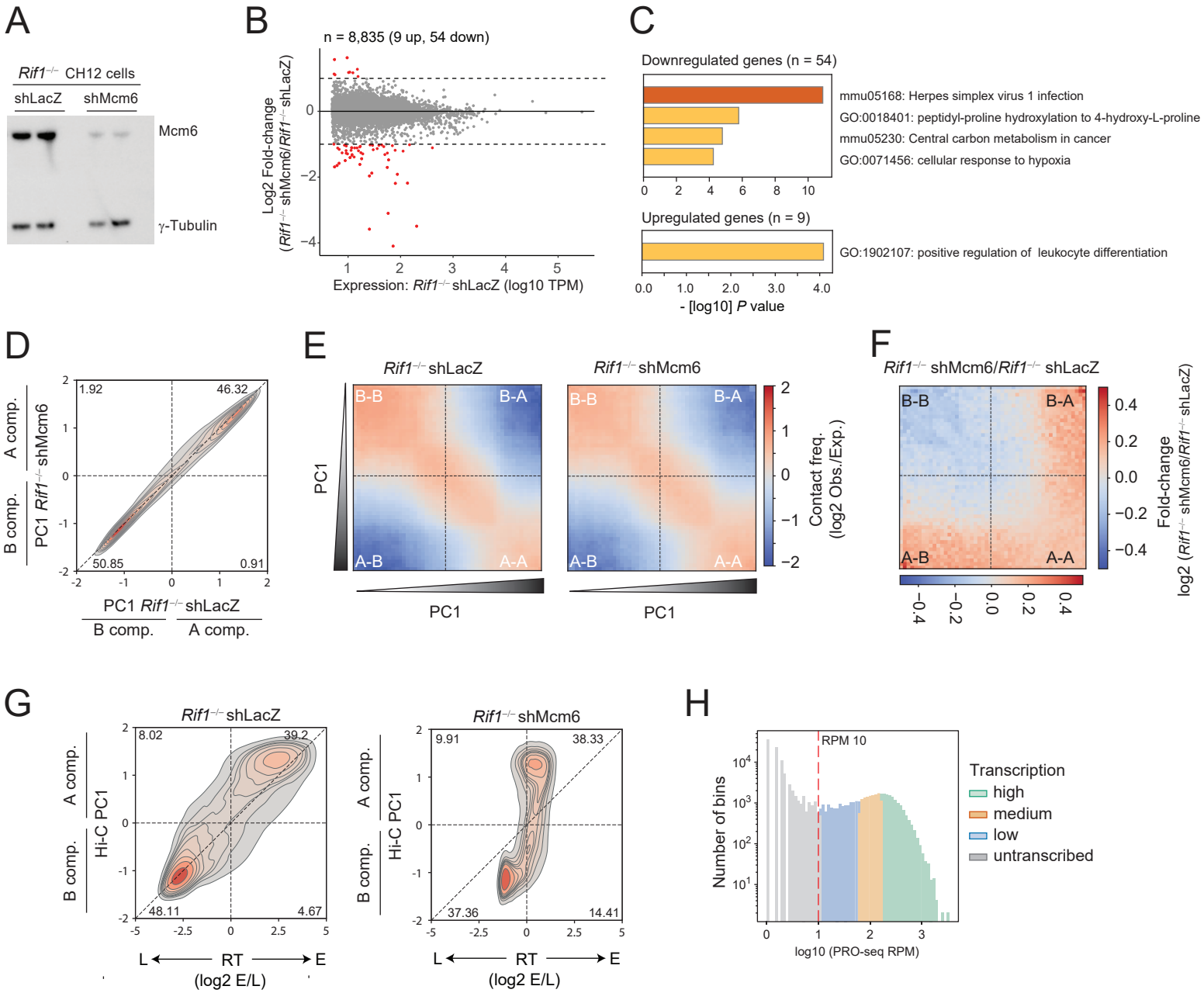


Figure S4: Gene expression and genome compartmentalization analyses in *Rif1*$^{-/-}$ shMcm6 cells

A. Western blot for Mcm6 protein levels from nuclear extracts obtained from two replicates of *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells. In all cases, 20 μg of extract was loaded. γ-Tubulin serves as the loading control.

B. Scatter plot of RNA-seq read densities in TPM versus fold-change (*Rif1*$^{-/-}$ shMcm6/Rif1*Rif1*$^{-/-}$ shLacZ) at 8,835 expressed genes (defined as TPM > 5 in *Rif1*$^{-/-}$ shLacZ cells). The dotted lines mark log2 FC 1 and -1 corresponding to 2-fold upregulated and 2-fold downregulated genes, respectively.

C. GO term analysis of the 54 downregulated and 9 upregulated genes from B.

D. Density-contour plot comparing PC1 compartment signals in 20 kb genomic bins from *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells.

E. Saddle plots comparing compartmental interactions between *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells.

F. Fold-change saddle plots based on data from E.

G. PC1 versus RT density-contour plots in 20 kb genomic bins comparing the changes in either feature in *Rif1*$^{-/-}$ shLacZ and *Rif1*$^{-/-}$ shMcm6 cells.

H. Histogram showing the distribution of PRO-seq densities (RPM) in 20 kb genomic bins from shLacZ cells. High, Medium, Low and Untranscribed groups are color-coded to match the analysis in Fig. 7B. The dotted red line indicates RPM 10 which marks the boundary between transcribed and untranscribed regions in our analyses.