A *Mycobacterium tuberculosis* secreted virulence factor disrupts host snRNP biogenesis

Komal Chauhan[1], Nishat Passricha[1], Yogita Kapoor[2], Dipanwita Datta[1], Naresh Arora[1], Mrinmoy Das[1], Kavya Rao[1], Lakshyaveer Singh[1], Ravi Dutt[1], Raman Deep Sharma[1], Binayak Sarkar[3], Basanti Malakkar[3], Prince Saini[4], Lalita Mehra[5], Prasenjit Das[5], Vineet Ahuja[6], Amit Singhal[7,8], Vinay Nandicoori[2, 3] and Dhiraj Kumar[1]*

[1]Cellular Immunology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India
[2]Centre for Cellular and Molecular Biology, Hyderabad, India
[3]National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi-110067, India
[4]Department of Biological Sciences, Indian Institute of Science Education and Research Mohali, Punjab 140306, India
[5]Department of Pathology, All India Institute of Medical Sciences, New Delhi-110012, India
[6]Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi-110012, India
[7]A*STAR Infectious Diseases Labs (A*STAR ID labs), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
[8]Singapore Immunology Network (SIgN), A*STAR, Singapore, Singapore.

Running title: Regulation of host RNA splicing by Mtb

*Correspondence- dhiraj@icgeb.res.in

Abstract

Alternative splicing is a fundamental eukaryotic process that fine-tunes cellular responses post-transcriptionally. *Mycobacterium tuberculosis* (Mtb) infection of human macrophages triggers extensive dysregulation of host alternative splicing (AS) to alter the host responses. However, the mechanism behind this remains unknown. Here, we present an unexpected mechanism exploited by the Mtb to achieve the same. We identify several bacterial secretory proteins that interact with the host splicing factors (SFs) and alter select RNA splicing events *in vitro*, *in cell* and during *ex vivo* infections. We show bacterial proteins disrupt AS by impairing Small nuclear ribonucleoprotein (snRNP) biogenesis, primarily by interacting with and sequestering U5snRNA and SNRPF, key spliceosome constituents. These interactions are driven, in part, by one of the identified Mtb proteins, Hsr1, whose access to the host cytosol is ESX1-dependent. The Mtb*Δhsr1* is unable to interact and sequester U5snRNA or SNRPF and fails to alter host AS. Consequently, Mtb*Δhsr1* shows compromised survival and pathogenesis in macrophages and mice. Lungs of infected mice show distinct SNRPF staining, which is hsr1 dependent. Similar distinctive SNRPF staining is also noted in human tuberculous granulomas. Moreover, monocytes from tuberculosis patients show a substantial splicing dysregulation indicating splicing defects as a key determinant behind systemic tuberculosis pathology. We propose mycobacterial exploitation of the evolutionarily conserved mechanism of snRNP biogenesis, aided its adaptation for human pathogenesis and provides a unique therapeutic opportunity.

Introduction

Tuberculosis (TB) is one of the leading causes of human mortality due to infectious diseases.[1] *Mycobacterium tuberculosis* (Mtb), the causative organism, is known to alter a variety of host cellular processes that eventually help establish the infection and cause disease. Thus, cellular mechanisms like signaling, autophagy, apoptosis, inflammation etc., have been extensively explored for potential host-directed therapy (HDT) against tuberculosis [2-5]. Mtb infection also alters host gene expression, which determines the long-term host responses against infection. In fact, the pathogen regulates host gene expression at multiple tiers, including epigenetic, transcriptional and post-transcriptional stages [6-9]. Previously, reports from our group and other showed that infection of human macrophages with Mtb triggers a massive alteration in the pattern of host RNA splicing[10,11]. Altered RNA splicing can impact the target proteins function, localization, stability and stoichimetry[12]. We reported increased instances of transcripts which either do not get translated or are substrate of a process called nonsense-mediated decay[13], were upregulated as a result of alternative splicing in the infected macrophages[8,11,13]. Curiously, genes typically belonging to functional classes that are activated upon infection are impacted by AS[8,9,12]. This targeted blunting of host response by bacterial infection favours the bacteria[11]. Incidentally, infection of macrophages with *Salmonella typhimurium* or *Lysteria monocytogenes* were also reported to alter isoform usage, in addition to a widespread shortening of 3'UTR[14]. The macrophage transcriptome also appears to get spliced differently during other inflammatory conditions [15,16]. However, the mechanism determining differential isoform usage in most of these conditions is unclear. Given the fundamental function of RNA splicing in shaping the host proteome and its perturbation by Mtb during macrophage infections, insights into the mechanism of infection-dependent alternative RNA is critical to understand molecular details of tuberculosis pathogenesis[9]

94 Host RNA splicing is a complex and fundamental physiological process in all eukaryotes.
95 RNA splicing is performed by a multimolecular and dynamic spliceosome complex, which
96 employs a large number of proteins to ensure the accuracy of splicing to form mature
97 mRNAs [17-19]. One of the key components is U-rich U snRNAs, which form protein-RNA
98 complexes known as small nuclear ribonucleoproteins (snRNPs)[20-22]. The snRNPs are
99 recruited stepwise onto the pre-mRNA exon-intron boundaries to form a functional
100 spliceosome, join the adjoining exons, and splice out the intervening intron [18,23].
101
102 Like other bacterial pathogens, Mtb also secretes virulence factors to alter host physiology [24-
103 29]. Since mycobacterial secretory proteins in the infected macrophages have access to both
104 cytosol and the nucleus [24-26,30], we asked whether secreted Mtb proteins could impact
105 spliceosome complex assembly and, consequently, host RNA splicing. This study extensively
106 maps the interactions between mycobacterial secretory proteins and host proteins involved in
107 RNA splicing. We identify the Mtb proteins capable of altering the assembly of spliceosome
108 complex by exploiting an evolutionarily conserved quality control mechanism of RNA
109 splicing in the eukaryotes.
110
111

112 **Results**

113 **A targeted high-throughput yeast-2-hybrid (Y2H) screen to identify *Mtb*-human protein**
114 **interactions**

115 Eukaryotic RNA splicing is a highly conserved, multi-step process involving a large number
116 of proteins and protein-RNA complexes (herein referred to as splicing factors or SF). Our
117 hypothesis was to test whether Mtb secretory proteins could directly interact with the SFs to
118 interfere in the host RNA splicing. For this, we performed a high-throughput one-to-one
119 yeast-2-hybrid assay. The bait library (binding domain or BD) comprised of mycobacterial
120 secreted proteins while the prey library (activation domain or AD) comprised the protein-
121 coding isoforms of genes belonging to functional classes "RNA splicing" and "Spliceosome
122 function", making it a unique custom library of human transcripts (Table S1). Briefly, our BD
123 library contained 384 *Mtb* genes and also included 37 putative negative controls (Fig. S1A-B
124 Table S1 and S2)[24,27-29,31].

125 Four independent mating reactions were performed between 384 BD and 162 AD clones to
126 obtain high-quality interaction data, leading to 2,48,832 mating reactions. The schematic of
127 the Y2H strategy followed for the study is shown in Fig. 1A. In total, 123 Mtb proteins
128 showed interactions with one or more SFs, together consisting of 439 incidences of
129 interactions, including 121 autoactivation incidences (Table S3). The BD clones showing
130 autoactivation were excluded from the list of interactors. A few negative control list proteins
131 also showed interactions, including 56 autoactivation incidences (Table S3).

132 The interactome between *Mtb* secretory proteins and host SFs, as observed by the Y2H
133 screen, is shown in Fig. 1B (Table S4). The interactions spanned across all classes of host
134 genes involved in RNA splicing, including RNA polymerases, Sm proteins, SMN complex
135 proteins, U2 snRNPs, LSm proteins, HnRNPs, U4/U6.U5 snRNP tricomplex proteins and
136 non-snRNP proteins (Fig. 1C). The Venn-Diagram shows the number of Mtb proteins

interacting with the host SFs (Fig. 1C). Overall, host proteins from RNA Pol II subunits, Sm Core, catalytic step II factors, HnRNPs and SMN complex (Fig. S1C) and Mtb proteins from information pathway, virulence factors and conserved hypothetical proteins functional class showed most interactions (Fig. S1D). To note, there were many instances where different proteins of similar functional sub-class shared their interacting partner with *Mtb* (Fig. 1D).

Interaction of Mtb proteins with the splice variants of human proteins

Since the human library consisted of all protein-coding isoforms of the corresponding genes, among the 162 AD clones that were screened, a total of eighteen genes were represented by more than one transcript (Fig. S2A). Among them POLR2H, SYF2, SNRPD2, LSm7, RBMX, LUC7L2, DCAF4, PPIH and U2AF1 showed interaction with Mtb protein in Y2H (Fig. S2A). The analysis of interaction patterns observed for the different transcripts of the same gene revealed unique insights (Fig. S2B). It shows the gain or loss of interacting partners as a consequence of the gain or loss of exons (Fig. S2B). Most strikingly, truncated isoforms for three genes showed an unusually high number of interactions with the Mtb proteins (Fig. S2B). In our earlier report, we noted a significant increase in the expression of truncated protein-coding isoforms of many host genes, which appeared to help Mtb. Possibly, the truncated spliced variants take confirmations that allow it to interact with more partners. However, this remains a preliminary observation and needs a more focused study to specifically understand the underlying mechanisms and consequences.

***In vitro* splicing assays point to a direct role of *Mtb* proteins in regulating host-RNA splicing.**

We asked whether the mycobacterial proteins interacting with the host SFs, either through Y2H or MS approach, could also alter host RNA splicing. For functional assays, we selected transcripts from a handful of genes like *ACSL1*, *ACADVL*, *DDX11*, *DDX42* and *NFYA*, each representing a specific exon exclusion event upon H37Rv infection (Fig 1E)[11]. These

162 transcripts were specifically upregulated in H37Rv infected macrophages but not in H37Ra
163 infected ones[11]. We designed primers to screen for the specific splicing events (Fig. S3A).
164 The selected transcripts from the above genes showed an exon exclusion event specifically
165 upregulated upon H37Rv infection (Fig. S3B). The increase was not due to an increased
166 transcriptional event since the exon inclusion primers did not show any change in expression
167 (Fig. S3C). Moreover, using primers that spanned the junction between flanking exons after
168 excluding the middle exon, we reconfirmed that the increased expression of the transcripts
169 noted above was indeed due to an increased exon exclusion event (Fig. S3D). The first set of
170 functional assays involved *in vitro* splicing assays. For *in vitro* splicing assays, we generated
171 minigene constructs for four specific isoforms for the genes *ACSL1*, *ACADVL*, *DDX11* and
172 *DDX42*. The schematic details for *in vitro* splicing assay are shown in Fig. 1E and Fig.S3E.
173 We verified the substrate RNAs produced via *in vitro* transcribing the minigene constructs,
174 undergo splicing and form the desired product in the presence of splicing-competent nuclear
175 extract from THP-1 macrophages (Fig. S3F).
176 Next, to test whether Mtb proteins could directly interfere with RNA splicing, we purified
177 several Mtb proteins identified in the Y2H screen and spliceosome complex pulldown assays
178 (Fig. 1, Fig. S1, Fig. S3G) and used them in independent in *vitro* splicing assays. Most Mtb
179 proteins did not have any impact on the inclusion or exclusion of the test exon in the *in vitro*
180 assays, a few mycobacterial proteins appeared to cause the exclusion of test exon in the case
181 of ACADVL minigene substrate (Fig. 1F). Thus, Mtb secretory proteins could indeed
182 selectively alter the choice of inclusion/exclusion of specific exons. While the *in vitro* assay
183 provided evidence for the possible role of Mtb proteins in altering the host RNA splicing, it
184 did not capture the mechanism for global impact on RNA splicing as reported earlier and
185 which was our major interest [11].

186 ***Mycobacterium smegmatis*-based knock-in system to verify the role of *Mtb* proteins in**
187 **regulating host RNA splicing**
188 As an alternate approach, we recapitulated the infection-like scenario, by generating 19 non-
189 pathogenic *M. smegmatis* (*Msmeg*) lines inducibly expressing Mtb proteins identified through
190 Y2H (*Rv1390, Rv0170, Rv0639, Rv0780, Rv0288, Rv1094, Rv3864, Rv1566c, Rv1643,*
191 *Rv3849* and *Rv1435c*) (Fig S4A). These were sub-cloned under the IVN-inducible promotor
192 to develop corresponding knock-in *Msmeg::Rv* lines (Fig. 2A). We next used each of the
193 *Msmeg::Rv* lines to infect THP-1 macrophages and analyzed specific RNA splicing events of
194 a set of target genes through RT-PCR using the exclusion primers (Fig. S3A), against
195 transcripts from *ACSL1*, *ACADVL*, *DDX11*, *DDX42* and *NFYA* (Fig. S4A). After the first
196 biological experiment, out of 11 genes screened 6 genes (Rv1566c, Rv1643, RV1435c,
197 Rv3849, Rv1390) showed IVN-induction dependent AS event in one or more of the five test
198 exons analysed (Fig. S4A). Out of these Rv1390 did not show consistent patterns during the
199 subsequent biological replicates (Fig. S4B). The *Msmeg::Rv* lines expressing *Rv1566c*,
200 *Rv1435c* or *Rv1643* showed significant exon skipping events for one or more test exons
201 analyzed across three biological replicates (Fig. 2B). Another *Msmeg::Rv* line, expressing
202 Rv3849, also showed consistent AS events across three biological replicates (Fig. S4B);
203 unlike Rv1566c, Rv1435c and Rv1643, we could not confirm its secretion from the
204 *Msmeg::Rv* line and therefore Rv3849 was not taken forward for further studies (Fig. S4C).
205 Most importantly, we observed insignificant changes in the AS for the target transcripts when
206 *Msmeg* alone or upon IVN-induction were used for infection (Fig. S4D). Thus, with the
207 limitation of using only five test exons, each Mtb protein expressed in *Msmeg* impacted the
208 splicing of a unique and overlapping subset of test exons analyzed (Fig. 2B). Despite the
209 limitations, the results from this surrogate approach suggests that select Mtb proteins were
210 able to alter specific host RNA splicing events during infection in THP-1 macrophages.

211 ***In-cell* splicing reporter assays illustrate the direct role of *Mtb* proteins in regulating**
212 **host-RNA splicing.**
213 We next developed an *in-cell* splicing reporter assay to test the three Mtb proteins identified
214 in the results in Fig. 2B. The reporters included a constitutive pRint vector and a set of test
215 pGint vector[32]; the latter was modified to incorporate the custom exon-intron boundaries from
216 our list of splicing events (Fig. 2C). In HEK293T cells, stably expressing pRint, pGint
217 constructs (pGint_ACADVL, pGint_DDX11, pGint_DDX42, pGint_ACSL1 and
218 pGint_NFYA) were co-transfected with plasmids encoding *Rv1566c, Rv1435c and Rv1643*
219 respectively, in separate experiments (Fig. 2D). We used a plasmid encoding *Rv1252c* as a
220 non-interacting control (NIC) for this experiment. We measured the %pGint$^+$ cells (GFP$^+$) in
221 the pRint$^+$ gated population (Fig. 2D). As shown in Figure 2E, transfecting NIC marginally
222 but consistently increased the splicing event for *ACADVL* and *ACSL1*. However, transfecting
223 *Rv1566c*, *Rv1435c,* and *Rv1643* further increased the exon exclusion event in *ACADVL* but
224 not in *ACSL1* (Fig. 2E). NIC transfection had no impact on the exon exclusion event for
225 *DDX11*, *DDX42* and *NFYA* (Fig. 2E). In contrast, expressing *Rv1566c*, *Rv1435c* or *Rv1643*
226 led to a significant increase in the exon exclusion event for *DDX11*, *DDX42* and *NFYA* (Fig.
227 2E). We also analyzed pRint expression in pGint$^+$-population across all the conditions, which
228 did not show any significant differences, suggesting no impact on the basal splicing event as
229 a result of ectopically expressing Mtb proteins in 293T cells (Fig. S5A). Taken together,
230 through the validation experiments in figures 1-2, *in vitro* assays, *ex vivo* infection assays,
231 and *in-cell* reporter-based assays, we demonstrate that the Mtb proteins identified in this
232 study were capable of systemically altering host RNA splicing in a targeted fashion.
233 **Verification of physical interaction between Mtb proteins and host splicing factors**
234 The three Mtb proteins identified above (Rv1435c, Rv1566c and Rv1643) were next verified
235 for their physical interaction with the human SFs in a mammalian system. Mtb proteins

236 (Flag-tagged) and human SFs (3X-HA-tagged) were co-transfected in HEK293T cells, and
237 co-immunoprecipitation was performed using anti-HA and anti-FLAG antibodies,
238 respectively. The Y2H interaction combinations thus verified included PRPF18-Rv1435c,
239 SNRPD2-Rv1435c, GEMIN3-Rv1435c, POLR2H-Rv1566c, and SNRPD2-Rv1643 (Fig. 2F).
240 The combination of SNRPD2-Rv1566c, while showed interactions in the Y2H screen, did not
241 show interaction in the co-IP experiment (Fig. S5B). We also included a few combinations
242 that did not show interaction in the Y2H screen as the additional negative control. Thus, co-
243 transfecting *SNRPF* with *Rv1566c* in HEK293T cells did not result in their co-
244 immunoprecipitation (Fig. S5B). On the other hand, co-transfecting *SNRPF* with *Rv1435c*
245 showed only limited interaction (Fig. S5B). While Rv1435c could pulldown SNRPF, the
246 reverse was not true (Fig. S5B). For a couple of Rv1435c interactions, we also verified the
247 interaction by exchanging the BD and AD clones in Y2H (Fig. S5C). Overall, from the above
248 set of results, we conclude that the Mtb proteins, which were able to alter targeted host-RNA
249 splicing in the functional assays, also showed physical interactions with the host splicing
250 factors in the mammalian system. Finally, each of these three proteins was able to alter exon
251 inclusion for ACADVL test transcript in the *in vitro* splicing assay (Fig. S5D).

252 **Mtb infection impacts levels and distribution of splicing-associated proteins in**
253 **macrophages.**

254 The results above together show that select Mtb proteins could interact with host SFs and
255 alter the host RNA splicing. However, these results did not provide any insights into the
256 possible mechanisms. We next analyzed *in vivo* assembled spliceosome complexes from
257 Mtb-infected macrophages by label-free quantitative proteomics. The spliceosome complexes
258 were pulled down using anti-SNRPB (Y12) antibody from nuclear and cytosolic lysates of
259 the infected and uninfected macrophages to analyse the abundance and distribution of host
260 spliceosome proteins (Table S5). In the nuclear fractions of the Mtb-infected cells, there was

261 a significant depletion of SR proteins like SRSF3, SRSF7, Sm Core proteins like SNRPC,
262 SNRPB, SNRPN and SNRPD2 and some HNRNP proteins, as compared to the uninfected
263 ones (Fig. 3A). Other proteins like RPL4, SF3B2 and SYNCRIP followed a reversed pattern
264 in the infected cells, getting more enriched in the cytosol unlike the uninfected cells (Fig.
265 3A). Similarly, proteins like EIF5A and EIF5AL1 showed increased nuclear enrichment in
266 the infected macrophages, in contrast to the uninfected cells (Fig. 3A). With the limited
267 example of proteins identified in the immunoprecipitated spliceosome complexes, it was
268 evident that a diverse set of SFs were either getting depleted or mislocalized in the Mtb-
269 infected macrophages.

270 **Mtb infection dramatically alters the levels and distribution of proteins involved in**
271 **RNA splicing.**

272 To specifically investigate the levels and distribution of host SFs in Mtb-infected cells, we
273 systematically analyzed a representative set of SFs involved at different stages of RNA
274 splicing (Fig. 3B). We monitored three different parameters- MFI/cell, foci/cell and the
275 cytosolic to the nuclear ratio (Cyt/Nuc) for these proteins. (Fig. 3C-H). We also included
276 some transcription-related proteins for this analysis, since splicing begins while the RNA is
277 still being transcribed [33,34]. Both POLR2G and RBP8 (transcription related proteins) showed,
278 unexpectedly, increased nuclear retention (S6A). The U1 snRNP protein SF1 showed
279 increased MFI and increased cytosolic retention upon infection (Fig. 3C). The U5 snRNP
280 component DDX23 showed a decrease in MFI upon infection (Fig. 3D). The SMN complex
281 proteins GEMIN8 and GEMIN3 showed differential MFI and foci/cell (Fig. 3E-F). A couple
282 of SFs like PRPF4 (U4/U5.U6 associated) and SNRP200 (U5 associated) did not show
283 alteration in any of the three parameters analyzed (Fig. S6B). Among the Sm proteins, LSM7
284 showed decreased MFI/cell in the infected macrophages and SNRPF showed an increased
285 cytosolic retention (Fig. 3G-H). However, a more perplexing observation was the direct

286 interaction of SNRPF with Mtb phagosomes (Fig. 3H). It also appeared that SNRPF was
287 getting sequestered in the cytoplasm due to its strong interaction with Mtb phagosomes (Fig.
288 3I). The results in Figure 3 together indicated that the components of the RNA splicing
289 machinery, specifically snRNP complex biogenesis, could be dysregulated in the Mtb-
290 infected macrophages.

291 **U5 snRNA directly interacts with *Mtb* phagosomes**

292 The nuclear-cytosolic transport of U-rich snRNAs is fundamental to their assembly into U
293 snRNPs, which, together with different snRNPs, form the spliceosome complex[35]. The
294 snRNP complex assembly requires cytosolic transport of the snRNAs, which then interact
295 with SMN complex and Sm Core proteins to form snRNP before entering back into the
296 nucleus to form spliceosome complexes (Fig. 4A). Since SNRPF, a part of the pentameric
297 partial Sm ring, which first recognizes the Sm-sites present on snRNAs before the assembly
298 of the full heptameric ring takes place with the help of the SMN complex [20], was increasingly
299 retained in the cytoplasm of Mtb-infected macrophages, we anticipated that some snRNAs
300 could also show dysregulation in the infected macrophages. We analyzed the U2 and U5
301 snRNAs by fluorescence in situ hybridization (FISH) in the Mtb-infected macrophages. The
302 U2 FISH did not show any aberrant pattern for the parameters- MFI/cell, foci/cell and
303 Cyt/Nuc ratio, except for an increased MFI/cell at 24 hours post-infection (hpi, Fig. 4B).
304 Unlike U2 FISH, U5 FISH showed remarkable patterns in the Mtb-infected macrophages
305 (Fig. 4C). There was a decrease in MFI/cell and foci/cell at 24 hpi; however, a substantial
306 increase was noted in all three parameters at 48 hpi (Fig. 4C). At 48 hpi, there was significant
307 cytosolic retention of U5snRNA (Fig. 4C). More surprisingly, as in the case of SNRPF, we
308 noted direct interaction of U5snRNAs with Mtb in the cytosol (Fig. 4D), which quantitatively
309 increased from 24 to 48 hpi (Fig. 4D). The direct interaction of Mtb with SNRPF and
310 U5snRNA suggests binding and cytosolic sequestration of these specific splicing components

311 by Mtb. This would also mean stoichiometric disbalance among the components of snRNPs
312 in the infected macrophages, a condition which is known to cause widespread splicing defects
313 [9,36]. Interestingly, we verified that the expression of both U5 snRNA (by qPCR) and SNRPF
314 (by immunoblot) consistently increased post-infection, which could be a compensatory
315 mechanism to sustain the snRNP biogenesis in the Mtb-infected cells (Fig. 4E-F). Therefore,
316 disruption of snRNP biogenesis appears to serve as the major mechanism that triggers large-
317 scale splicing alterations in the Mtb-infected macrophages.

318 ***Rv1435c* deletion strain of Mtb fails to alter the splicing of target transcripts.**

319 We next asked whether the Mtb proteins identified in this study were directly responsible for
320 dysregulating snRNP biogenesis. To test the above, we deleted one of the three identified
321 Mtb proteins, *Rv1435c,* in H37Rv using recombineering (Fig. S6C)[37]. This protein is
322 annotated as a secretory protein and showed positive results in most of the interaction and
323 splicing assays. We confirmed the deletion of *Rv1435c* by PCR (Fig. S6D) and also generated
324 a complemented strain *ΔRv1435c::WT* using an integrative vector (see Methods). We first
325 verified the impact of *Rv1435c* deletion on the splicing of five target transcripts used in this
326 study. All five target transcripts, *ACADVL, DDX11, DDX42, ACSL1* and *NFYA*, showed
327 significantly reduced test exon exclusion events for these transcripts upon infection with the
328 H37Rv*ΔRv1435c* strain compared with the WT Mtb (Fig. 5A). Moreover, in the cells infected
329 with H37Rv*ΔRv1435c*::*WT*, the exon exclusion event was restored to the same extent or
330 higher than the WT infected cells (Fig. 5A). These results conclusively show that Rv1435c
331 can directly alter host RNA splicing during Mtb infection. We designated the Mtb protein
332 Rv1435c as the host splicing regulator 1 or Hsr1.

333 **H37Rv*Δhsr1* alters infection-induced global alternative splicing.**

334 We next wanted to see whether the loss of *hsr1* causes global recovery in the infection-
335 induced RNA splicing events. We performed RNAseq analysis from the THP-1

336 macrophages, uninfected or infected with WT or H37Rv*Δhsr1* strains and analyzed the
337 relative abundance of different splice variants of the protein-coding genes. The clustergram in
338 figure 5B shows that H37Rv infection causes increased expression of a large number of
339 transcripts, most of which were effectively restored in H37Rv*Δhsr1* infected cells (Fig. 5B).
340 Interestingly, H37Rv*Δhsr1*-infected cells show increased expression of another unique set of
341 transcripts, likely represented those splicing events, which were possibly inhibited during WT
342 H37Rv infection (Fig. 5B). Curiously, a large number of transcripts showing upregulation in
343 the Mtb-infected macrophages actually represented NMD substrates or processed transcripts
344 (Table S6), which usually don't form a functional protein. Expression of splice variants that
345 are either truncated in length or do not code for a functional protein is one of the ways
346 through which infection-induced alternative splicing helps Mtb evade host responses [11]. The
347 results here show a direct role of Mtb protein Hsr1 in manipulating host RNA splicing events
348 favoring the pathogen.

349 **Mtb *hsr1* directly perturbs snRNP biogenesis by sequestering U5snRNAs to alter host**
350 **RNA splicing.**

351 We next investigated how the loss of Mtb *hsr1* contributed to the reversal in the splicing
352 events in the macrophages. We analyzed U5snRNA interaction and distribution in Mtb*Δhsr1*
353 infected cells by FISH. In concurrence with the altered RNA splicing events reported above,
354 Mtb*Δhsr1* infected cells did not show cytosolic retention of U5snRNA (Fig. 5C). On the
355 similar line, the interaction between Mtb and U5snRNA was almost entirely lost in Mtb*Δhsr1*
356 infected cells (Fig. 5C). Like U5 snRNA, even SNRPF interaction with Mtb*Δhsr1* was
357 significantly lower than that observed with the WT Mtb (Fig. 5D). The effect of *hsr1* deletion
358 on the regulation of U5snRNA and SNRPF was specific since we did not observe any
359 difference in any of the parameters analysed for U2snRNA (Fig. S6E). These results imply

360 that Mtb Hsr1 perturbs host snRNP biogenesis by sequestering the U5snRNA and SNRPF in
361 the cytosol, leading to widespread RNA splicing defects.

362 **Hsr1-mediated impact on host RNA splicing is ESX1-dependent**

363 While we did observe secretion of Hsr1 from Mtb as well as when expressed in *Msmeg*, it
364 was not evident whether it also gets secreted inside the macrophages during infection. We
365 generated anti-sera against Hsr1 in mice to check its secretion and localization during
366 infection in macrophages. Interestingly, we could observe significant Hsr1 secretion in the
367 cytosol of THP-1 macrophage infected with the WT or *Δhsr1::WT* Mtb (Fig. 5E). The Hsr1
368 signal was almost completely lost in the cells infected with *Δhsr1* (Fig. 5E). Curiously, when
369 we used the *ΔCE* Mtb strain (lacking both *Esx1*/ESAT6 and *Esx2/CFP10*), we could not see
370 any cytosolic staining for Hsr1. The Hsr1 staining in THP-1 cytosol was restored when the
371 complemented strain *ΔCE::WT* was used for infection (Fig. 5E). Expectedly, THP-1
372 macrophages infected with Mtb strains lacking the *esx1* system failed to alter RNA splicing
373 for the target genes like ACSL1 and ACADVL (Fig. 5F). Consistent with the above
374 observation, the Mtb *ΔCE* strain could not alter SNRPF cytosolic/nuclear distribution (Fig.
375 S6F). These results suggest that the Hsr1-mediated impact on host RNA splicing reported in
376 this study is ESX1-dependent.

377 **Evolutionary perspective of hsr1 in mycobacterial virulence**

378 On UniProtKB, Rv1435c is characterized as a conserved proline, glycine, valine-rich secreted
379 protein. There are several orthologs present in diverse mycobacterial species, like *M.*
380 *marinum*, where it is classified as a conserved hypothetical protein; in opportunistic
381 pathogenic species like *M. paratuberculosis, M. avium,* and *M. helveticum,* the identified
382 orthologue codes for secreted beta-xylosidase enzyme (Fig. S7A-B). *In M. smegmatis*, the
383 orthologue, with very poor similarity, encodes a translation initiation factor (IF-2). In a few
384 pathogenic species like *M. lacus and M. heidelbergense,* it codes for a beta-xylosidase (Fig.

385 S7A-B). Thus, most *hsr1* orthologues in the genus *Mycobacterium* encode a secreted protein,
386 and their functions are better characterized in non-pathogenic or opportunistic pathogens. It is
387 likely that the ability of hsr1 to alter host splicing events was likely a gain-of-function event
388 during the course of Mtb co-evolution as a human pathogen.

389 **Mtb Hsr1 is a virulence factor.**

390 Since Mtb*Δhsr1* infection could not alter RNA splicing events as in WT Mtb-infected cells,
391 we wondered whether loss of *hsr1* also impacts bacterial pathogenicity. The Mtb*Δhsr1*
392 showed reduced intra-macrophage survival compared to the WT Mtb (Fig. 6A, S8A). The
393 Mtb*Δhsr1* strain, as such, did not have a growth defect *in vitro*. Rather, it grew faster than the
394 WT cells in liquid culture which was consistent with a decrease in the bacterial cell length
395 (Fig. S8B-C). Curiously, *Δhsr1::WT* grew inside the THP-1 macrophages even better than the
396 WT Mtb, pointing to Hsr1 being an important virulence factor (Fig. 6A, Fig. S8A).
397 Interestingly, *Δhsr1* showed increased targeting to the lysosomes and also the cell had higher
398 cathepsin D staining (Fig. S8D). Since Mtb*Δhsr1* was unable to bind and sequester SNRPF,
399 we wondered whether overexpressing SNRPF during WT Mtb infection could have an impact
400 on Mtb CFU. Overexpressing SNRPF in human macrophages caused a relative decline in
401 cytosolic retention of SNRPF and a reduced intracellular CFU of WT Mtb (Fig. S8E),
402 indicating that SNRPF sequestration by Hsr1 serves as the virulence mechanism of Hsr1. On
403 the contrary, overexpression of SNRPD2 did not have any impact on Mtb CFU (Fig. S8F).
404 Next, to assess the role of Hsr1 in bacterial virulence *in vivo*, we infected BALB/c mice with
405 WT, Mtb*Δhsr1* and Mtb*Δhsr1::WT* strains and analyzed bacterial CFU at two and four weeks
406 post-infection. At both two- and four-weeks post-infection, Mtb*Δhsr1* had fewer CFUs in the
407 lungs of the animals (Fig. 6B). Unlike in THP-1 macrophages, the complemented strain
408 Mtb*Δhsr1::WT* was not able to rescue the bacterial CFU (Fig. 6B). At both 2- and 4- weeks
409 post-infection, analysis of a panel of eight cytokines in the lung homogenates showed an

410 increase in cytokine levels in WT infected animals, which, to some extent, were diminished
411 in the Mtb*Δhsr1* infected mice (Fig. S8G). The lung homogenates from mice infected with
412 Mtb*Δhsr1::WT* did not revert the expression to the WT levels (Fig. S8G). The lung tissue
413 sections from mice that were infected for four weeks showed strong SNRPF staining in case
414 of WT and *Δhsr1::WT* Mtb infection (Fig. 6C). The SNRPF staining was strikingly lower in
415 uninfected mice or those infected with Mtb*Δhsr1* (Fig. 6C). Remarkably, SNRPF staining in
416 the WT or *Δhsr1::WT* Mtb infected mice appeared mostly cytosolic in nature. Together, the
417 *ex-vivo* assays in macrophages and *in vivo* assays corroborate Mtb Hsr1 as a virulence factor
418 which helps its survival in the animals, largely by sequestering the host SNRPF.

419 **Macrophages in human TB granulomas show altered snRNP complex component**
420 **localization.**

421 Specific SNRPF cytosolic retention in THP1 macrophage and in the lungs of infected mice
422 prompted us to explore the clinically relevant human samples. Tissue sections from the
423 confirmed human TB granulomas belonging to abdominal TB cases were stained with
424 SNRPF. To our surprise, all granuloma lesions showed strong SNRPF staining (Fig. 6D,
425 S9A). Moreover, it was evident that there were more cytosolic SNRPF than in the nucleus
426 (Fig. 6D). Lesions that were highly necrotic showed even higher SNRPF levels (Fig. 6D).
427 Multinucleated giant cells, a hallmark of TB granulomas, showed intense staining for
428 cytosolic SNRPF (Fig. S9A). Unlike the granuloma lesions, in healthy tissue sections,
429 SNRPF staining was not remarkable (Fig. S9B). Finally, we also stained for another sm Core
430 protein, i.e., SNRPD2, in the granulomatous lesions or healthy sections. In both cases,
431 SNRPD2 staining was conspicuously nuclear and indistinctive (Fig. S9C). Thus, these results
432 highlight an unusual SNRPF-dependent dysregulation in human TB lesions. No previous
433 report of such rampant dysregulation of snRNP biogenesis in human TB granulomas exists.

434 **Monocytes from human pulmonary TB patients show pervasive regulation of**
435 **alternative splicing.**
436 Next, from an RNAseq dataset from monocytes of 20 PTB cases and 19 healthy contacts, we
437 analyzed and compared the relative abundance of different transcripts [38]. We focused our
438 analysis on four major functional classes- RNA splicing, immune response, inflammation and
439 metabolism (Fig. 6E, Table S7). As can be seen from the clustergrams, transcript level
440 expression of genes belonging to each of these functional classes was able to classify the PTB
441 and HC subjects (Fig. 6E). In each of the cases, there were two major clusters of transcripts-
442 one showing increased expression in the TB cases (Cluster 1) and the other in the control
443 cases (Cluster 2). We analyzed the biotypes of the transcripts present in these clusters across
444 all the groups. In the RNA splicing class, cluster 2 was much bigger than cluster 1, showing
445 more downregulation events in the TB cases (Fig. 6E). Moreover, the small number of
446 transcripts that were upregulated in TB subjects were either not the protein-coding isoforms
447 or encoded a truncated protein isoform (Fig. 6F). A similar analysis performed for other
448 functional classes revealed systemic splicing defects (Fig. 6F). In inflammation, immune
449 response and metabolism classes, while a large number of transcripts were upregulated in the
450 PTB subjects, a significant number of those genes did not code for a protein or encoded a
451 truncated isoform of the corresponding protein (Fig. 6F). For example, the upregulated genes
452 in the immune response class show increased expression of the transcripts of genes like
453 TLR4, THBS1, OAS, OASL3, GRN, CXCL16, etc. (Fig. 6G), however for each of them one
454 or more splice variants with highly truncated protein length were also upregulated, which
455 could significantly blunt the host responses. Comparison of monocyte expression data with
456 those from H37Rv-infected THP-1 (Fig. 5B) showed around 45 commonly upregulated
457 transcripts; however, almost a third of those genes did not code for a protein-coding variant
458 (Fig. S9D, Table S8). Taken together, these observations make it evident that by perturbing

459 host RNA splicing, the mechanism for which we report in this study, Mtb exerts a pervasive
460 influence on host physiology and contributes towards systemic pathologies witnessed in TB
461 patients.

462 **Discussions**

463 RNA splicing is a conserved eukaryotic process which impacts cellular proteome diversity
464 [10,39]. Also, its dysregulation is associated with diverse pathologies [40]. This study was designed
465 to understand the mechanism of Mtb infection-mediated alteration of host RNA splicing[9].
466 Hijacking of host RNA splicing machinery by viral pathogens, including influenza virus and
467 SARS-CoV2 is well-known [41-44]. However, very little is understood in the context of
468 bacterial pathogens. Identification of several Mtb proteins that could directly interact with
469 and alter host RNA splicing is a significant leap in our understanding of host-pathogen
470 interactions. Similarly, the observation that different spliced variants could interact
471 differently with Mtb proteins adds a new dimension to host-pathogen interactions. There are
472 only limited efforts to systematically study and map out how alternative splicing events could
473 contribute to the functional divergence in the proteome [45]. However, our results underscore
474 that more efforts are needed to understand the consequences of this fundamental aspect of
475 eukaryotic regulation.

476 The dysregulation in the abundance and distribution of SFs in the infected macrophages
477 provided crucial insights into the underlying mechanism. The U snRNPs biogenesis is
478 fundamental to eukaryotic gene expression, given their critical role in RNA splicing [46-48]. For
479 snRNPs to form, pre-snRNAs from the nucleus are transported to the cytoplasm, where they
480 interact with the multi-molecular SMN complex and Sm Core ring [46-48]. The snRNA-protein
481 complex comprising SMN complex and Sm Core ring, is transported back to the nucleus
482 where they mature into snRNPs and participate in the splicing reactions [49,50]. SMN complex,
483 including SMN1/2 dimer and Gemin proteins, charge snRNAs with the Sm core ring [51]. A
484 loss or altered stoichiometry in either the Gemins or SMN1 or 2, can lead to widespread
485 splicing defects in different tissues [20,52-54].

486 Thus, while a few Mtb proteins could alter RNA splicing of test transcripts in the *in vitro*
487 assays, the dysregulation of SMN complex proteins and Sm core proteins reported in this
488 study appears to be the major driver for the global alteration of AS in Mtb-infected
489 macrophages. To our surprise, we also noted direct interactions between SNRPF and U5
490 snRNA with Mtb, which suggests potential sequestration of these key regulatory components
491 by Mtb. While several viral proteins are reported to directly interact with the SFs [9,36,44,55],
492 there is no precedence of such interaction of SFs or snRNAs directly with a viral or bacterial
493 pathogen.
494 There could be concerns about the specificity of this regulation since targeted dysregulation
495 of splicing factors (SFs) can lead to a pervasive impact on alternative splicing. However,
496 studies suggest it does not alter the splicing of all introns in the cell [19,56,57]. Rather, the impact
497 of one splicing factor dysregulation could be very distinct and selective from another SF
498 dysregulation [19]. At the same time, there is some functional congruence, too, in that a specific
499 SF could regulate AS of functionally related genes [19,56,57].
500 Experiments using the *hsr1* deleted Mtb strain provide definite proof that Mtb Hsr1 regulates
501 host RNA splicing in an ESX1-dependent manner. More strikingly, the loss of *hsr1* impacted
502 the ability of Mtb to interact and sequester U5snRNA or SNRPF. It is pertinent to ask why
503 Mtb targets U5snRNA and no other snRNAs or at least the U2snRNAs, which we analyzed.
504 During splicing, U5snRNPs recognize the acceptor splice site [58]. Several viral proteins
505 interact with the U5 components to disrupt the normal RNA splicing [36,55]. Thus, targeting of
506 U5snRNA through interaction reveals somewhat conserved pathogenic mechanisms across
507 diverse human pathogens. The decline in Mtb*Δhsr1* survival in macrophages and in mice
508 clearly indicates that the ability to disrupt snRNP biogenesis is a desirable phenotype for
509 tuberculosis pathogenesis. To note, the orthologues of *hsr1* are secreted xylose hydrolysing
510 enzymes, which is probably required by non-pathogenic, or opportunistic pathogenic bacteria

511 in search of diverse carbon sources. Therefore, the gain of function to interact with host SFs
512 appears to be an important adaptation during the evolution of Mtb as a human pathogen.
513 Interestingly, all snRNAs, with a few exceptions, while synthesized in the nucleus, must
514 travel to the cytosol for assembly before they are transported back to the nucleus for
515 participation in the splicing reaction [49]. The cytosolic export of pre-snRNAs ensures that
516 immature snRNAs do not get incorporated into the spliceosomal complexes in the nucleus,
517 which can alter host RNA splicing catastrophically [35,59]. This evolutionarily conserved quality
518 control step appears to be co-opted by the pathogenic mycobacteria during its evolution as the
519 human pathogen.
520 Tuberculosis granulomas evolve through a complex interplay between the bacterial factors
521 and host machinery, which largely benefits the pathogen [60]. Therefore, the intense SNRPF
522 staining of human tuberculosis granulomas is of great interest. Similarly, the monocyte
523 RNAseq data from the patients emphasize the pervasiveness of infection-induced alteration in
524 host RNA splicing. Incidentally, differentially expressed transcript isoforms were recently
525 reported in an LTBI cohort and identified those associated with resistance to TST/IGRA
526 conversion [61]. There are systemic inflammation, metabolic dysregulation and muscle loss
527 reported in TB patients [62], some of which could potentially be a consequence of the large-
528 scale splicing defects observed in these individuals.
529 To conclude, with the limitation of only select splicing events analyzed, we present a new
530 paradigm of pathogen-mediated perturbation of host physiology and show the hijacking of
531 this evolutionarily conserved mechanism by Mtb to blunt host responses. We propose that
532 inhibition of molecular interactions between the host and the pathogenic partners identified in
533 this study could be exploited for developing a novel class of therapeutics.
534
535

Acknowledgements:

The work was supported by the DBT-Wellcome Trust India Alliance Senior Fellowship to DK (IA/S/17/1/503071). The work was also partly supported by grants from SERB (EMR/2016/005296 to DK) and the Department of Biotechnology, Govt of India (BT/IC-06/003/91-Flagship program to DK). All work on virulent Mtb was performed in the Tuberculosis Aerosol Challenge Facility (TACF), a national facility at ICGEB, New Delhi, supported by the Department of Biotechnology, Govt. of India.

Author Contributions:

Original conceptualization: DK; Experiments: KC, NP, YK, DD, NA, PS, BM, HK, LS, RD, RDS, LM, BS, PY; Investigation: KC, DD; Analysis and Visualization: KC, MD, PY, DD, NSB, DK; Resources: VN, AS, VA, PD; Supervision: DK; Writing-original draft: DK, KC, MD, NP; Writing—review & editing: VN, AS, NSB, PD, DK

549 **Figures and Figure Legends**

550

551 **Figure 1: Mycobacterial secretory proteins and host spliceosome proteins interactome.**

552 **A)** The schematic flow chart for high-throughput one-to-one Y2H assay in 1536-well
553 format. Protein pairs showing interaction in 3 or 4 biological replicates were considered
554 as positive interactions.

555 **B)** Mtb proteins -host splicing factor interactome- the positive interactions from the Y2H
556 screen were used to generate the network shown here. The diamond and circle represent
557 the host and Mtb proteins, respectively, and the size corresponds to the number of
558 interactions observed. The proteins coloured yellow in the network are discussed in
559 detail in the text.

560 **C)** The Venn diagrams show proteins belonging to the major spliceosome components and
561 their corresponding overlap with the interacting Mtb proteins.

562 **D)** A more descriptive representation of the Venn diagrams, shown above in Fig. 1C is
563 presented in the table as a list of interactions.

564 **E)** Transcripts that were used for making the minigene constructs to test the role of Mtb
565 proteins in host RNA splicing. Only exons are shown, and two different transcripts for
566 each of the five genes are shown. The red bars indicate the overlapping sequences, and
567 the skipped exons are shown in yellow.

568 **F)** On the left, a schematic of the minigene cassettes for DDX42 and ACADVL is shown.
569 The constitutive exons are shown in green, while the alternative exons are shown in
570 blue. *In vitro* transcribed products of minigenes were used for in-vitro splicing assays
571 in the presence of purified Mtb proteins Rv3867, Rv0360, Rv1566c and Rv1471.

572

573

Figure 2: *Ex vivo* infection assays and reporter assays to test the role of Mtb proteins in host RNA splicing

A) The schematic representation of infecting THP-1 macrophages with *M. smegmatis* strains that express mycobacterial proteins under IVN inducible promotor. Isoform-specific qPCR for the truncated isoforms of ACADVL, DDX11, DDX42, ACSL1 and NFYA was used to assess the impact on alternative splicing.

B) Expression of spliced isoforms of the five target genes (as indicated) in THP-1 macrophages upon infection with WT or *Rv1566c, Rv1435c* and *Rv1643* knock-in *M smegmatis* lines. *M smegmatis* lines were induced with IVN (indicated dose) for 12 hours before infection. Data from three independent biological experiments. The p-values shown are from One-way ANOVA.

C) Designing of the custom *in vivo* fluorescent splicing reporter. Schematic for the constitutive splicing reporter pRint with a short adenovirus intron intervening the DsRed gene with a strong 5' and 3' splice sites. In the pGint splicing constructs, the alternative exon of ACADVL, DDX11, DDX42, ACSL1 and NFYA, along with the flanking 100bp intronic sequences, were inserted within eGFP reading frame. The exon disrupts the reading frame, however, skipping of the test exon reconstitutes EGFP fluorescence.

D) The pRint expressing stable HEK293T cell lines were co-transfected with a pGint construct and pDest-FLAG construct encoding proteins Rv1252c (NIC), Rv1566c, Rv1435c, Rv1643, or empty vector control, followed by Flow cytometry. In the $DsRed^+$ gate, $\%FITC^+$ population was calculated to measure the extent of the test exon-skipping event.

E) $\%FITC^+$ population for each of the indicated reporter constructs in the presence of indicated mycobacterial protein, non-interacting control or vector controls. Each dot represents an independent biological replicate (N=3), indicating p values from One-way ANOVA.

F) Validation of physical interactions between indicated pairs of *Mtb* proteins (FLAG-tagged) and host proteins (HA-tagged) in HEK293T cells. Data is representative of two or more biologically independent transfection and pulldown experiments.

606 **Figure 3: Dysregulation of host RNA splicing machinery in Mtb-infected macrophages**

607 **A)** Heatmap shows massive dysregulation in the level and cellular distribution of human
608 proteins involved in RNA splicing and identified in MS of the spliceosome complex
609 pulldown from the Mtb infected THP-1 macrophages 48 hours post-infection.

610 **B)** A schematic model depicting the ordered pathway of mRNA transcription by RNAP II
611 and splicing catalyzed by spliceosome complex. The U1snRNP binds the 5' splice site
612 and U2 snRNP binds the branch point adenine to form complex A. U4/U5.U6 snRNP
613 binds to form complex B and causes conformational changes to form activated complex
614 B. This is followed by the first transesterification reaction releasing U1 and U4 snRNP
615 and the second transesterification reaction forming complex C releasing the lariat intron
616 and U2, U5 and U6 snRNP. The U snRNP are the major players in splicing catalysis
617 and their biogenesis is mediated by the SMN complex and the Sm proteins in the
618 cytoplasm.

619 **C-H)** The proteins were visualized by indirect immunofluorescence (green), the nucleus
620 was visualized by DAPI staining (blue) and mCherry-expressing bacteria (red) was
621 used for infection. Each dot in the bar graphs represents a field analyzed, and an average
622 of a hundred cells were quantified. The confocal images were analyzed for three
623 different parameters- overall MFI (MFI/cell), number of Foci (Foci/cell) and cytosolic
624 to the nuclear ratio (Cyt/Nuc) for all major splicing-related proteins upon Mtb infection
625 such as **C)** U1 associated proteins (SF1), **D)** U5 associated proteins (DDX23), **E)** SMN
626 complex proteins (GEMIN8) **F)** SMN complex proteins (GEMIN3) and **G)** Sm Core
627 proteins (LSM7) and **H)** Sm Core proteins (SNRPF)

628 **I)** Image showing SNRPF colocalizing with bacterial phagosomes. The 3D reconstruction
629 of the nucleus, SNRPF foci and bacteria (see methods) was done to show the SNRPF
630 Foci at <0.2 microns distance from the bacterial phagosomes. Scale bar: 2 μm and p
631 values indicate One-way ANOVA.

632

633

Figure 4: Dysregulation in SMN complex and Sm proteins impact U snRNA localization and overall assembly of U snRNPs.

A) Schematic representation of U snRNA nucleo-cytoplasmic trajectory and assembly of Sm core ring on snRNA by SMN complex to form snRNPs

B) Representative images for RNA FISH against U2 snRNA in THP-1 macrophages upon Mtb infection 24- and 48-hrs post-infection. The U2 snRNA probe was labelled with Alexa 488 (green), the nucleus was stained with DAPI (blue) and mCherry-expressing H37Rv (red) was used for infection. The plots corresponding to overall MFI (MFI/cell), number of Foci (Foci/cell) and cytosolic to nuclear ratio (Cyt/Nuc) of U2 snRNA are shown here. Each dot represents a field analyzed, and an average of a hundred cells are quantified. The p-values is from one-way ANOVA.

C) Representative images for RNA FISH against U5 snRNA in THP-1 macrophages upon Mtb infection 24- and 48-hrs post-infection. The U5 snRNA probe was labelled with Alexa 488 (green), the nucleus was stained with DAPI (blue) and mCherry-expressing H37Rv (red) was used for infection. The plots corresponding to overall MFI (MFI/cell), number of Foci (Foci/cell) and cytosolic to nuclear ratio (Cyt/Nuc) of U5 snRNA are shown here. Each dot represents a field analyzed, an average of a hundred cells are quantified, and p values are from one-way ANOVA.

D) Direct interaction between H37Rv (red) and U5snRNA at 24- and 48-hrs post-infection. For the plot at the right, the number of U5snRNA puncta within 0.2 microns from H37Rv surface was counted, p values from one-way ANOVA. Representative images for colocalization of Sm core protein SNRPF (green) to the H37Rv phagosome (red). The 3D reconstruction and spot analysis are done as previously described.

E) Expression of U5 snRNA in THP-1 macrophages, 12 and 48 hpi with H37Rv. Data from three independent biological experiments.

F) Immunoblot showing SNRPF protein expression in the whole cell lysates of uninfected (UI) and H37Rv-infected THP-1 macrophages, 48 hpi. Data is representative of two independent biological experiments.

Figure 5: Role of H37Rv hsr1 protein in snRNP biogenesis and altering host RNA splicing.

A) Q-PCR for the target transcripts as mentioned after infection with H37Rv, H37Rv*Δhsr1* and H37Rv*Δhsr1::WT* strain. Data from three biological replicates (N=3), *p<0.05, **p<0.01 (one-way ANOVA).

B) Clustergram of RNAseq data obtained from uninfected, or THP1 macrophages infected with H37Rv or H37Rv*Δhsr1* strains. Transcripts showing an infection-dependent increase (red box), hsr1-triggered increase (purple) or hsr1-blocked transcripts (turquoise) are shown. The distribution of different transcript biotypes across the three groups is shown in the aligned plot towards the right.

C) Representative image for the U5snRNA FISH in THP-1 macrophages infected with WT H37Rv or H37Rv*Δhsr1* strains. The plots at the right show the cytosolic to nuclear ratio and the ratio of interacting to non-interacting U5snRNA with WT H37Rv or H37Rv*Δhsr1*. **p<0.01 (one-way ANOVA).

D) Representative images of SNRPF staining in THP-1 macrophages infected with WT H37Rv or H37Rv*Δhsr1* strains. For the plot at the right, the number of SNRPF puncta within 0.2 microns from the H37Rv surface was counted as interacting and shows the ratio of interacting to non-interacting SNRPF with WT H37Rv or H37Rv*Δhsr1*. *p<0.05 (one-way ANOVA).

E) Representative images of THP1-macrophages stained with DAPI (nucleus) and Rv1435c antisera (green) upon infection with WT H37Rv, H37Rv*Δhsr1,* H37Rv*ΔCE* and H37Rv*ΔCE::WT* and H37Rv*Δhsr1::WT.* Images are representative of two independent biological experiments. Scale bar-2 μm.

F) Expression of truncated isoform as indicated upon infection with WT H37Rv, H37Rv*ΔCE* and H37Rv*ΔCE::WT* at 48hpi from three biologically independent experiments. The p values are from One-way ANOVA.

692 **Figure 6: Regulation of host alternative splicing as a virulence mechanism**

693 A) THP-1 macrophages were infected with WT H37Rv, H37Rv*Δhsr1,* H37Rv*Δhsr1::WT*
694 strains. At 48 hours post-infection, bacterial load was examined using CFU plating.
695 Data represent fold change in CFU at 48 hrs compared to 0 hr post-infection. (N=3), p
696 values are from one-way ANOVA.

697 B) BALB/C mice (6-8 weeks old) were infected with 200 CFU of WT H37Rv,
698 H37Rv*Δhsr1,* H37Rv*Δhsr1::WT* in separate experiments. At 2- and 4 weeks post-
699 infection, bacterial load in the lungs of the animals were determined by CFU plating. P
700 values are from one-way ANOVA.

701 C) The 4 week post-infection lung tissue sections of WT H37Rv, H37Rv*Δhsr1,*
702 H37Rv*Δhsr1::WT* were stained with DAPI (nucleus) and SNRPF. Scale bar: 20 µm

703 D) Human intestinal tuberculosis granulomatous lesions were stained with SNRPF. The
704 arrows indicate multinucleated giant cells with intense staining for cytosolic SNRPF.
705 Black arrows show areas of intense SNRPF staining in the granulomas and in the giant
706 cells.

707 **E)** RNAseq data from the monocytes of pulmonary TB cases (N=20) and healthy contacts
708 (N=19) was analyzed for expression of differential alternatively spliced transcripts.
709 Genes belonging to the indicated functional classes were selected for the cluster
710 analysis. Cluster 1 in each case shows transcripts upregulated in PTB cases.

711 F) For each of the functional classes shown in Fig. 6D, the relative distribution of transcript
712 biotypes (as indicated) is plotted and binned separately for those that were up- (cluster
713 1, red) or down- (cluster 2, blue) regulated in PTB cases.

714 G) The transcripts from the list of genes in the immune response class (Fig. 6D) were
715 further analyzed for their length with respect to the longest encoded isoform of the
716 corresponding gene. Only genes with more than one transcript in the list were chosen
717 for this analysis. The red dotted line separates cluster 1 and cluster 2 transcripts.
718 Transcripts of DYSF and RNASET2 are present in both clusters.

719

720

721

References

1 in *WHO consolidated guidelines on tuberculosis: Module 3: diagnosis - rapid diagnostics for tuberculosis detection WHO Guidelines Approved by the Guidelines Review Committee* (2021).

2 Pisu, D., Huang, L., Grenier, J. K. & Russell, D. G. Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions. *Cell Rep* **30**, 335-350 e334 (2020). https://doi.org/10.1016/j.celrep.2019.12.033

3 Matteucci, K. C., Correa, A. A. S. & Costa, D. L. Recent Advances in Host-Directed Therapies for Tuberculosis and Malaria. *Front Cell Infect Microbiol* **12**, 905278 (2022). https://doi.org/10.3389/fcimb.2022.905278

4 Young, C., Walzl, G. & Du Plessis, N. Therapeutic host-directed strategies to improve outcome in tuberculosis. *Mucosal Immunol* **13**, 190-204 (2020). https://doi.org/10.1038/s41385-019-0226-5

5 Kumar, D. & Rao, K. V. Regulation between survival, persistence, and elimination of intracellular mycobacteria: a nested equilibrium of delicate balances. *Microbes Infect* **13**, 121-133 (2011). https://doi.org/10.1016/j.micinf.2010.10.009

6 Karim, A. F. *et al.* Express path analysis identifies a tyrosine kinase Src-centric network regulating divergent host responses to Mycobacterium tuberculosis infection. *J Biol Chem* **286**, 40307-40319 (2011). https://doi.org/10.1074/jbc.M111.266239

7 Khan, N. *et al.* M. tuberculosis Reprograms Hematopoietic Stem Cells to Limit Myelopoiesis and Impair Trained Immunity. *Cell* **183**, 752-770 e722 (2020). https://doi.org/10.1016/j.cell.2020.09.062

8 Kalam, H., Singh, K., Chauhan, K., Fontana, M. F. & Kumar, D. Alternate splicing of transcripts upon Mycobacterium tuberculosis infection impacts the expression of functional protein domains. *IUBMB Life* **70**, 845-854 (2018). https://doi.org/10.1002/iub.1887

9 Chauhan, K., Kalam, H., Dutt, R. & Kumar, D. RNA Splicing: A New Paradigm in Host-Pathogen Interactions. *J Mol Biol* **431**, 1565-1575 (2019). https://doi.org/10.1016/j.jmb.2019.03.001

10 Liu, Y. *et al.* Impact of Alternative Splicing on the Human Proteome. *Cell Rep* **20**, 1229-1241 (2017). https://doi.org/10.1016/j.celrep.2017.07.025

11 Kalam, H., Fontana, M. F. & Kumar, D. Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLoS Pathog* **13**, e1006236 (2017). https://doi.org/10.1371/journal.ppat.1006236

12 Zhang, X. O., Fu, Y., Mou, H., Xue, W. & Weng, Z. The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLoS Genet* **14**, e1007579 (2018). https://doi.org/10.1371/journal.pgen.1007579

13 Hug, N., Longman, D. & Caceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44**, 1483-1495 (2016). https://doi.org/10.1093/nar/gkw010

14 Pai, A. A. *et al.* Widespread Shortening of 3' Untranslated Regions and Increased Exon Inclusion Are Evolutionarily Conserved Features of Innate Immune Responses to Infection. *PLoS Genet* **12**, e1006338 (2016). https://doi.org/10.1371/journal.pgen.1006338

15 Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279-290 (2012). https://doi.org/10.1016/j.cell.2012.05.043

16 Pandya-Jones, A. *et al.* Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19**, 811-827 (2013). https://doi.org/10.1261/rna.039081.113

17 Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718 (2009). https://doi.org/10.1016/j.cell.2009.02.009

18 Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem* **89**, 359-388 (2020). https://doi.org/10.1146/annurev-biochem-091719-064225

19 Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcarcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* **57**, 7-22 (2015). https://doi.org/10.1016/j.molcel.2014.10.030

20 Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* **146**, 384-395 (2011). https://doi.org/10.1016/j.cell.2011.06.043

21 Karijolich, J. & Yu, Y. T. Spliceosomal snRNA modifications and their function. *RNA Biol* **7**, 192-204 (2010). https://doi.org/10.4161/rna.7.2.11207

22 Strasser, A., Dickmanns, A., Luhrmann, R. & Ficner, R. Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J* **24**, 2235-2243 (2005). https://doi.org/10.1038/sj.emboj.7600701

23 Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**, 291-323 (2015). https://doi.org/10.1146/annurev-biochem-060614-034316

24 Penn, B. H. *et al.* An Mtb-Human Protein-Protein Interaction Map Identifies a Switch between Host Antiviral and Antibacterial Responses. *Mol Cell* **71**, 637-648 e635 (2018). https://doi.org/10.1016/j.molcel.2018.07.010

25 Portal-Celhay, C. *et al.* Mycobacterium tuberculosis EsxH inhibits ESCRT-dependent CD4(+) T-cell activation. *Nat Microbiol* **2**, 16232 (2016). https://doi.org/10.1038/nmicrobiol.2016.232

26 Sharma, G., Upadhyay, S., Srilalitha, M., Nandicoori, V. K. & Khosla, S. The interaction of mycobacterial protein Rv2966c with host chromatin is mediated through non-CpG methylation and histone H3/H4 binding. *Nucleic Acids Res* **43**, 3922-3937 (2015). https://doi.org/10.1093/nar/gkv261

27 Mehra, A. *et al.* Mycobacterium tuberculosis type VII secreted effector EsxH targets host ESCRT to impair trafficking. *PLoS Pathog* **9**, e1003734 (2013). https://doi.org/10.1371/journal.ppat.1003734

28 de Souza, G. A., Leversen, N. A., Malen, H. & Wiker, H. G. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *J Proteomics* **75**, 502-510 (2011). https://doi.org/10.1016/j.jprot.2011.08.016

29 Malen, H., Berven, F. S., Fladmark, K. E. & Wiker, H. G. Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv. *Proteomics* **7**, 1702-1718 (2007). https://doi.org/10.1002/pmic.200600853

30 Yaseen, I., Kaur, P., Nandicoori, V. K. & Khosla, S. Mycobacteria modulate host epigenetic machinery by Rv1988 methylation of a non-tail arginine of histone H3. *Nat Commun* **6**, 8922 (2015). https://doi.org/10.1038/ncomms9922

31 Malakar, B. *et al.* Phosphorylation of CFP10 modulates Mycobacterium tuberculosis virulence. *mBio* **14**, e0123223 (2023). https://doi.org/10.1128/mbio.01232-23

32 Bonano, V. I., Oltean, S. & Garcia-Blanco, M. A. A protocol for imaging alternative splicing regulation in vivo using fluorescence reporters in transgenic mice. *Nat Protoc* **2**, 2166-2181 (2007). https://doi.org/10.1038/nprot.2007.292

33 Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637-650 (2017). https://doi.org/10.1038/nrm.2017.63

34 Oesterreich, F. C. *et al.* Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**, 372-381 (2016). https://doi.org/10.1016/j.cell.2016.02.045

35 Becker, D. *et al.* Nuclear Pre-snRNA Export Is an Essential Quality Assurance Mechanism for Functional Spliceosomes. *Cell Rep* **27**, 3199-3214 e3193 (2019). https://doi.org/10.1016/j.celrep.2019.05.031

36 Boudreault, S. *et al.* Reovirus mu2 protein modulates host cell alternative splicing by reducing protein levels of U5 snRNP core components. *Nucleic Acids Res* **50**, 5263-5281 (2022). https://doi.org/10.1093/nar/gkac272

37 van Kessel, J. C. & Hatfull, G. F. Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4**, 147-152 (2007). https://doi.org/10.1038/nmeth996

38 Del Rosario, R. C. H. *et al.* Histone acetylome-wide associations in immune cells from individuals with active Mycobacterium tuberculosis infection. *Nat Microbiol* **7**, 312-326 (2022). https://doi.org/10.1038/s41564-021-01049-w

39 Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**, 153-165 (2013). https://doi.org/10.1038/nrm3525

40 Bradley, R. K. & Anczukow, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* (2023). https://doi.org/10.1038/s41568-022-00541-7

41 Liao, K. C. & Garcia-Blanco, M. A. Role of Alternative Splicing in Regulating Host Response to Viral Infection. *Cells* **10** (2021). https://doi.org/10.3390/cells10071720

42 Wang, C. *et al.* Abnormal global alternative RNA splicing in COVID-19 patients. *PLoS Genet* **18**, e1010137 (2022). https://doi.org/10.1371/journal.pgen.1010137

43 Manet, E. *et al.* Modulation of alternative splicing during early infection of human primary B lymphocytes with Epstein-Barr virus (EBV): a novel function for the viral EBNA-LP protein. *Nucleic Acids Res* **49**, 10657-10676 (2021). https://doi.org/10.1093/nar/gkab787

44 Thompson, M. G. *et al.* Viral-induced alternative splicing of host genes promotes influenza replication. *Elife* **9** (2020). https://doi.org/10.7554/eLife.55500

45 Louadi, Z. *et al.* DIGGER: exploring the functional role of alternative splicing in protein interactions. *Nucleic Acids Res* **49**, D309-D318 (2021). https://doi.org/10.1093/nar/gkaa768

46 Neuenkirchen, N. *et al.* Reconstitution of the human U snRNP assembly machinery reveals stepwise Sm protein organization. *EMBO J* **34**, 1925-1941 (2015). https://doi.org/10.15252/embj.201490350

47 Shaw, D. J., Eggleton, P. & Young, P. J. Joining the dots: production, processing and targeting of U snRNP to nuclear bodies. *Biochim Biophys Acta* **1783**, 2137-2144 (2008). https://doi.org/10.1016/j.bbamcr.2008.07.025

48 Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I. W. & Dreyfuss, G. The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol Cell Biol* **22**, 6533-6541 (2002). https://doi.org/10.1128/MCB.22.18.6533-6541.2002

49 Gubitz, A. K., Feng, W. & Dreyfuss, G. The SMN complex. *Exp Cell Res* **296**, 51-56 (2004). https://doi.org/10.1016/j.yexcr.2004.03.022

50 Pellizzoni, L. Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep* **8**, 340-345 (2007). https://doi.org/10.1038/sj.embor.7400941

51 Yong, J., Kasim, M., Bachorik, J. L., Wan, L. & Dreyfuss, G. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell* **38**, 551-562 (2010). https://doi.org/10.1016/j.molcel.2010.03.014

52 Schlaen, R. G. *et al.* The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proc Natl Acad Sci U S A* **112**, 9382-9387 (2015). https://doi.org/10.1073/pnas.1504541112

53 Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585-600 (2008). https://doi.org/10.1016/j.cell.2008.03.031

54 Cartegni, L., Hastings, M. L., Calarco, J. A., de Stanchina, E. & Krainer, A. R. Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am J Hum Genet* **78**, 63-77 (2006). https://doi.org/10.1086/498853

55 De Maio, F. A. *et al.* The Dengue Virus NS5 Protein Intrudes in the Cellular Spliceosome and Modulates Splicing. *PLoS Pathog* **12**, e1005841 (2016). https://doi.org/10.1371/journal.ppat.1005841

56 Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802-813 (2008). https://doi.org/10.1261/rna.876308

57 Grosso, A. R. *et al.* Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res* **36**, 4823-4832 (2008). https://doi.org/10.1093/nar/gkn463

58 Artemyeva-Isman, O. V. & Porter, A. C. G. U5 snRNA Interactions With Exons Ensure Splicing Precision. *Front Genet* **12**, 676971 (2021). https://doi.org/10.3389/fgene.2021.676971

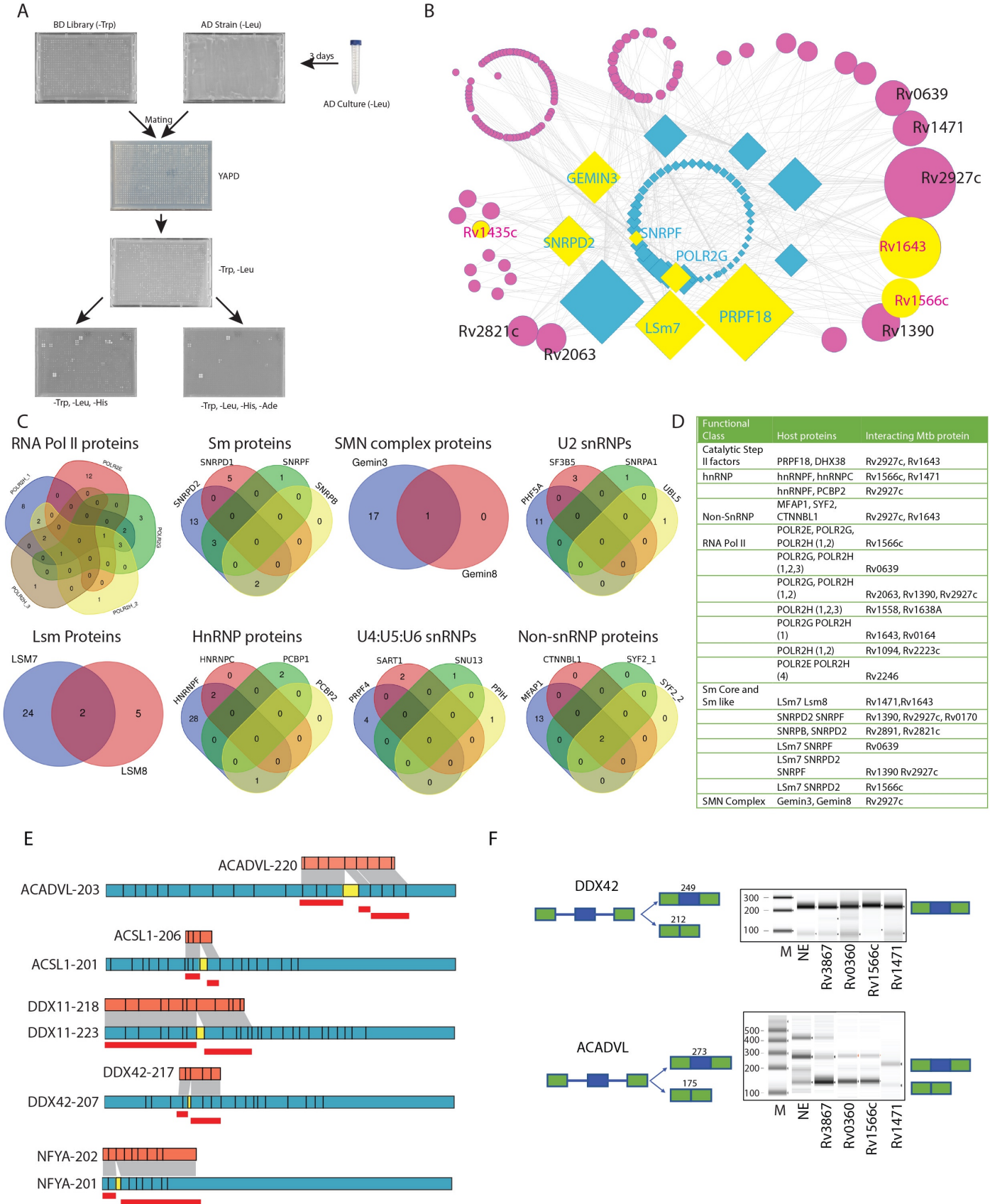
59 Jia, Y., Mu, J. C. & Ackerman, S. L. Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. *Cell* **148**, 296-308 (2012). https://doi.org/10.1016/j.cell.2011.11.057

60 Ramakrishnan, L. Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol* **12**, 352-366 (2012). https://doi.org/10.1038/nri3211

61 Simmons, J. D. *et al.* Differentially expressed transcript isoforms associate with resistance to tuberculin skin test and interferon gamma release assay conversion. *PLoS One* **18**, e0284498 (2023). https://doi.org/10.1371/journal.pone.0284498

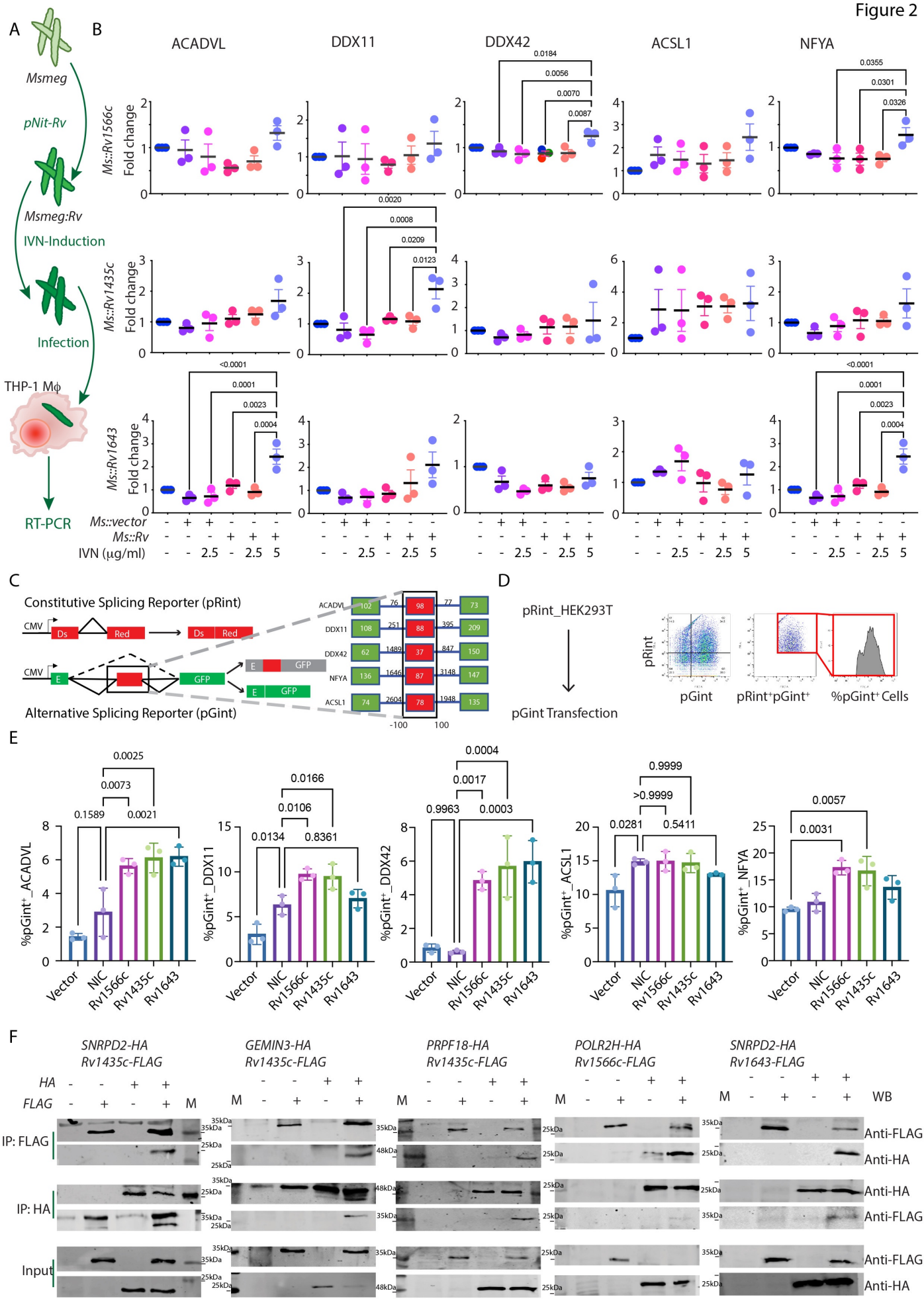
62 Mehdiratta, K. *et al.* Kupyaphores are zinc homeostatic metallophores required for colonization of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **119** (2022). https://doi.org/10.1073/pnas.2110293119

Figure 1



Functional Class	Host proteins	Interacting Mtb protein
Catalytic Step II factors	PRPF18, DHX38	Rv2927c, Rv1643
hnRNP	hnRNPF, hnRNPC	Rv1566c, Rv1471
	hnRNPF, PCBP2	Rv2927c
Non-SnRNP	MFAP1, SYF2, CTNNBL1	Rv2927c, Rv1643
RNA Pol II	POLR2E, POLR2G, POLR2H (1,2)	Rv1566c
	POLR2G, POLR2H (1,2,3)	Rv0639
	POLR2G, POLR2H (1,2)	Rv2063, Rv1390, Rv2927c
	POLR2H (1,2,3)	Rv1558, Rv1638A
	POLR2G POLR2H (1)	Rv1643, Rv0164
	POLR2H (1,2)	Rv1094, Rv2223c
	POLR2E POLR2H (4)	Rv2246
Sm Core and Sm like	LSm7 Lsm8	Rv1471,Rv1643
	SNRPD2 SNRPF	Rv1390, Rv2927c, Rv0170
	SNRPB, SNRPD2	Rv2891, Rv2821c
	LSm7 SNRPF	Rv0639
	LSm7 SNRPD2 SNRPF	Rv1390 Rv2927c
	LSm7 SNRPD2	Rv1566c
SMN Complex	Gemin3, Gemin8	Rv2927c

Figure 2


A
Msmeg
pNit-Rv
Msmeg:Rv
IVN-Induction
Infection
THP-1 Mϕ
RT-PCR
B
ACADVL
DDX11
DDX42
ACSL1
NFYA
Ms::Rv1566c Fold change
Ms::Rv1435c Fold change
Ms::Rv1643 Fold change
Ms::vector
Ms::Rv
IVN (μg/ml)
C
Constitutive Splicing Reporter (pRint)
CMV
Ds
Red
E
GFP
Alternative Splicing Reporter (pGint)
ACADVL
DDX11
DDX42
NFYA
ACSL1
-100
100
D
pRint_HEK293T
pGint Transfection
pRint
pGint
pRint+pGint+
%pGint+ Cells
E
%pGint+_ACADVL
%pGint+_DDX11
%pGint+_DDX42
%pGint+_ACSL1
%pGint+_NFYA
Vector
NIC
Rv1566c
Rv1435c
Rv1643
F
SNRPD2-HA Rv1435c-FLAG
GEMIN3-HA Rv1435c-FLAG
PRPF18-HA Rv1435c-FLAG
POLR2H-HA Rv1566c-FLAG
SNRPD2-HA Rv1643-FLAG
HA
FLAG
M
WB
IP: FLAG
IP: HA
Input
Anti-FLAG
Anti-HA

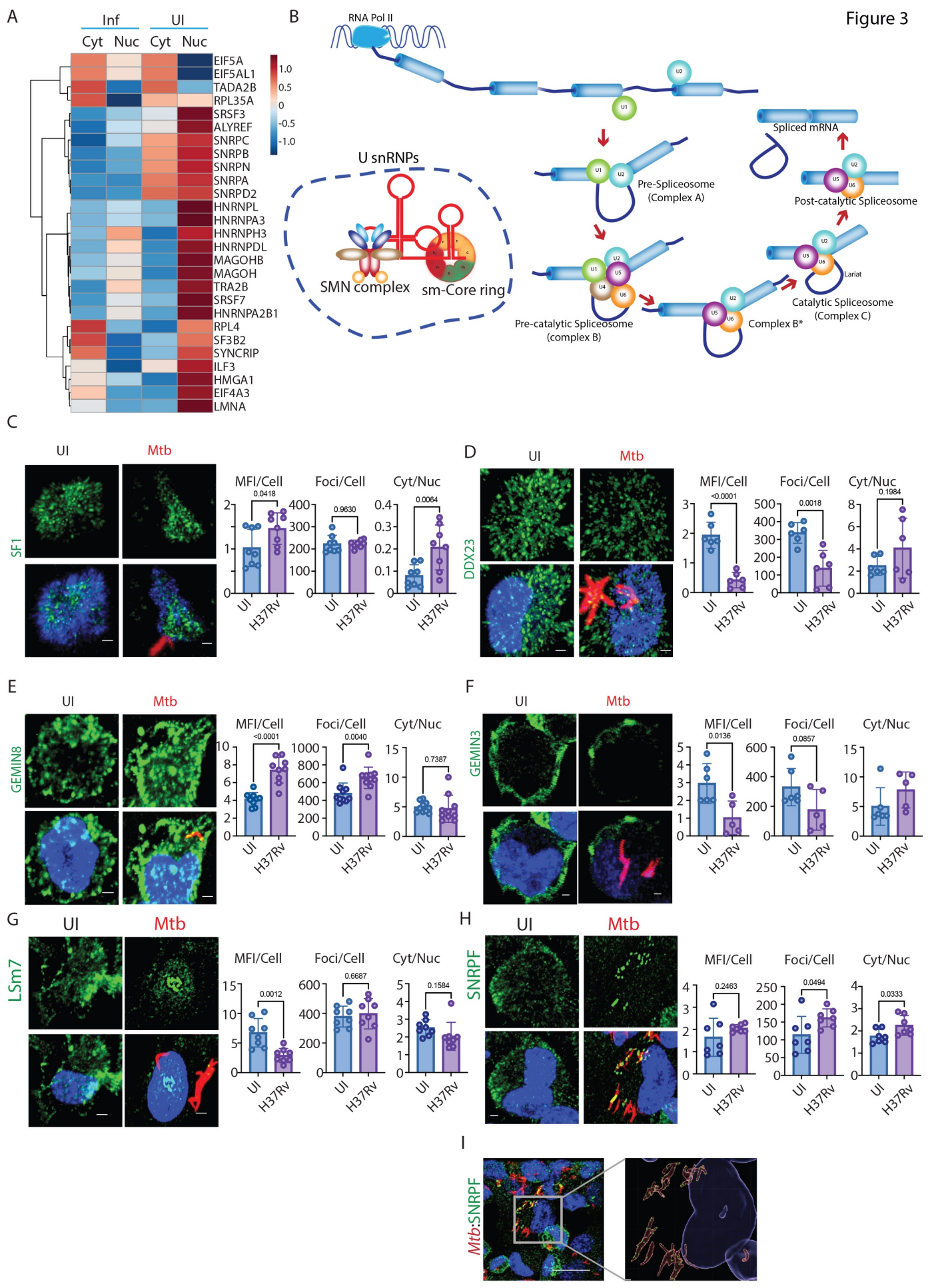
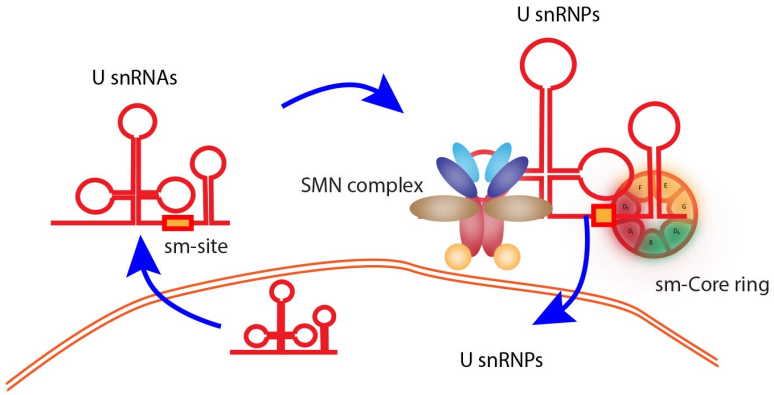


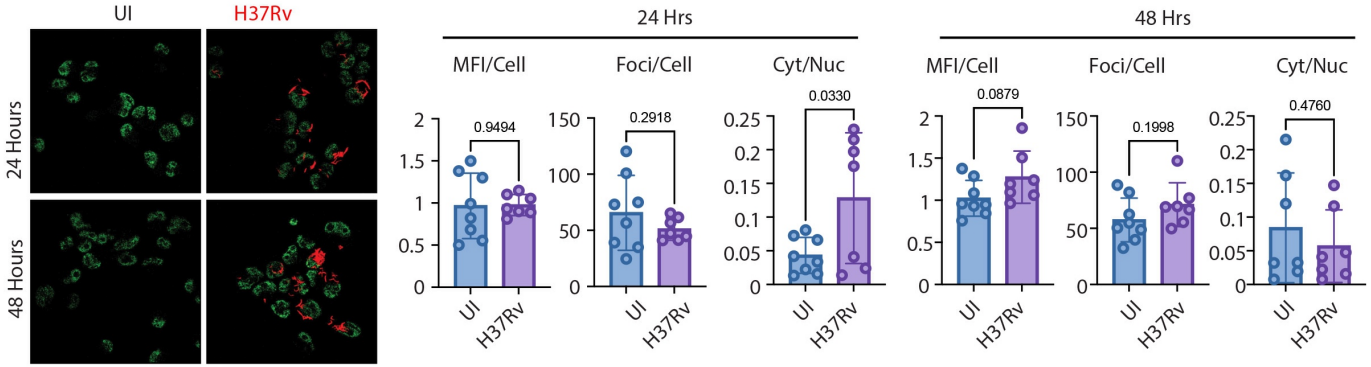
Figure 3
A
Inf
UI
Cyt
Nuc
Cyt
Nuc
EIF5A
EIF5AL1
TADA2B
RPL35A
SRSF3
ALYREF
SNRPC
SNRPB
SNRPN
SNRPA
SNRPD2
HNRNPL
HNRNPA3
HNRNPH3
HNRNPDL
MAGOHB
MAGOH
TRA2B
SRSF7
HNRNPA2B1
RPL4
SF3B2
SYNCRIP
ILF3
HMGA1
EIF4A3
LMNA
1.0
0.5
0
-0.5
-1.0
B
RNA Pol II
U snRNPs
SMN complex
sm-Core ring
Pre-Spliceosome
(Complex A)
Pre-catalytic Spliceosome
(complex B)
Complex B*
Catalytic Spliceosome
(Complex C)
Lariat
Post-catalytic Spliceosome
Spliced mRNA
C
UI
Mtb
SF1
MFI/Cell
Foci/Cell
Cyt/Nuc
0.0418
0.9630
0.0064
D
DDX23
<0.0001
0.0018
0.1984
E
GEMIN8
<0.0001
0.0040
0.7387
F
GEMIN3
0.0136
0.0857
G
LSm7
0.0012
0.6687
0.1584
H
SNRPF
0.2463
0.0494
0.0333
I
Mtb:SNRPF
H37Rv

Figure 4

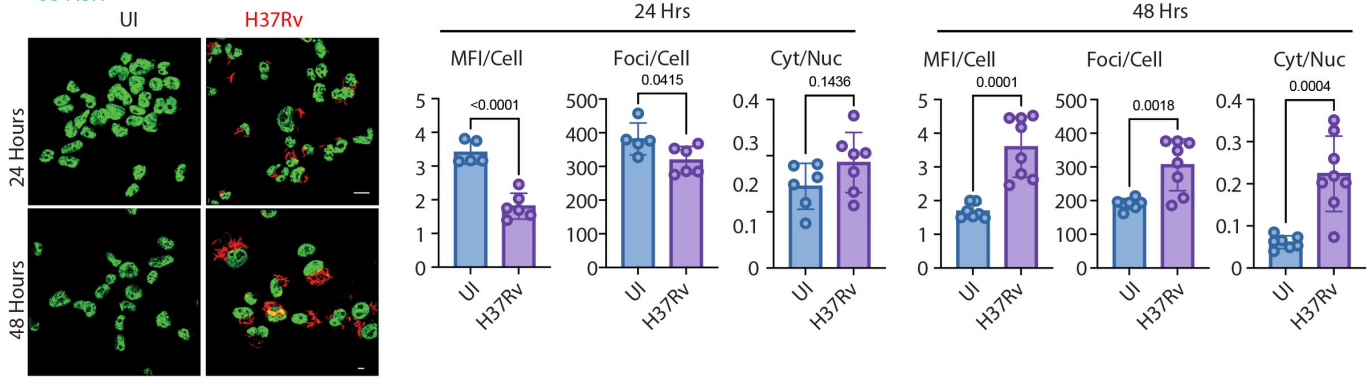
A



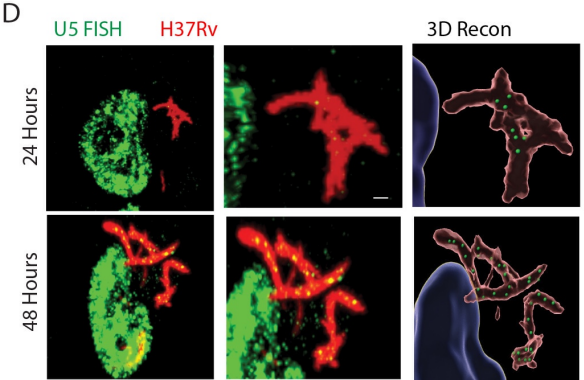
B



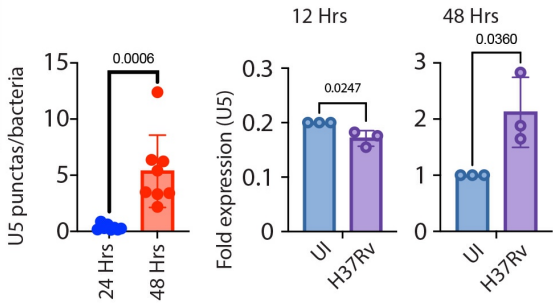
C



D



E



F

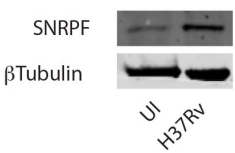


Figure 5

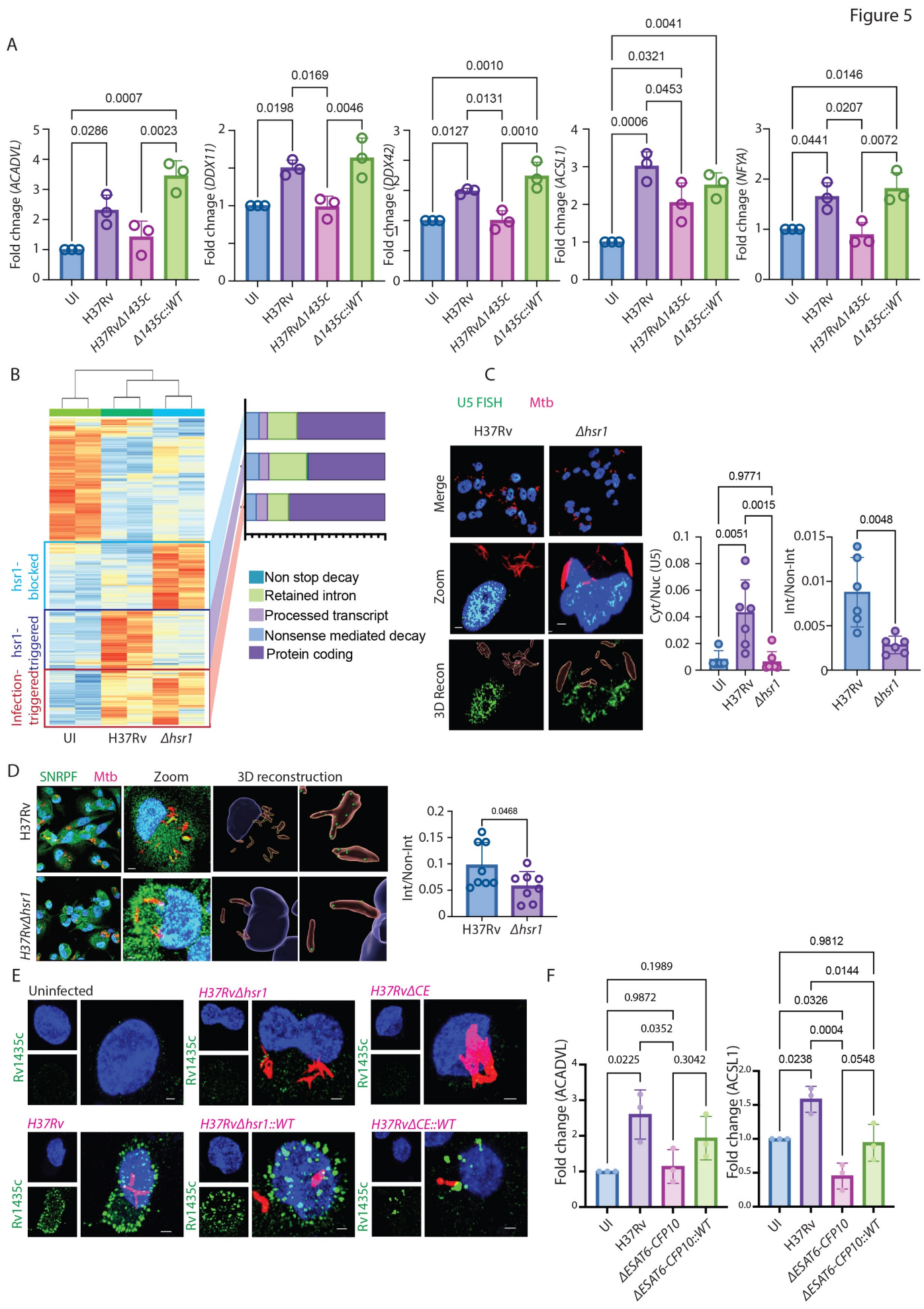
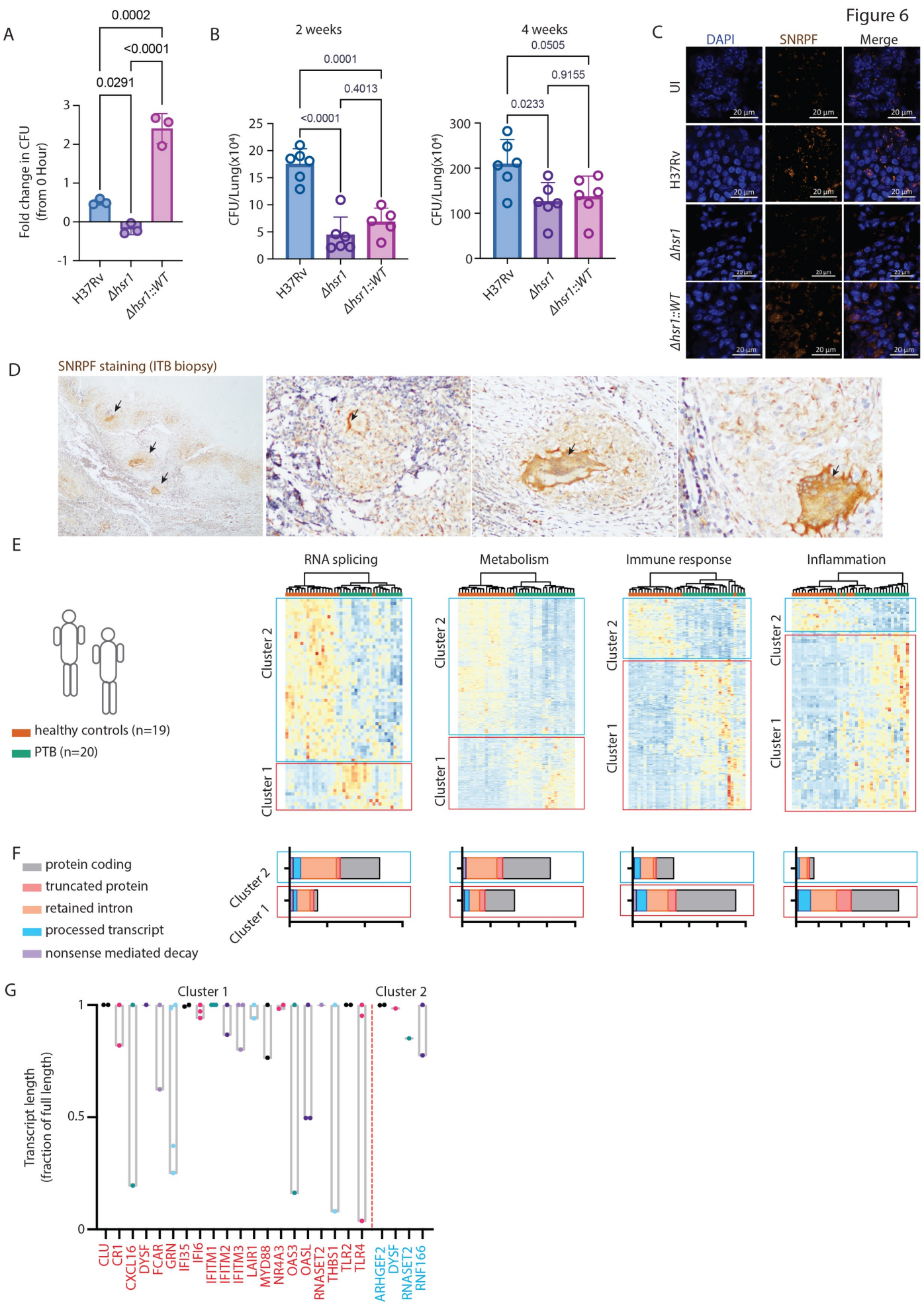

A
Fold chnage (ACADVL)
Fold chnage (DDX11)
Fold chnage (DDX42)
Fold chnage (ACSL1)
Fold chnage (NFYA)
0.0007
0.0286
0.0023
0.0169
0.0198
0.0046
0.0010
0.0127
0.0131
0.0010
0.0041
0.0321
0.0453
0.0006
0.0146
0.0207
0.0441
0.0072
UI
H37Rv
H37RvΔ1435c
Δ1435c::WT
B
hsr1-blocked
hsr1-triggered
Infection-triggered
UI
H37Rv
Δhsr1
Non stop decay
Retained intron
Processed transcript
Nonsense mediated decay
Protein coding
C
U5 FISH
Mtb
H37Rv
Δhsr1
Merge
Zoom
3D Recon
Cyt/Nuc (U5)
Int/Non-Int
0.9771
0.0015
0.0051
0.0048
D
SNRPF
Mtb
Zoom
3D reconstruction
H37Rv
H37RvΔhsr1
Int/Non-Int
0.0468
E
Uninfected
H37RvΔhsr1
H37RvΔCE
H37Rv
H37RvΔhsr1::WT
H37RvΔCE::WT
Rv1435c
F
Fold change (ACADVL)
Fold change (ACSL1)
0.1989
0.9872
0.0352
0.0225
0.3042
0.9812
0.0144
0.0326
0.0004
0.0238
0.0548
UI
H37Rv
ΔESAT6-CFP10
ΔESAT6-CFP10::WT

Figure 6


A
0.0002
<0.0001
0.0291
Fold change in CFU (from 0 Hour)
H37Rv
Δhsr1
Δhsr1::WT
B
2 weeks
0.0001
0.4013
<0.0001
CFU/Lung(x10⁴)
4 weeks
0.0505
0.9155
0.0233
C
DAPI
SNRPF
Merge
UI
H37Rv
Δhsr1
Δhsr1::WT
20 µm
D
SNRPF staining (ITB biopsy)
E
healthy controls (n=19)
PTB (n=20)
RNA splicing
Metabolism
Immune response
Inflammation
Cluster 2
Cluster 1
F
protein coding
truncated protein
retained intron
processed transcript
nonsense mediated decay
G
Cluster 1
Cluster 2
Transcript length (fraction of full length)
CLU
CR1
CXCL16
DYSF
FCAR
GRN
IFI35
IFI6
IFITM1
IFITM2
IFITM3
LAIR1
MYD88
NR4A3
OAS3
OASL
RNASET2
THBS1
TLR2
TLR4
ARHGEF2
DYSF
RNASET2
RNF166

Supplementary information for the manuscript:

A *Mycobacterium tuberculosis* secreted virulence factor disrupts host snRNP biogenesis

Chauhan, K et al.

Supplementary methods
Supplementary Figures and Legends to Figures S1-S9
Supplementary tables S1-S9

Ethics statement: Studies on human samples were approved by the IEC of AIIMS, New Delhi (Ref no. IEC-304/02-06-2017). For animal experiments, approval was obtained from the Institutional Animal Ethics Committee, ICGEB, New Delhi (ICGEB/IAEC/22092022/CI-28).

Biological materials: Tissue sections from Intestinal TB biopsy samples used in the study were obtained from the Department of Pathology, AIIMS, New Delhi. Access to these materials is subject to institutional guidelines. **Clinical diagnosis of ITB subjects:** The diagnoses depend on clinical features of nocturnal fever, weight loss, diarrhoea, and radiological suspicion of short-segment bowel thickenings, narrowed ileocaecal region, significant mesenteric lymphadenopathy, and/or necrosis. In such cases, an intestinal biopsy is taken, or a segment of the intestine is resected if there is a feature of intestinal obstruction due to fibrosis. All our cases underwent intestinal resection. Histological examination is critical in such cases, and demonstration of large, confluent, multiple, necrotic granulomas with multinucleated giant cells in deeper layers of the intestine are suggestive of ITB, as we have demonstrated in our cases. ZN stain has a very low sensitivity, varying from 5-10%. Such cases are confirmed by either the Gene-Xpert test, Mycobacterial culture, or where such tests are not possible, by ATT trial followed by re-evaluation at two months and six months. All our cases underwent either a *Mycobacterial* culture or a gene-Xpert test.

Antibodies and reagents

The antibodies used in the immunofluorescence experiments were as follows: POLR2G (#NBP2-47329), RBP8 (#NBP2-93995), SNRPD2 (#NBP2-95236), SNRPF(#NBP1-57463), LSm7 (#NBP1-92084), Gemin3 (#NBP2-52454) and Gemin8(#NBP1-83925) from Novus biologicals. SF1(#A303-213A), PRPF4B(#A301-665A), SNRP200(#A303-453A), DDX23(#A300-693A) from Bethyl Laboratories were used. PKH26 red and PKH67 green fluorescent labelling dyes were used from Sigma-Aldrich. The spliceosome complex pull-

down was done using an anti-SNRPB antibody (Y-12, MA513449). The antibodies used in immune-blotting experiments were as follows DHX8 (bethyl laboratories; A300-624A), tubulin (Abcam plc; ab6046), histone (Abcam plc; ab18521), IRDye 800CW Goat anti-mouse IgG secondary antibody (LI-COR Biosciences)and IRDye 680RD goat anti-rabbit IgG secondary antibody(LI-COR Biosciences). The co-immunoprecipitation was done using EZView Red Anti-HA affinity gel (E6779), Anti-FLAG M2 Affinity gel (A2220) and immunoblots were detected using Anti-HA antibody, Mouse monoclonal (H3663) and Monoclonal Anti-FLAG M2 antibody mouse (F1804) from Sigma-Aldrich. Bio-11-UTP (#AM8450) from Invitrogen was used for *in-vitro* splicing assays.

Cell culture

<u>*Mammalian cells*</u>: THP1 monocytes (ATCC) were maintained in RPMI 1640 media (Cell Clone), and HEK293T cells (ATCC) were maintained in DMEM media (Cell Clone). Cells were grown at 37°C with a 5% CO2 atmosphere. All the media was supplemented with 10% fetal bovine serum. HEK293T cells were transfected with plasmids at a 30-50% confluency using jetPRIME buffer (Polyplus #712-60) as per manufacturer protocol. The siRNA was transfected using the Dharmafect reagent. HEK293T constitutively expressing PRINT plasmid was selected in the presence of blasticidin (Gibco) 20 μg/ml concentration.

<u>*Bacterial Cells*</u>: *M. tuberculosis H37Rv*, *M. smegmatis* mc^2155 and mutant strains were grown at 37°C to log phase in Middlebrook 7H9 media with 0.05% Tween80, 10% ADC Enrichment, and 0.2% glycerol. Plasmids were selected with 50 mg/ml kanamycin or hygromycin depending upon the resistance marker.

Infection

THP-1 cells were activated into macrophages by treatment with 32 nM PMA (Sigma) for 24 hours, followed by another 24 hours for recovery. The THP-1-derived macrophages were infected with *Mtb* strains at an MOI of 10 for 4 hours, followed by 2 hours of amikacin (200 μg/μl) treatment. *M. smegmatis* infection was performed at an MOI of 10 for 1 hour, followed by 2 hours of amikacin treatment. Amikacin treatment clears all the extracellular bacteria, which are washed following the treatment, and this time point is considered as 0 hours of infection throughout the manuscript.

63 **RNA isolation and qPCR**

64 RNA was isolated at indicated time points using MDI mammalian Total RNA isolation kit
65 (MTRK250) as per the manufacturer's protocol. The RNA was quantified using a nanodrop
66 spectrophotometer. 1000ng RNA was used to make cDNA using the Iscript cDNA synthesis
67 kit (BioRad #170-8891), and 2 µl of 1:5 diluted cDNA was used to set qPCR on the BioRad
68 CFX384 well plate qPCR machine. Primers were custom synthesized from Sigma Aldrich
69 Chemicals Ltd.

70 **Transformation of yeast cell**

71 The yeast competent cells were prepared by harvesting yeast cells by centrifugation at 400 rpm
72 for 5 min at 25 °C. The supernatant was discarded, and the pellet was washed and resuspended
73 in filter-sterilized SORB buffer (100 mM Lithium acetate, 100 mM Tris HCl, pH 8.0, 1 mM
74 EDTA, pH 8.0, 1 M sorbitol pH 8.0) to which denatured salmon sperm DNA was added (10
75 mg/ml). The bait (pGBDKT7 vector) plasmids were mixed gently with the Y187 competent
76 cells. Then PEG mixture was added and incubated at 30 °C for 30 min, 42 °C for 30 min and
77 4 °C for 30 min in a thermocycler machine. Tubes were taken out and plated on 15 cm diameter
78 SC-TRP plates (Cat No. Clonetech 630413). Plates were incubated at 30°C for 2 to 7 days. A
79 similar protocol was used for the prey (pGADT7 vector) plasmid in the AH109 yeast strain on
80 SC-Leu medium (Cat No. Clonetech 630414).

81 **Y2H assay between bait and prey libraries**

82 The bait library was prepared by plating the Y187 yeast strain (bait) transformed with Mtb
83 proteins in 384 well patterns on the Singer Robot Plate. The AH109 (prey) yeast strain
84 transformed with host spliceosome protein was grown in SC-leu broth medium and equally
85 plated on a robotic compatible square plate. Using 384 pinned robotic pad bait libraries were
86 patched on another YAPD plate. Each plate was then patched with individual cells containing
87 prey plasmid and mixed well using the robotic system. This step was performed 4 times as four
88 independent mating steps, thereby making it a total of 1536 spots per plate. The plate containing
89 mixed colonies was kept for mating of both strains at 30 °C for 36 hours. After 36 hours these
90 colonies were selected on SC-Trp-Leu (Cat No. Clonetech 630417) plates for the selection of
91 those cells which contain both plasmids (bait as well as prey). After 72 hours these plates were
92 replicated on SC-Trp-Leu-His (3 drops-out, Clonejet 630419) and SC-Trp-Leu-His-Ade (4
93 drop-outs, Clonetech 630428) for the screening of interaction between the protein which was

94 fused with bait and prey respectively. The colonies that appear on both 3-DO and 4-DO plates
95 represented very specific and strong interactions between the bait and prey proteins.

96 **Co-immunoprecipitation and Immunoblotting**

97 The cells were lysed using a non-denaturing IP lysis Buffer (20 mM Tris-Cl, 137 mM NaCl, 2
98 mM EDTA, 10% glycerol, 1% NP-40, pH 8) and centrifuged at 14000 rpm for 10 min at 4 °C.
99 The supernatant was precleared using IgG beads, and 300 µg of protein was used for IP. The
100 IP was performed using FLAG-tagged beads (Sigma) and HA-tagged beads (Sigma) at 4 °C
101 for 4 hours, followed by three washes with IP buffer. The beads were then boiled in 2X SDS
102 loading dye. The protein samples were subjected to SDS PAGE followed by transfer on
103 nitrocellulose membrane (BioRad) using BioRad Turbo transfer blot. The blocking was
104 performed by Odessey blocking buffer (LI-COR Biosciences) in 1:1 dilution with 1X PBS at
105 room temperature. The blots were immunoblotted with primary antibodies, followed by IR-
106 tagged secondary antibodies (LI-COR Biosciences). Blot imaging was done using Odessey IR
107 Imaging System (LI-COR Biosciences).

108 **Immunofluorescence**

109 For immunofluorescence experiments, ~0.2x 10^6 cells were plated on 12 mm coverslips per
110 well in a 24- well plate. Cells were infected with m-cherry expressing *Mtb*, and other strains
111 were labelled with PKH67 fluorescent dye (Sigma). The cells were fixed using 4% v/v
112 paraformaldehyde for 20 min at room temperature for indicated time points. The cells were
113 incubated with blocking buffer (0.02% Triton-X100, 3% w/v BSA in 1XPBS) for 1 hour. The
114 cells were washed thrice with 1X PBST and incubated with primary antibodies for 3 hours.
115 The cells were washed thrice with 1XPBST and incubated with secondary Alexa Flour (488,
116 555, 590 or 670nm) tagged antibodies (Invitrogen). The nucleus was stained using DAPI (300
117 nM) for 10 min. The cells were mounted on slides with Pro-Long Gold antifade reagents (Life
118 Technologies). All the incubation and staining were done at room temperature. Images were
119 acquired by NIS-Elements software using the Nikon A1R laser scanning confocal microscope
120 equipped with a Nikon Plan Apo VC ×20, NA 0.75, and Plan Apo VC ×100 oil, NA 1.40
121 objectives were used. Serial confocal sections, 0.5 µm thick, were acquired with a z-stack
122 spanning 10–15 µm to form a composite image. Images were analysed using Imaris 9.6.6
123 imaging software using several tools like deconvolution, 3D-reconstruction and shortest
124 distance calculation amongst others, NIS-Elements and image J software.

125 **Fluorescent reporter splicing assay using flow cytometry**

126 For Flow cytometry experiments, ~0.5 x 10^6 HEK293T cells constitutively expressing pRINT
127 vector (Addgene plasmid # 24218) were plated on 12 well plates. These were co-transfected
128 with 800ng of modified pGINT vector (Addgene plasmid # 24217) with exon skipping cassette
129 and *Mtb* protein-expressing vectors at a ratio of 1:1 using jetPRIME buffer. The cells were
130 fixed using 4% v/v paraformaldehyde in 1X PBS for 20 minutes at room temperature. Then,
131 the cells were scraped and resuspended in 1X PBS. The data were acquired using BD
132 FACSDiva acquisition software in BD FACS Fortussa cytometer (BD).

133 **Spliceosome complex Co-IP**

134 For the pulldown experiment, ~200 million THP-1 derived macrophages were infected with
135 H37Rv at an MOI of 10 and samples were processed 48-hour post-infection. The nuclear and
136 cytoplasmic extract was prepared using the Active Motif Nuclear Complex Co-IP kit
137 (catalogue No.54001) per the manufacturer's guidelines. The nuclear extract and cytoplasmic
138 extract were passed through a 0.2-micron filter. 500 µg of nuclear extract and cytoplasmic
139 extracts were used for pulldown using 10 µg Y-12 antibody (MA5-13449, Thermo Fisher) or
140 IgG-control (Sigma) antibody for each condition. The Y-12 antibody was crosslinked using
141 DMP (20 mM) to Protein A-Sepharose beads for 1 hour (beads were washed prior with 0.2 M
142 sodium borate, pH 9.0) followed by 0.2 M ethanolamine incubation for 1 hour at room
143 temperature. The Nuclear and Cytosolic extracts were then incubated with crosslinked beads
144 overnight at 4 °C. The beads were washed thrice with wash buffer (0.025 M Tris, 0.15 M NaCl,
145 pH 7.2). The protein was finally eluted in 50 µl of an elution buffer containing primary amine
146 pH 2.8. To neutralize the low pH of the elution buffer, 1 M Tris (pH 9.5) was added to the
147 collection tube.
148

149 **Mass Spectrometric Data analysis for pulled down spliceosome complex**
150 *Sample Preparation*- An equal volume of eluted protein sample (50ul) was used for digestion
151 and reduced with 5 mM TCEP and further alkylated with 50 mM iodoacetamide and then
152 digested with Trypsin (1:50, Trypsin/lysate ratio) for 16 h at 37 °C. Digests were cleaned using
153 a C18 silica cartridge to remove the salt and dried using a speed vac. The dried pellet was
154 resuspended in buffer A (2% acetonitrile, 0.1% formic acid). An equal amount of peptide (1ug)
155 was injected into the analytical column. *Mass Spectrometric Analysis of Peptide Mixtures*- All
156 the experiments were performed using EASY-nLC 1000 system (Thermo Fisher Scientific)
157 coupled to Thermo Fisher-QExactive plus equipped with nano electrospray ion source. 1ug
158 peptide was loaded on C18 column 50 cm, 3.0µm Easy-spray column (Thermo Fisher

159 Scientific) at column heating temp 40 C. Peptides were eluted with a 0–40% gradient of buffer
160 B (80% acetonitrile, 0.1% formic acid) at a flow rate of 300 nl/min) and injected for MS
161 analysis. LC gradients were run for 90 minutes MS1 spectra were acquired in the Orbitrap
162 (Max IT = 50ms, AGC target = 3e6; R=70K, mass range = 350-1750; Profile data). Dynamic
163 exclusion was employed for 30s excluding all charge states for a given precursor. MS2 spectra
164 were collected for top 15 peptides. MS2 (Max IT= 100ms, R= 17.5K, AGC target 1e5). *Data*
165 *Processing*- All samples were processed and RAW files generated were analysed with
166 Proteome Discoverer (v2.2) against the Uniprot Human database. For dual Sequest and
167 Amanda search, the precursor and fragment mass tolerances were set at 10 ppm and 0.5 Da,
168 respectively. The protease used to generate peptides, i.e. enzyme specificity was set for
169 trypsin/P (cleavage at the C terminus of "K/R: unless followed by "P"). Carbamidomethyl on
170 cysteine as fixed modification and oxidation of methionine, N-terminal acetylation were
171 considered as variable modifications for database search. Both peptide spectrum match and
172 protein false discovery rate were set to 0.01 FDR. Only statistically significant proteins (Using
173 ANNOVA p-value cut off 0.05) that were found in all three biological replicates across and in
174 all experimental groups were used to generate heatmap and for further analysis.

175 **Splicing competent nuclear extract preparation**

176 For the *in-vitro* splicing assay, the splicing-competent nuclear extract was prepared from THP-
177 1 macrophages. The cells were gently scraped, washed with 1X PBS and centrifuged at 1000
178 rpm for 10 min. The pellet was washed with 5X PCV (packed cell volume) of hypotonic buffer
179 (10 mM HEPES, 1.5 mM MgOAc, 10 mM KOAc, 0.5 mM DTT, 0.2 mM PMSF) and
180 centrifuged at 1850 g for 10 min. The cell pellet was incubated in 3X PCV of hypotonic buffer
181 on ice for 60 min, then douncing for 50-150 plunges. Cell lysis was confirmed using trypan
182 blue, and the lysate was centrifuged at 3300 g for 15 min at 4°C. The supernatant was stored
183 as cytoplasmic extract and the nuclear pellet was resuspended in 0.5X PNV (packed nuclear
184 volume) of Low salt buffer (20 mM HEPES, 1.5 mM MgOAc, 0.02 mM KOAc, 0.2 mM
185 EDTA, 0.5 mM DTT, 0.2 mM PMSF, 25% glycerol). An equal volume of high salt buffer (20
186 mM HEPES, 1.5 mM MgOAc, 1.2 M KOAc, 0.5 mM DTT, 0.2 mM PMSF, 25% glycerol)
187 was added in a drop-wise manner while continuous stirring on ice for 30 minutes. It was then
188 centrifuged at 25000g for 30 min. The supernatant was collected, aliquoted and stored at -80
189 °C. These aliquots were used for *in-vitro* splicing reactions. Preparation of splicing-competent
190 extract was confirmed by immunoblotting for Histone H3 and tubulin.

191 ***In-vitro* splicing assay**

192 Unique exons from the alternatively spliced isoforms were selected for cloning the minigenes.
193 The entire native intronic sequence was amplified from THP-1 genomic DNA and cloned in
194 the pcDNA3.1 vector under the T7 promoter. These minigenes were used for T7 run-off
195 transcription in the presence of Biotin-UTP using MEGAscript® T7 Kit as per the
196 manufacturer's protocol (Thermo Fisher). For the splicing reaction, one µg of biotinylated *in*
197 *vitro* transcribed mRNA was added to the 50 µg of nuclear extracts in splicing buffer (Creatine
198 Phosphate 20 mM, 3.2 mM MgOAc, 1 mM DTT, 72 mM KOAc, 12 mM HEPES, 1 mM ATP).
199 The splicing reaction was incubated at 30°C. The biotinylated RNA was isolated using
200 Dynabeads™ M-270 Streptavidin (65305), followed by proteinase K digestion from the
201 splicing reaction using a Qiagen RNA isolation kit. The primers designed from the flanking
202 exons were used to perform a one-step RT-PCR reaction (Qiagen). Splicing events were
203 observed by running the PCR products on Agilent Tape Station 3000. A list of all primers used
204 in the study is provided in Table S9.

205 **RNA FISH**

206 RNA-fluorescence *in situ* hybridization (RNA FISH) was performed using previously
207 described snRNA probes [1]. The sequences of the oligonucleotides used for FISH were as
208 follows: U2 snRNA, 5′-GAA CAG ATA CTA CAC TTG ATC TTA GCC AAA AGG CCG
209 AGA AGC-3′; U5 snRNA, 5′-CTC TCC ACG GAA ATC TTT AGT AAA AGG CGA AAG
210 ATT TAT ACG ATT TGA AGA G-3′; The U2 snRNA and U5 snRNA probes were labelled
211 with Alexa Fluor 488 and Alexa Flour 647 (Invitrogen) respectively at their 5' end. Next, the
212 cells were fixed with 4% v/v paraformaldehyde for 20 min at room temperature. Next, cells
213 were permeabilized and incubated with RNA-FISH probes for three hours at 40 °C by using
214 the ViewRNA™ ISH Cell Assay Kit system (Thermo Fischer). Subsequently, cells were
215 immune-stained and mounted using a Pro-Long gold antifade (Life technologies).

216 **Bacterial transformations**

217 For the generation of *M. smegmatis* mutants expressing *H37Rv* proteins, 50 ml of an early
218 exponential phase culture of *M. smegmatis* mc^2155 was washed twice and concentrated ice-
219 cold 10% ultrapure glycerol (Invitrogen). Concentrated cells were electroporated with
220 approximately 10 µg of pNit plasmid. Electroporation was carried out at 2500 V, 25 µF, and
221 1000 Ω in the Biorad gene pulsar. Electroporated cells were allowed to recover at 37 °C in 7H9

222 Middlebrook broth and plated on 7H11 supplemented with 10% OADC and kanamycin (25
223 µg/ml; Sigma).

224 ***Mtb* survival assay**

225 For CFU experiments, ~0.01 million THP-1 macrophages were seeded in a 96-well plate and
226 *Mtb* infection was done at an MOI of 10. The lysis was done by incubating with 0.06% SDS
227 for 5 min at indicated time points. The lysed cells were diluted to 1:10 and 1:50 in 7H9 media
228 and 10 µl of the dilutions were plated in a drop-wise fashion on a 7H11 plate supplemented
229 with 10% OADC. The plates were incubated for approximately around 3 weeks for colony
230 enumeration at 37 °C.

231 **Expression and purification of recombinant proteins**

232 The Mtb proteins were cloned in the pEZY19 vector and transformed into E. coli BL21(DE3)
233 competent. Protein expression was induced with 0.1 mM isopropyl b-D-1-
234 thiogalactopyranoside (IPTG) for 16-19 hours. Collected frozen cells were lysed in a lysis
235 Buffer (50 mM Tris-HCl pH. 8.0 containing 300 mM NaCl,10% Glycerol, 1mm DTT, 1mm
236 PMSF,1mm BHCL,0.1% Triton X-100, 10mM imidazole and lysozyme 1mg/ml) followed by
237 sonication (0.5 s pulse and 0.9 s rest) at 2 min intervals to prevent overheating. After
238 centrifugation, the proteins were purified from the supernatant using nickel-nitrilotriacetic acid
239 (Ni-NTA) affinity chromatography (Cat. No.- 30410, Sigma). The recombinant proteins were
240 eluted from Ni-NTA resins by elution buffer (50 mM Tris-HCl pH. 8.0 and 300 mM NaCl)
241 containing 300mm Imidazole. The purified protein was concentrated using 10Kda Centricon
242 (ACS501024, Millipore) subjected to gel permeation chromatography on GE Healthcare
243 Superdex 75 Increase 10/300 GL where the monomeric fraction of the protein was eluted. The
244 final protein was stored in 50 mM Tris-HCl pH. 8.0 containing 300 mM NaCl,10% Glycerol
245 buffer at -80˚C.
246

247 **BALB/c aerosol challenge**

248 All mice experiments were carried out in the Tuberculosis Aerosol Challenge Facility (TACF,
249 ICGEB, New Delhi, India). BALB/c mice have been housed in individually ventilated cages
250 contained within the biosafety level 3 enclosure. The animal holding area of the BSL3 lab is
251 maintained at 20–25 °C, 30–60% humidity and 12–12 h of the light–dark cycle. An aerosol
252 challenge of ~100 CFUs was given to the animals in a Wisconsin–Madison chamber according
253 to the standardized protocol in the TACF facility. Two animals were selected randomly and

254 humanely euthanized 24 h post aerosol challenge to check for infection establishment. The
255 lungs and spleen tissues were harvested and homogenized to enumerate CFU. Tissue lysates
256 were serially diluted and plated on Petri plates containing Middlebrook 7H11 agar (Difco)
257 supplemented with 10% OADC (Becton, Dickinson) and 0.5% glycerol.
258
259 **RNA-Seq analysis**
260 Reads were quality filtered using Trimmomatic (v.0.39, LEADING:10 TRAILING:10
261 SLIDING WINDOW:4:15 MINLEN:75) [2] and were assessed for quality using FastQC (v.10.1)
262 [3] and visualised with MultiQC (v.1.9)[4]. Salmon (v.1.8, –numGibbsSamples 30) [5] was used to
263 quantify quality filtered reads against the human genome (GRCh38, primary assembly) with
264 Gencode annotation (v.40) [6]. Profiles were imported using tximeta (v.1.10) [7] and [8] (v.1.22) and
265 filtered to retain all transcripts from only protein-coding genes. Statistical analysis was done
266 using swish methodology from fishpond (v.1.8) [9] and significantly differentially expressed
267 transcripts (DETs) were defined based on *qvalue* ≤ 0.1. Furthermore, Isoform usage was
268 analysed using IsoformSwitchAnalyzeR (v.1.18) [10] with reduceToSwitchingGenes = FALSE,
269 reduceFurtherToGenesWithConsequencePotential = FALSE, No dIFcutoff and Isoform
270 fraction cuffoff. Heatmap of the DETs was generated with [11] using mean z-scaled normalised
271 scaled TPM values and clustered based on Euclidean distance and Ward.D2 methodology.
272 Clusters were defined manually. Enrichment analysis was done separately on overexpressed
273 and repressed DETs from each pairwise comparisons using enrichR (v.3.0) [12] with KEGG
274 Human 2021 and Gene Ontology (GO) [13] (Biological Process (BP), Cellular Component (CC)
275 and Molecular Function (MF)) 2021 database figures were generated using ggplot2 (v.3.3.5)
276 [14]. Analyses were done in R statistical software environment (v.4.1) [15].
277
278 **Generation of Rv1435c mutant**
279 *Rv1435c* mutant was generated using recombineering methodology which requires an Allelic
280 exchange substrate [16]. To construct that, 836bp upstream (LHS) and 1Kb downstream (RHS)
281 regions of the *Rv1435c* locus (retaining ~170 bp of the gene at both ends) were amplified from
282 *Mtb* (*H37Rv* genome) with the help of primers having appropriate restriction endonuclease sites
283 (LHS-DraIII; RHS-SfiI). The flanks were then digested and ligated with other two fragments-
284 Hygromycin resistant cassette and *oriE*+cosλ with compatible ends to create a four-piece AES.
285 The latter was then digested with SnaBI sites to release the linearised fragment consisting of
286 LHS-*Hyg*^{r}-RHS and electroporated into recombineering proficient *Rv::pNitET* (Kan^{r}) strain
287 that overexpresses RecET recombinases upon over-expression using Isovaleronitrile. Mutant

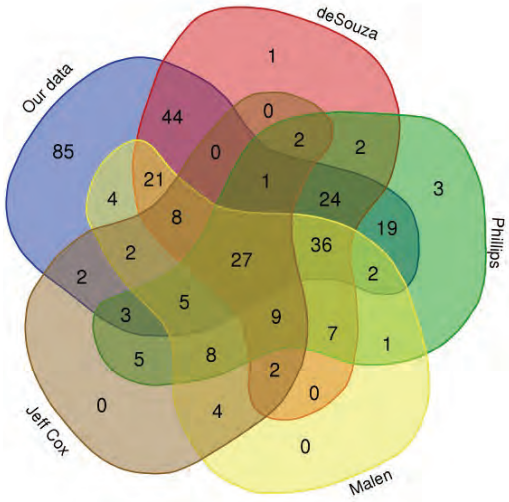
288 colonies were analysed using different primer sets to confirm legitimate recombination at the
289 required locus and cured of the pNitET plasmid by negative selection on 2% sucrose-containing
290 7H11-hyg^r agar plates.
291
292 Primers used:
293 *Rv1435c* 5' Flank forward primer with DraIII & SnaBI site:
294 5'CACCTTTTCACAAAGTGTACGTACGTGGTGCACGACGCATTGGA 3'
295 *Rv1435c* 5' Flank reverse primer with DraIII site:
296 5' TTTTTTTTCACTTCGTGCGCCGGCGGCTACCGGCGCGT 3'
297 *Rv1435c* 3' Flank forward primer with SfiI site:
298 5' CACCTTTTGGCCTAGATGGCCGGGCCGGTCCTGGCTGGTGCT 3'
299 *rv1435c* 3' Flank reverse primer with SfiI site and SnaB1 site:
300 5' TTTTTTTTGGCCACTTAGGCCTACGTAGGGATCGCTGGGAGCGTGGTT 3'
301
302 **Generation of Rv1435c complementation construct pSCW-F-Rv1435c**
303 *Rv1435c* was cloned along with its native promoter (500bp upstream of the gene) between ScaI
304 and HindIII sites using an integrative vector, pSCW-F (C-terminal 3X FLAG tag), which
305 facilitates the integration of the gene at Tweety attP sites in the *Mtb* genome.
306 Primers used:
307 FP *Rv1435c* (with UPSTREAM 500bp) ScaI
308 5' FP : 5' CACCAGTACTATGCGCCAGCGTGCTGTTG 3'
309 RP *Rv1435c* HindIII, w/o stop codon
310 5'TTTTAAGCTTGCGCGCACCTGCGGGC 3'
311
312 **Immunohistochemical staining:** The paraffin-embedded specimens were cut into 4–5 µm-
313 thick sections, deparaffinized in xylene for 5–10 min, and dehydrated through a series of graded
314 alcohols (100%, 100%, 95% and 80%). Tissue sections were boiled at 100°C for 30 min in
315 10mM citrate buffer (pH 6.0), followed by cooling and treatment with blocking solution (100%
316 methanol containing 0.3% H_2O_2) for 30 min to block any endogenous peroxidase activity. They
317 were incubated with SNRPF or SNRPD2 primary antibody (1:600, pH 9) overnight at 4°C,
318 followed by incubation with secondary antibody (Ultra trek anti polyvalent kit with HRP
319 conjugated secondary kit, Scy Tek Laboratories, Logan, USA) for 30 mins at room
320 temperature. The slides were thereafter stained with 3,3′-Diaminobenzidine (DAB). Finally,
321 the sections were counterstained with Hematoxylin and Eosin and observed under a

322 microscope. **Interpretation of IHC results:** SNRPF staining in >10% of cells was considered
323 positive. Both cytoplasmic and nuclear staining were looked for. The spatial distribution of the
324 stained sections, in terms of granulomas areas and areas away from granulomas, were noted
325 down to identify polarization of staining, if any.
326

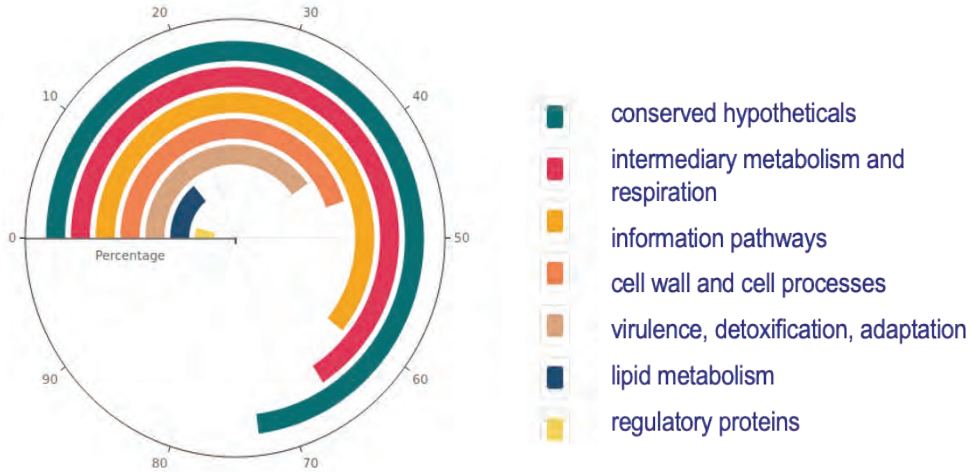
Legends to Figures S1-S9:

Figure S1

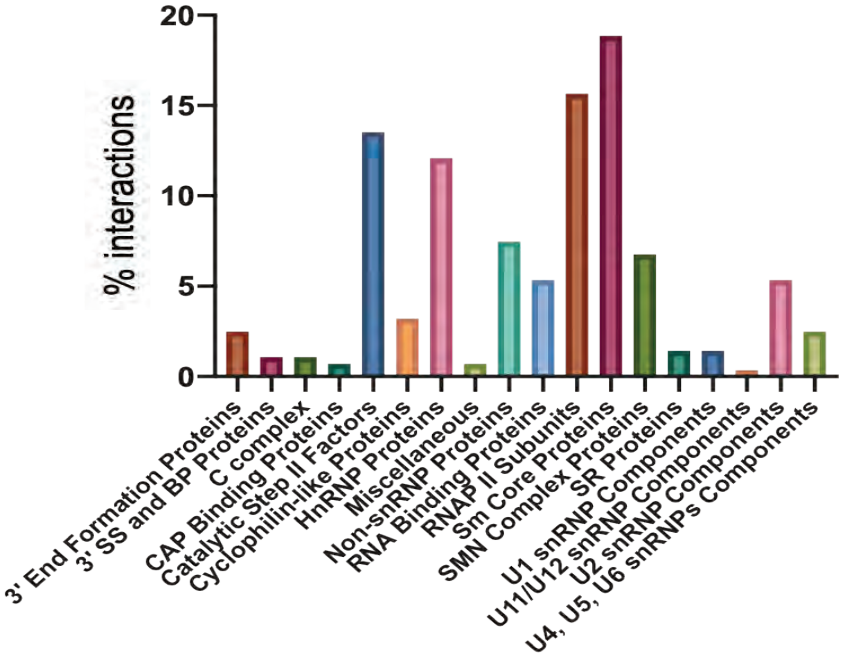
A



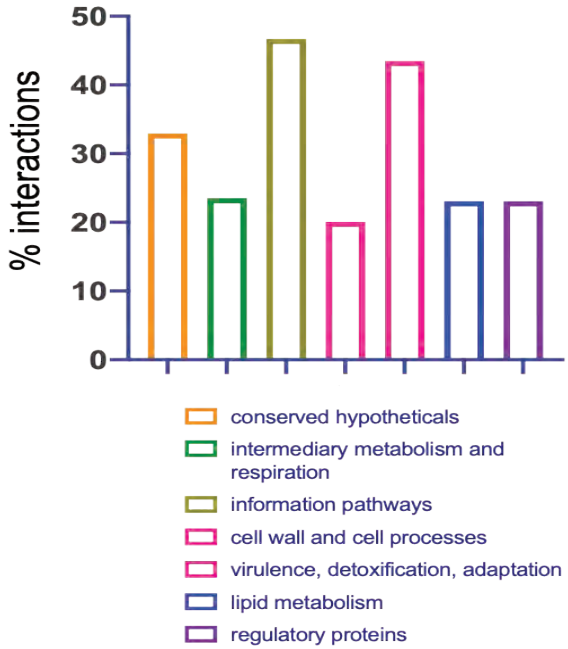
B



C



D



330 **Figure S1: Mycobacterial secretory proteins and host spliceosome proteins interactome.**
331 **A)** The Venn diagrams show the overlap of Mtb secretory proteins present in the bait
332 library with the secretory Mtb proteins identified in other studies.
333 **B)** The percentage of Mtb proteins in the Y2H bait library belongs to the different
334 functional classifications.
335 **C)** The percentage of the interaction shown by spliceosome-related proteins belonging
336 to different spliceosome protein categories
337 **D)** The percentage of the interacting Mtb proteins belonging to the different functional
338 classifications.
339
340

Gene	No. of isoform in Y2H	Interacting transcript	Interaction
SNRP40	2	0	No
PQBP1	2	0	No
NCBP2	2	0	No
POLR2H	4	4	19/10/4/1
SF3B1	2	0	No
SYF2	2	2	2 (Rv1643, Rv2927c)
DCAF4	2	1	2 (Rv1643, Rv2927c)
SNRPD2	2	1	16 (Rv1643, Rv2927c, Rv1566c)
LSm7	2	1	26 (Rv3864, Rv1643, Rv2927c, Rv1566c)
LSm4	2	0	No
SNRPG	3	0	No
RBPMS	2	0	No
SF1	3	0	No
CSTF	2	0	No
LUC7L2	2	1	4 (Rv1643, Rv2927c)
RBMX	2	1	8 (Rv3849)
U2AF1	3	1	2 (Rv1643, Rv2927c)
PPIH	3	1	1

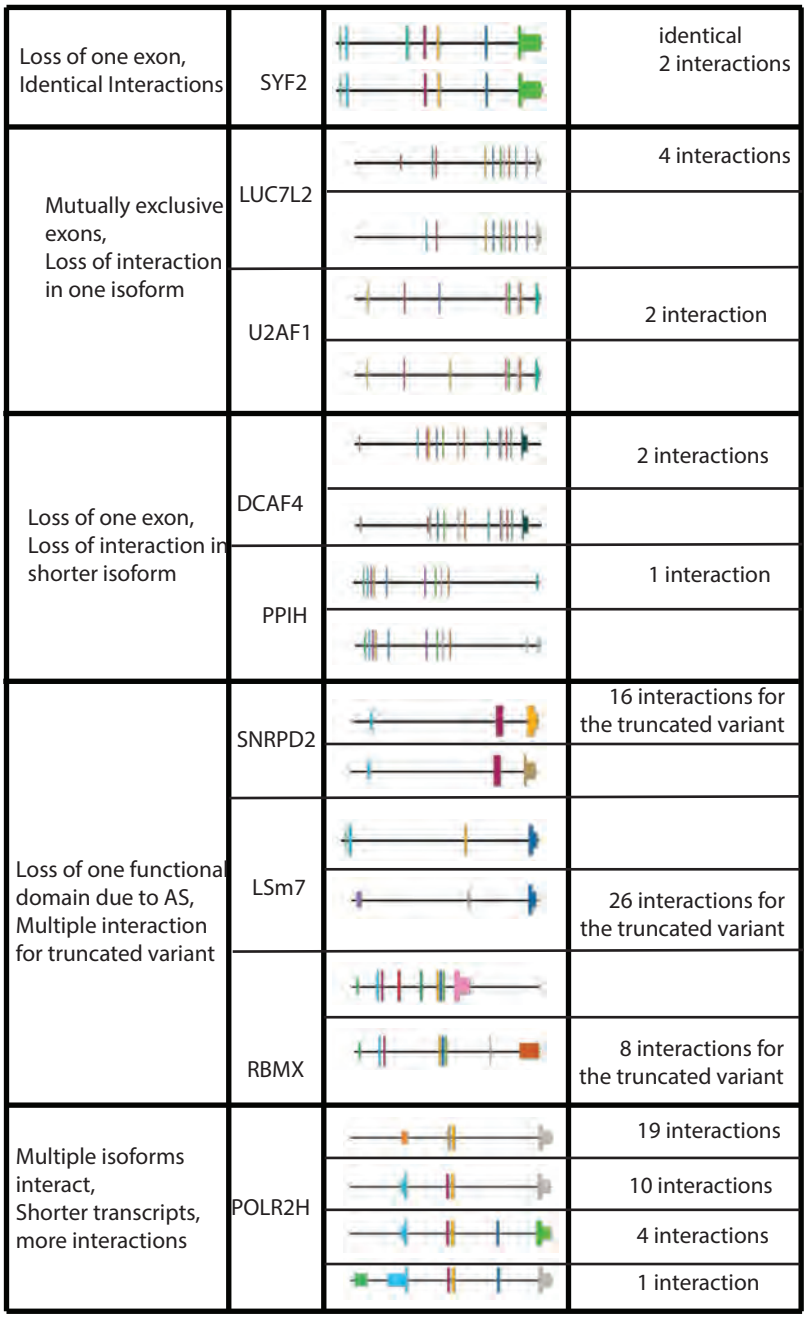
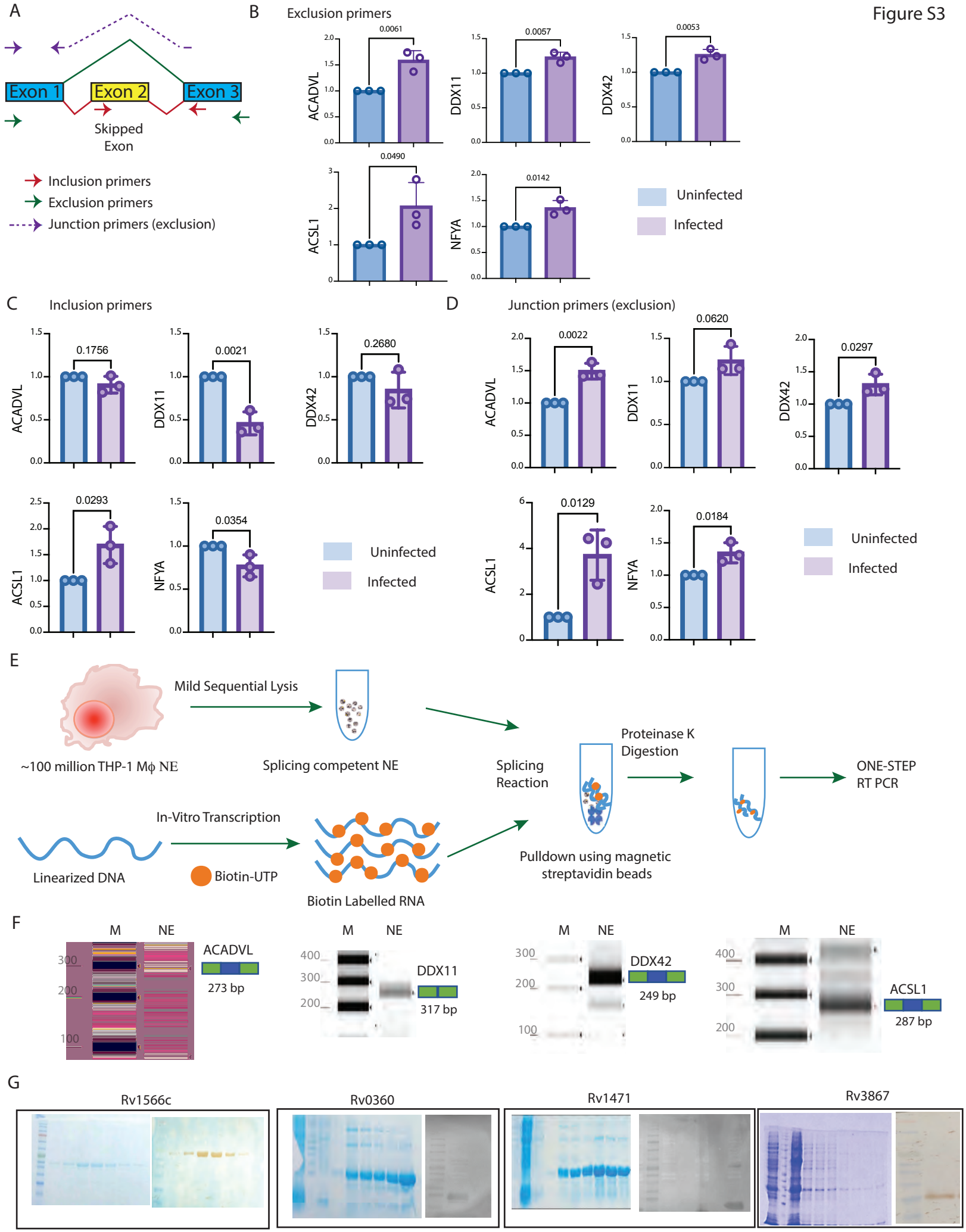


Figure S2: Interaction of Mtb proteins with the splice variants of human proteins.

A) A descriptive list of the interactions between the human spliceosome protein isoforms and the corresponding Mtb protein is presented in the table.

B) The table shows the distinct spliced mRNA isoforms that are classified as loss of exon, mutually exclusive exon and loss of the functional domain. The isoforms of SYF2, differing by only one exon, show interaction with the same Mtb proteins, suggesting no contribution of alternatively spliced exons towards interaction. Only one of the isoforms of genes which show exon skipping like DCAF4 and PPIH or mutually exclusive exons like U2AF1 and LUC7L2 showed interaction with any Mtb proteins suggesting the involvement of alternative exons in the interactions. The truncated isoforms of SNRPD2, LSm7 and RBMX show a loss of functional domains and multiple interactions with Mtb proteins while the longer isoforms for these genes did not show any interactions. Similarly, the truncated isoforms of POLR2H show more interactions.

Figure S3


A
Exon 1
Exon 2
Exon 3
Skipped Exon
Inclusion primers
Exclusion primers
Junction primers (exclusion)
B
Exclusion primers
ACADVL
DDX11
DDX42
ACSL1
NFYA
0.0061
0.0057
0.0053
0.0490
0.0142
Uninfected
Infected
C
Inclusion primers
0.1756
0.0021
0.2680
0.0293
0.0354
D
Junction primers (exclusion)
0.0022
0.0620
0.0297
0.0129
0.0184
E
Mild Sequential Lysis
~100 million THP-1 Mϕ NE
Splicing competent NE
Splicing Reaction
Proteinase K Digestion
ONE-STEP RT PCR
In-Vitro Transcription
Linearized DNA
Biotin-UTP
Biotin Labelled RNA
Pulldown using magnetic streptavidin beads
F
M
NE
ACADVL
273 bp
DDX11
317 bp
DDX42
249 bp
ACSL1
287 bp
G
Rv1566c
Rv0360
Rv1471
Rv3867

357 **Figure S3: *In vitro* splicing assays point to a direct role of *Mtb* proteins in regulating host-**
358 **RNA splicing.**

359 **A)** The schematic shows the primer design strategy for exon-inclusion, exon-
360 exclusion and exon-junction primers.

361 **B)** Expression of five target genes (as indicated) in THP-1 macrophages upon
362 infection with H37Rv using exon exclusion primers at 48 hpi. Data from three
363 independent biological experiments.

364 **C)** Expression of five target genes (as indicated) in THP-1 macrophages upon
365 infection with H37Rv using exon inclusion primers at 48 hpi. Data from three
366 independent biological experiments.

367 **D)** Expression of five target genes (as indicated) in THP-1 macrophages upon
368 infection with H37Rv using exon junction primers at 48 hpi. Data from three
369 independent biological experiments.

370 **E)** The schematic flow chart for In vitro splicing assay. The splicing-competent
371 nuclear extract (NE) and biotin-labelled substrate are used for the reaction. The
372 spliced products are amplified by one-step RT PCR.

373 **F)** The splice products for minigene cassettes of *ACADVL, DDX11, DDX42* and
374 *ACSL1* are analyzed on the Agilent tape station 3000.

375 **G)** His tagged *Mtb* protein Rv1566c, Rv0360, Rv1471 and Rv3867 are purified using
376 Ni-NTA affinity and size exclusion chromatography for *in vitro* splicing assay and
377 confirmed using immunoblotting.

378

379

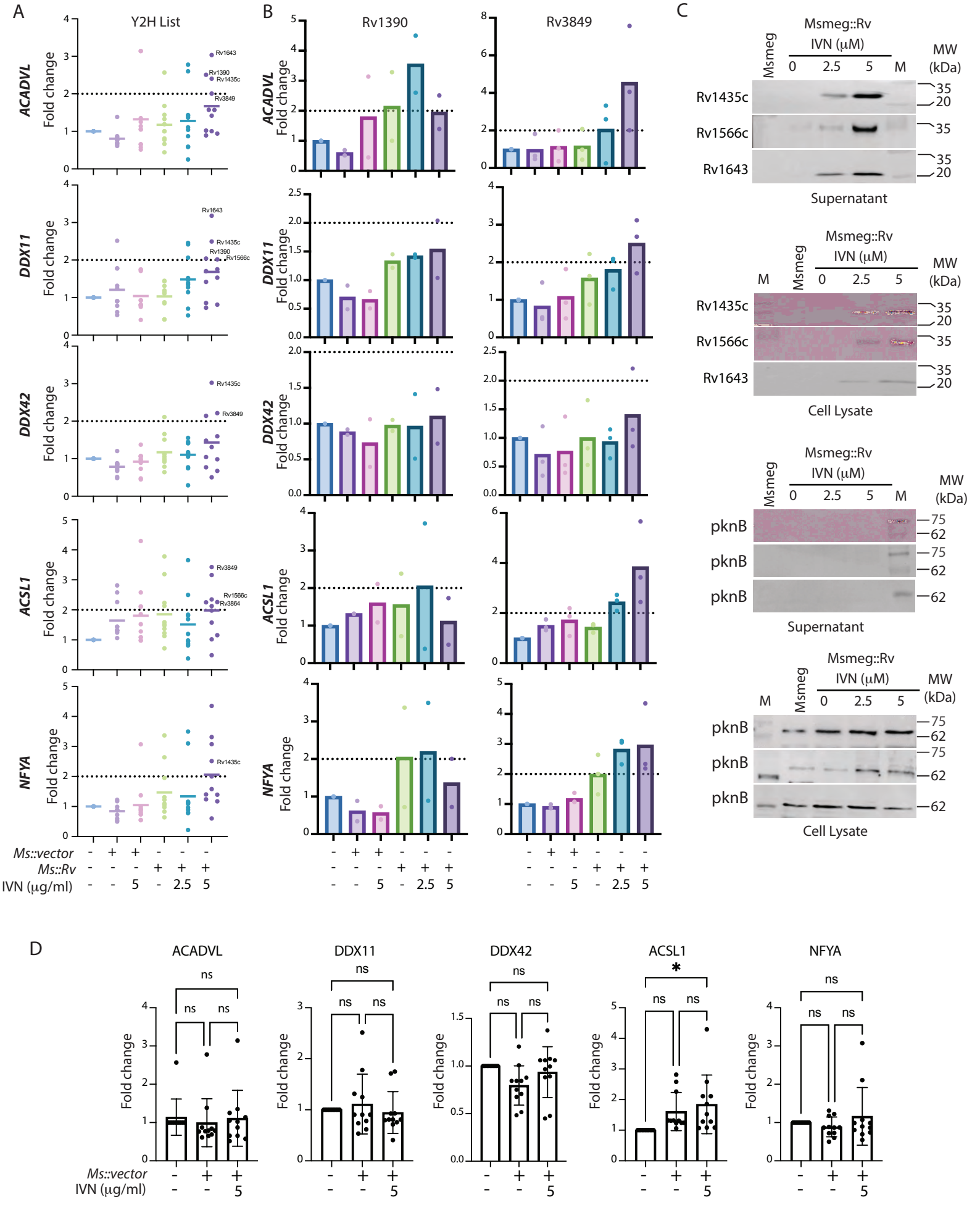

A
Y2H List
ACADVL
DDX11
DDX42
ACSL1
NFYA
Fold change
Rv1643
Rv1390
Rv1435c
Rv3849
Rv1566c
Rv3864
Ms::vector
Ms::Rv
IVN (μg/ml)
B
Rv1390
Rv3849
C
Msmeg::Rv
IVN (μM)
Msmeg
M
MW (kDa)
Supernatant
Cell Lysate
pknB
D
ACADVL
DDX11
DDX42
ACSL1
NFYA
ns
*

Figure S4: Mycobacterial proteins can influence RNA splicing during infection.

A) The M. smegmatis strains expressing the Mtb proteins identified through Y2H and spliceosome pulldown are used for infecting THP-1 macrophages followed by qPCR for truncated isoforms of *ACADVL, DDX11, DDX42, ACSL1* and *NFYA*. The plot represents the groups in the following order: Uninfected, M. smegmatis WT, M. smegmatis IVN induced, *M. smegmatis::Rv* uninduced, *M. smegmatis::Rv* 2.5µM IVN induced and *M. smegmatis::Rv* 5µM IVN induced. Each dot represents a different strain used for infection. The strains showing more than 2-fold increase in expression of truncated transcripts (dotted line) uniquely upon IVN induction are labelled.

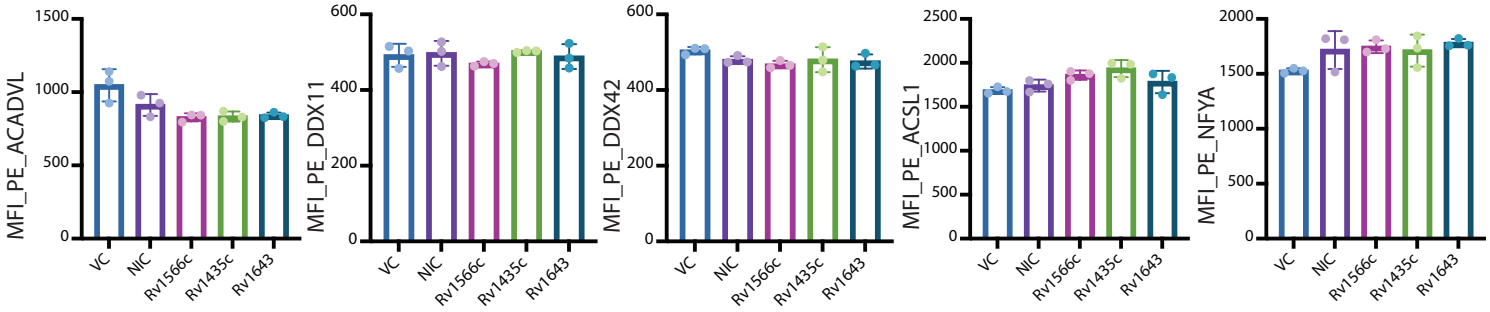
B) Expression of five target genes (as indicated) in THP-1 macrophages upon infection with Rv1390 and Rv3849 expressing M. smegmatis strains. The plot represents the groups in the following order: Uninfected, M. smegmatis WT, M. smegmatis IVN induced, *M. smegmatis::Rv* uninduced, *M. smegmatis::Rv* 2.5µM IVN induced and *M. smegmatis::Rv* 5µM IVN induced. Each dot represents an independent biological replicate.

C) The protein expression in modified *M. smegmatis* strains expressing Rv1566c, Rv1435c and Rv1643 upon induction with 2.5µM and 5 µM IVN was checked by immunoblotting in whole cell lysate and culture supernatant (SN). As control, the modified *M. smegmatis* strains expressing pknB were also induced with IVN and its expression in the supernatant or whole cell lysate was measured by immunoblotting.

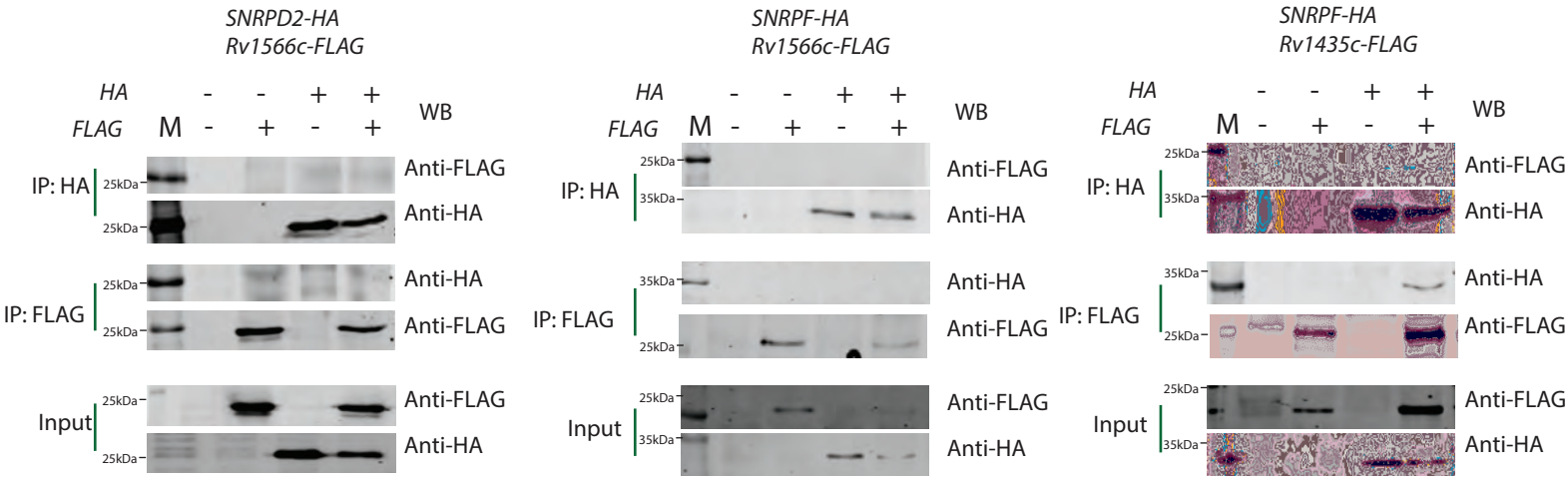
D) The plot represents the expression of truncated isoforms of *ACADVL, DDX11, DDX42, ACSL1* and *NFYA* upon infection with *pNit*-transfected *M. smegmatis* (vector control) or those where *M. smegmatis::Rv* lines without IVN-induction were used for infection. The p values are from One-way ANOVA.

Figure S5

A



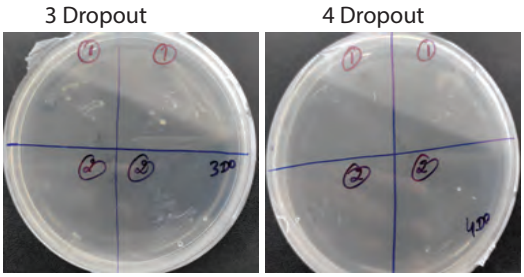
B



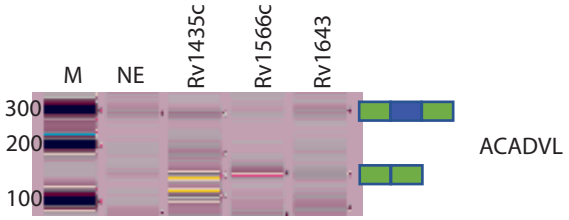
C

BD clones:
1. *pGBDKT7-MFAP1*
2. *pGBDKT7-PRPF18*

AD clone:
1. *pGADT7-Rv1435c*



D



408 **Figure S5: Mycobacterial secretory proteins and host spliceosome proteins interactome.**

409 **A)** The PE MFI in PE^+ $FITC^+$ dual positive population for each indicated reporter
410 construct in the presence of indicated mycobacterial protein or vector controls. Each
411 dot represents an independent biological replicate (N=3)

412 **B)** Co-immunoprecipitation assays to test the physical interactions between indicated
413 pairs of *Mtb* proteins (FLAG-tagged) and host proteins (HA-tagged) in HEK293T
414 cells.

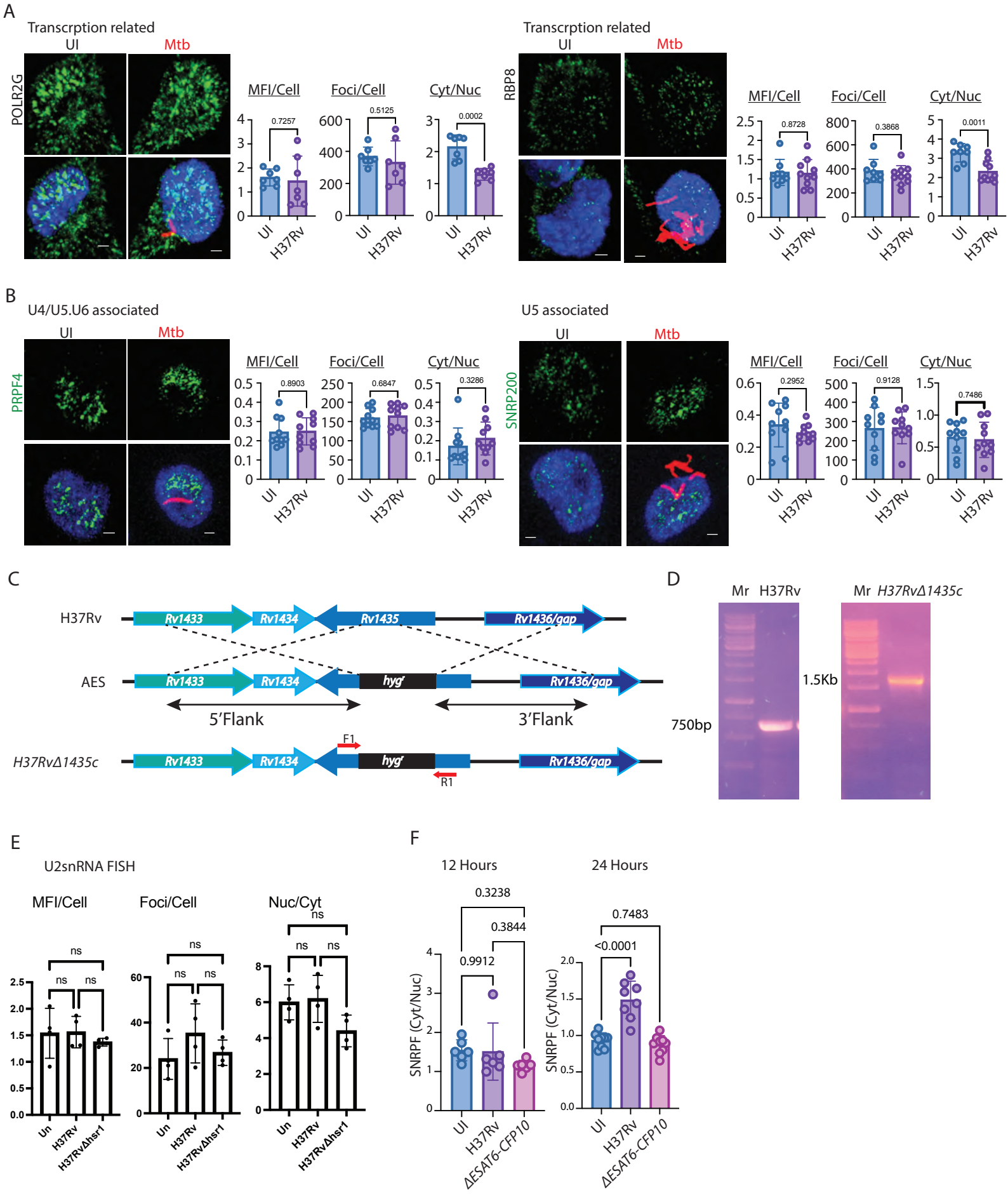
415 **C)** The 3-dropout and 4-dropout plates for reverse Y2H showing interaction for two
416 pairs: Rv1435c (in AD vector) and two human splicing-related proteins PRPF18
417 and MFAP1 (in BD vector).

418 **D)** In-vitro splicing assay in the presence of purified Mtb proteins Rv1435c, Rv1566c
419 and Rv1643 for ACADVL minigene cassette.

420

421

Figure S6


A
Transcrption related
UI
Mtb
POLR2G
MFI/Cell
Foci/Cell
Cyt/Nuc
0.7257
0.5125
0.0002
Transcrption related
RBP8
0.8728
0.3868
0.0011
B
U4/U5.U6 associated
PRPF4
0.8903
0.6847
0.3286
U5 associated
SNRP200
0.2952
0.9128
0.7486
UI
H37Rv
C
H37Rv
Rv1433
Rv1434
Rv1435
Rv1436/gap
AES
hyg'
5'Flank
3'Flank
H37RvΔ1435c
F1
R1
D
Mr H37Rv
Mr H37RvΔ1435c
1.5Kb
750bp
E
U2snRNA FISH
MFI/Cell
Foci/Cell
Nuc/Cyt
ns
Un
H37Rv
H37RvΔhsr1
F
12 Hours
24 Hours
0.3238
0.3844
0.9912
0.7483
<0.0001
SNRPF (Cyt/Nuc)
UI
H37Rv
ΔESAT6-CFP10

422 **Figure S6: Deletion of Hsr1 does not impact U2 snRNA localization.**

423 **A)** Representative image for staining of transcription-related proteins (POLR2G and
424 RBP8).The confocal images were analyzed and three different parameters were
425 plotted- overall MFI (MFI/cell), number of Foci (Foci/cell) and cytosolic to the
426 nuclear ratio (Cyt/Nuc) for SNRPF. Each dot represents a field analyzed, and an
427 average of hundred cells are quantified. Scale bar: 2 µm. The indicated p values are
428 from One-way ANOVA.

429 **B)** Representative image for staining of U4/U5.U6 related protein (PRPF4) and U5
430 associated protein (SNRP200). Each dot represents a field analyzed, and an average
431 of hundred cells are quantified.

432 **C)** The schematic of knocking out the *Rv1435c* or *hsr1* gene in H37Rv using the
433 recombineering approach. Positive colonies were screened with the F1 (gene
434 forward) & R1 (gene reverse) primer set shown by red arrows

435 **D)** Agarose gel electrophoresis to confirm the replacement of 609bp product
436 (*Rv1435c*) with 1.4Kb hygromycin cassette in the knockout line.

437 **E)** The plots corresponding to overall MFI (MFI/cell), number of Foci (Foci/cell) and
438 cytosolic to nuclear ratio (Cyt/Nuc) of U2 snRNA 48 hpi are shown here.

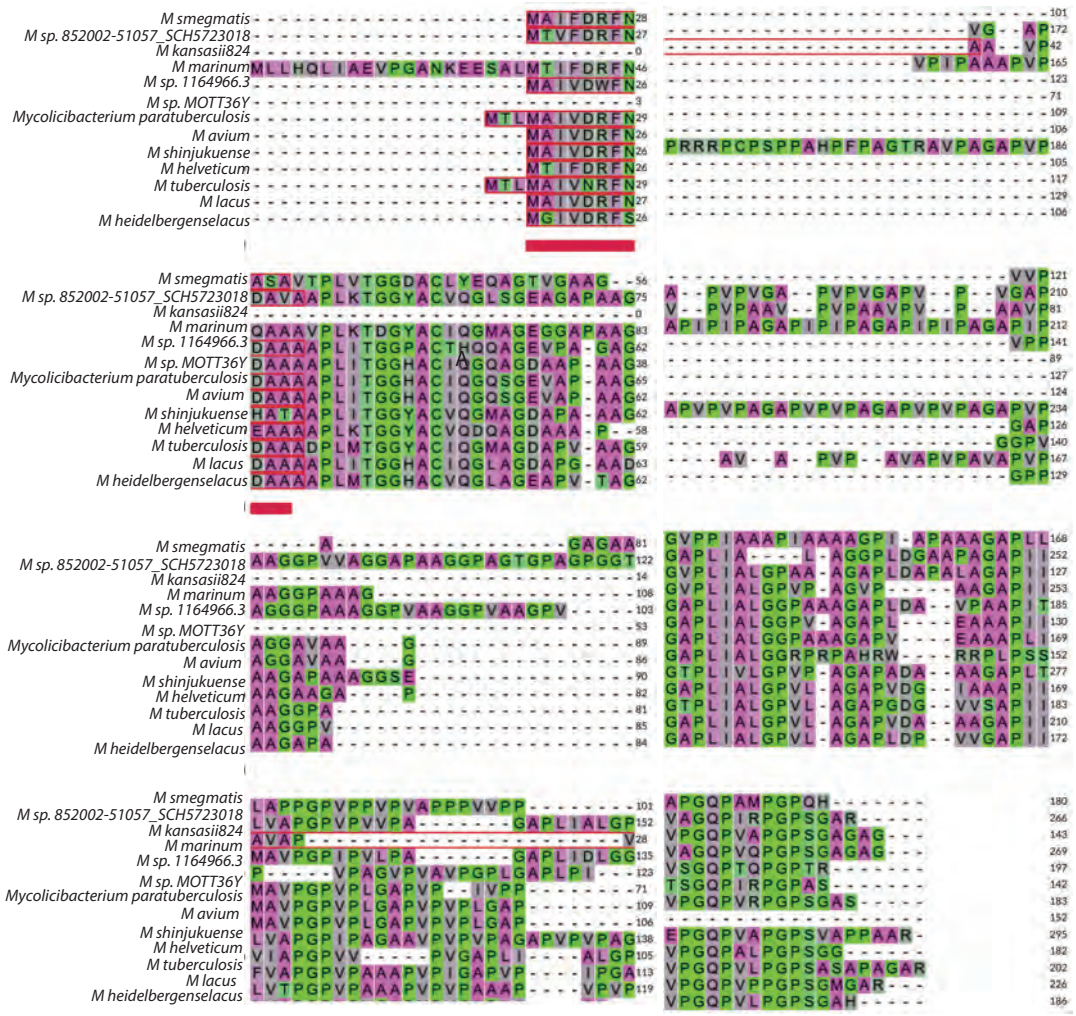
439 **F)** The plot show cytosolic to the nuclear ratio (Cyt/Nuc) for SNRPF in THP-1
440 macrophages infected with WT H37Rv, H37Rv*ΔCE* and H37Rv*ΔCE::WT* at 12hpi
441 and 24 hpi. The indicated p values are from One-way ANOVA.

442

443

Figure S7

A


M smegmatis
M sp. 852002-51057_SCH5723018
M kansasii824
M marinum
M sp. 1164966.3
M sp. MOTT36Y
Mycolicibacterium paratuberculosis
M avium
M shinjukuense
M helveticum
M tuberculosis
M lacus
M heidelbergenselacus

B

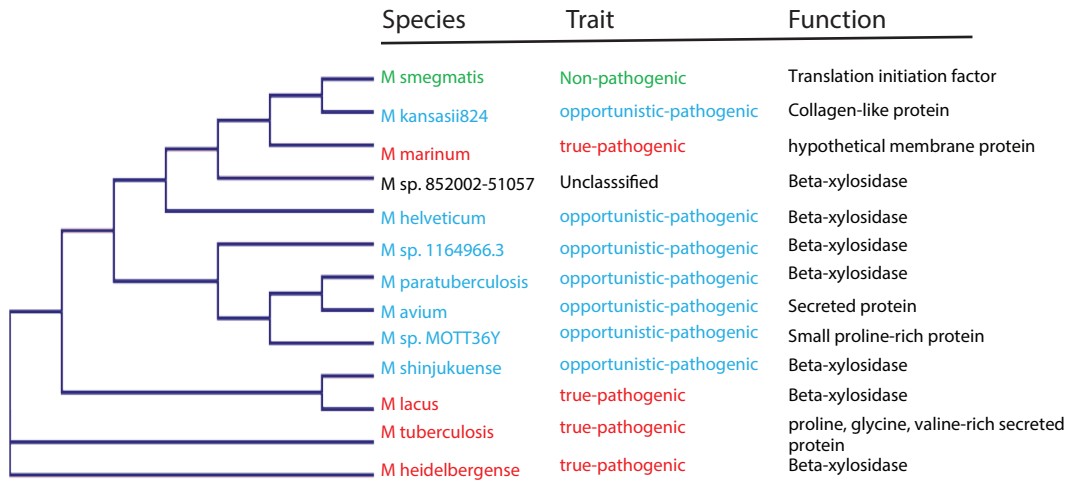
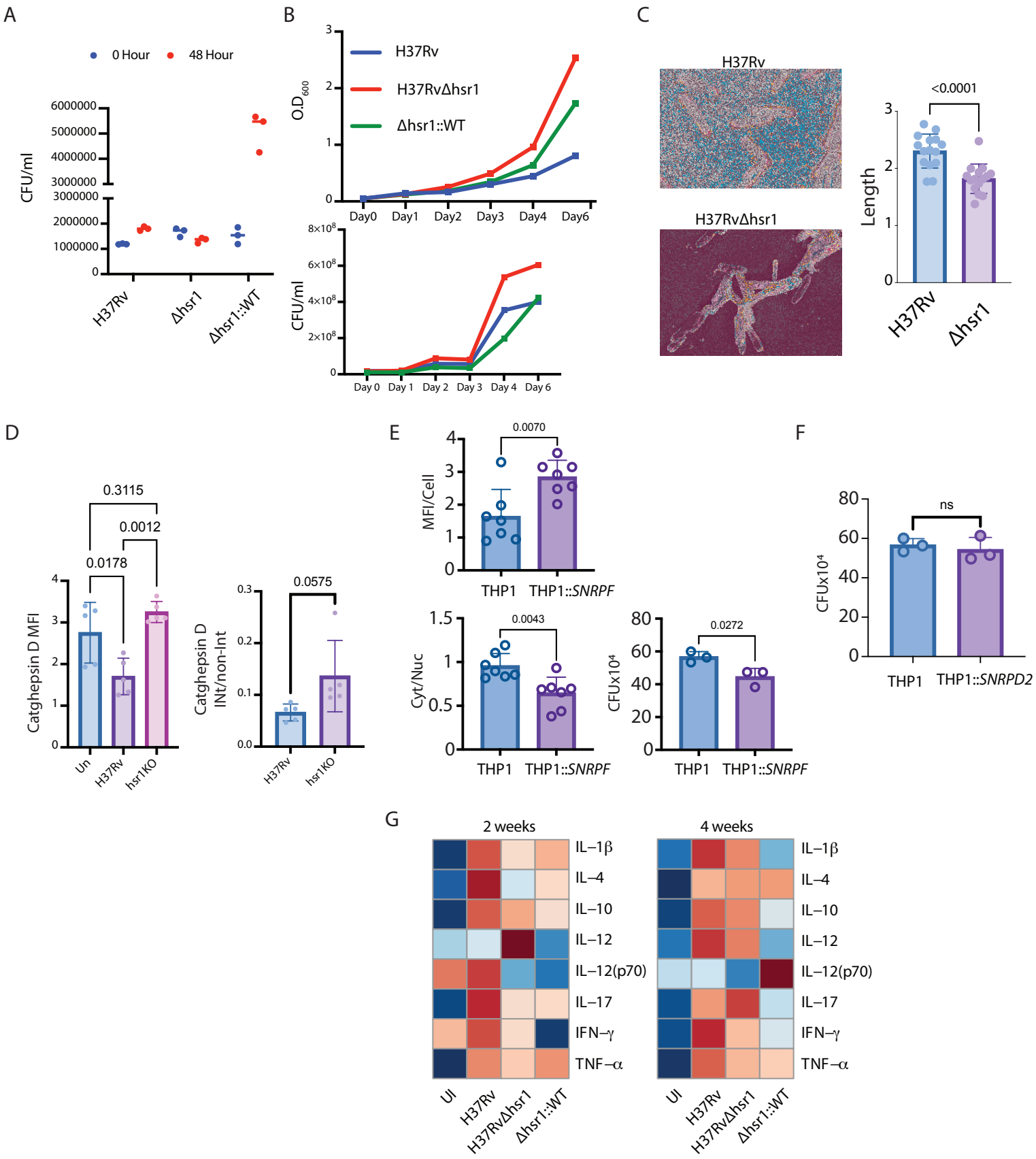

Species
Trait
Function
M smegmatis
Non-pathogenic
Translation initiation factor
M kansasii824
opportunistic-pathogenic
Collagen-like protein
M marinum
true-pathogenic
hypothetical membrane protein
M sp. 852002-51057
Unclasssified
Beta-xylosidase
M helveticum
opportunistic-pathogenic
Beta-xylosidase
M sp. 1164966.3
opportunistic-pathogenic
Beta-xylosidase
M paratuberculosis
opportunistic-pathogenic
Beta-xylosidase
M avium
opportunistic-pathogenic
Secreted protein
M sp. MOTT36Y
opportunistic-pathogenic
Small proline-rich protein
M shinjukuense
opportunistic-pathogenic
Beta-xylosidase
M lacus
true-pathogenic
Beta-xylosidase
M tuberculosis
true-pathogenic
proline, glycine, valine-rich secreted protein
M heidelbergense
true-pathogenic
Beta-xylosidase

Figure S7: Hsr1 show dynamic interaction with U5snRNA.

A) Multiple sequence alignment of Hsr1 protein in different mycobacterial species using CLUSTAL Omega. The red bar indicates the conserved sequence among different species.

B) A phylogenetic tree corresponding to the multiple sequence alignment is shown in S6C.


A
0 Hour
48 Hour
CFU/ml
6000000
5000000
4000000
3000000
3000000
2000000
1000000
0
H37Rv
Δhsr1
Δhsr1::WT
B
H37Rv
H37RvΔhsr1
Δhsr1::WT
O.D600
Day0
Day1
Day2
Day3
Day4
Day6
CFU/ml
8×10^8
6×10^8
4×10^8
2×10^8
Day 0
Day 1
Day 2
Day 3
Day 4
Day 6
C
H37Rv
H37RvΔhsr1
<0.0001
Length
H37Rv
Δhsr1
D
0.3115
0.0012
0.0178
Catghepsin D MFI
Un
H37Rv
hsr1KO
0.0575
Catghepsin D INt/non-Int
H37Rv
hsr1KO
E
0.0070
MFI/Cell
THP1
THP1::SNRPF
0.0043
Cyt/Nuc
THP1
THP1::SNRPF
0.0272
CFUx10⁴
THP1
THP1::SNRPF
F
ns
CFUx10⁴
THP1
THP1::SNRPD2
G
2 weeks
4 weeks
IL–1β
IL–4
IL–10
IL–12
IL–12(p70)
IL–17
IFN–γ
TNF–α
UI
H37Rv
H37RvΔhsr1
Δhsr1::WT

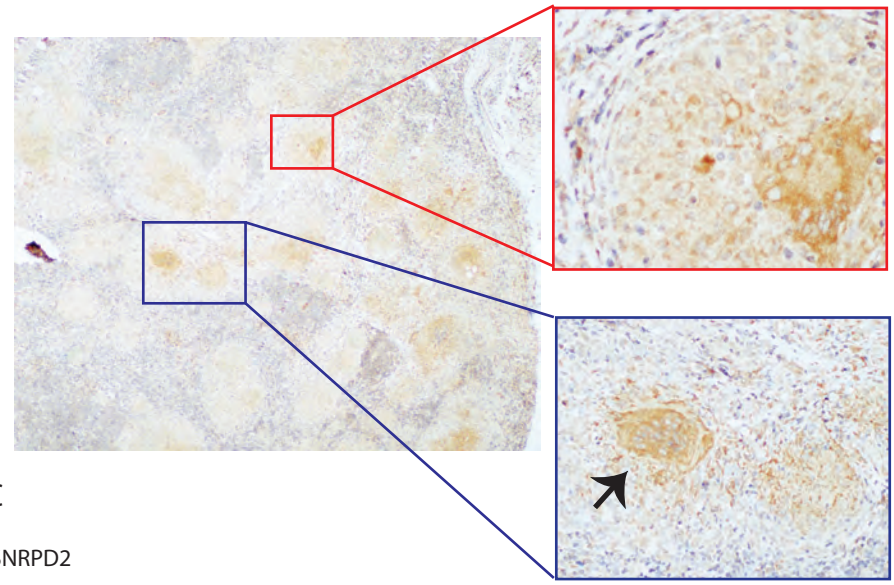
452 **Figure S8: Regulation of host alternative splicing as a virulence mechanism.**
453
454 A) The bacterial CFU of WT H37Rv, H37Rv*Δhsr1* and H37Rv*Δhsr1::WT* strains in
455 THP-1 macrophages at 0hr and 48hr post-infection.
456 B) The graph showing growth kinetics of WT H37Rv, H37Rv*Δhsr1* and
457 H37Rv*Δhsr1::WT* strains over the course of 6 days at 37°C to log phase in
458 Middlebrook 7H9 media with 0.05% Tween80, 10% ADC Enrichment, and 0.2%
459 glycerol.
460 C) The representative image for SEM of WT H37Rv and H37Rv*Δhsr1* and the plot
461 indicate quantification of the bacterial length.
462 D) The plot shows quantification for staining of cathepsin D in THP1 macrophages
463 infected with WT H37Rv and H37Rv*Δhsr1* at 48hpi. The different parameters
464 plotted show overall MFI (MFI/cell) and the ratio of cathepsin D foci interacting
465 with bacterial phagosomes (int/non-int). Each dot represents a field analyzed, and
466 an average of a hundred cells are quantified. The values represent p-values (One-
467 way ANOVA).
468 E) The plot shows quantification for staining of SNRPF in WT THP1 macrophages
469 and THP1::SNRPF infected with WT H37Rv at 48hpi. The different parameters
470 plotted show overall MFI (MFI/cell) and the ratio of cytosolic localization
471 (Cyt/Nuc) Each dot represents a field analyzed, and an average of a hundred cells
472 are quantified. The plot on the lower right panel shows the bacterial survival after
473 48 hours post-infection in both WT THP1 and THP1::SNRPF. Each dot represent
474 an independent biological experiment. Indicated p values represent One-way
475 ANOVA.
476 F) The plot represents the bacterial survival in WT THP-1 and THP-1::SNRPD2 at
477 48hr post-infection.
478 G) Heatmap showing the cytokine levels in the lung homogenates of BALB/c mice
479 after 2 weeks and 4 weeks of infection with WT H37Rv, H37Rv*Δhsr1* and
480 H37Rv*Δhsr1::WT.*
481
482

Figure S9

A

SNRPF

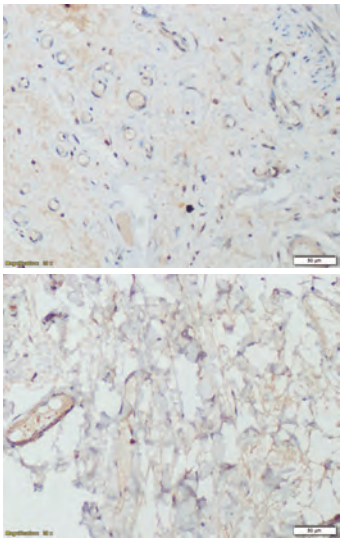
ITB sections



B

SNRPF

Control tissue

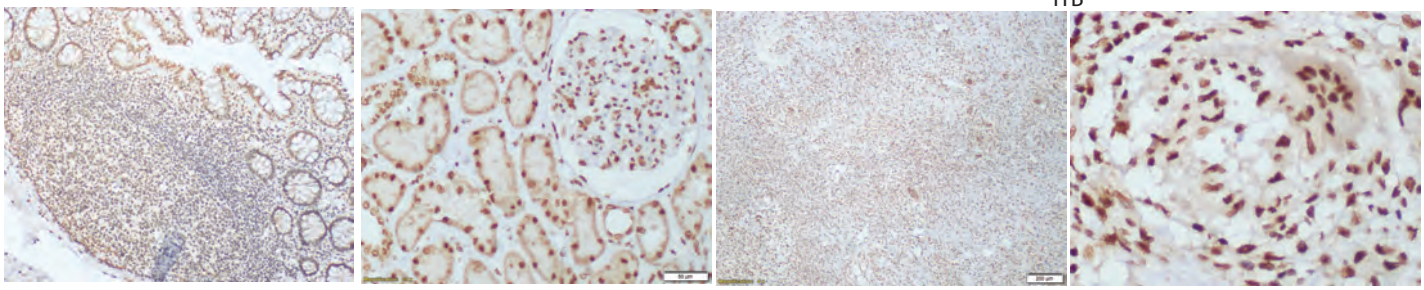


C

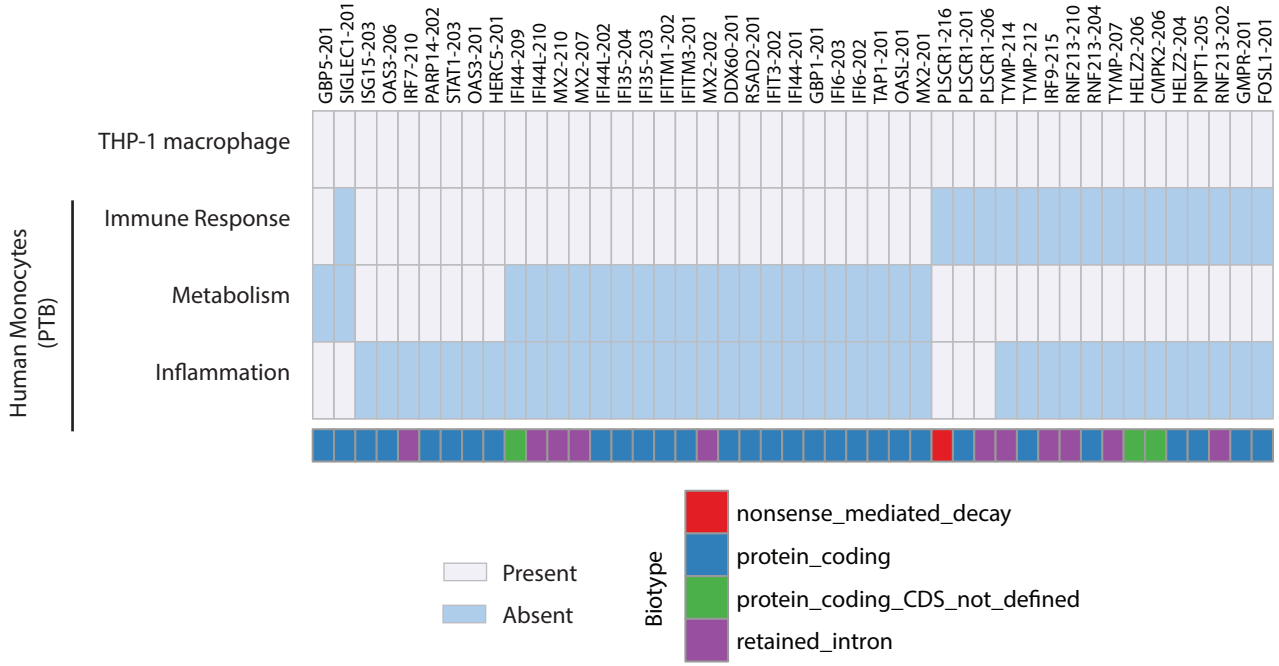
SNRPD2

Control tissue

ITB



D



483 **Figure S9: Regulation of host alternative splicing as a virulence mechanism**
484
485 A) Human intestinal tuberculosis granulomatous lesions were stained with SNRPF.
486 Intense cytosolic staining in multinucleated giant cells (black arrow) is shown.
487 B) Representative image for IHC staining of SNRPF in gut sections from healthy
488 controls.
489 C) Representative image for IHC staining of SNRPD2 in gut sections from healthy
490 controls and in intestinal tuberculosis granulomatous lesions from diseased
491 individuals.
492 D) Heatmap showing comparison of transcripts regulated in THP-1 macrophages upon
493 H37Rv infection and in the human monocytes from PTB subjects
494
495
496
497
498
499
500 **Supplemental Tables:**
501
502 **Table S1**: List of Mtb genes in the bait library (BD library) and human splicing factor coding
503 transcripts in the prey library (AD library).
504 **Table S2**: List of genes included in the BD library as negative controls
505 **Table S3**: Yeast-2-hybrid interaction matrix for AD and BD proteins
506 **Table S4**: Two-column file for the interaction network shown in Fig 1B.
507 **Table S5**: Spliceosome proteins identified in Y-12 pulldown MS analysis
508 **Table S6**: RNAseq data from THP-1 macrophages uninfected or infected with WT H37Rv or
509 H37Rv*Δhsr1* shown in Fig. 5B.
510 **Table S7**: RNAseq data shown in Fig. 6E.
511 **Table S8**: Transcripts present in THP-1 RNAseq and human monocyte RNAseq
512 **Table S9**: List of primers
513
514
515

518 **Supplementary References**

519

520 1 Liu, J. L. & Gall, J. G. U bodies are cytoplasmic structures that contain uridine-rich small
521 nuclear ribonucleoproteins and associate with P bodies. *Proc Natl Acad Sci U S A* **104**, 11655-
522 11659 (2007). https://doi.org:10.1073/pnas.0704977104
523 2 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
524 data. *Bioinformatics* **30**, 2114-2120 (2014). https://doi.org:10.1093/bioinformatics/btu170
525 3 Andrews, S. (2017).
526 4 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for
527 multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).
528 https://doi.org:10.1093/bioinformatics/btw354
529 5 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
530 aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
531 https://doi.org:10.1038/nmeth.4197
532 6 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes.
533 *Nucleic Acids Res* **47**, D766-D773 (2019). https://doi.org:10.1093/nar/gky955
534 7 Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in
535 RNA-seq. *PLoS Comput Biol* **16**, e1007664 (2020).
536 https://doi.org:10.1371/journal.pcbi.1007664
537 8 Martin Morgan, V. O., Jim Hester, Hervé Pagès. SummarizedExperiment:
538 SummarizedExperiment container. *R package version 1.28.0* (2022).
539 https://doi.org:10.18129/B9.bioc.SummarizedExperiment
540 9 Zhu, A., Srivastava, A., Ibrahim, J. G., Patro, R. & Love, M. I. Nonparametric expression
541 analysis using inferential replicate counts. *Nucleic Acids Res* **47**, e105 (2019).
542 https://doi.org:10.1093/nar/gkz622
543 10 Vitting-Seerup, K. & Sandelin, A. IsoformSwitchAnalyzeR: analysis of changes in genome-
544 wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**,
545 4469-4471 (2019). https://doi.org:10.1093/bioinformatics/btz247
546 11 Kolde, R. Pheatmap: pretty heatmaps. *R package version v.1.0.12* (2019).
547 12 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis
548 tool. *BMC Bioinformatics* **14**, 128 (2013). https://doi.org:10.1186/1471-2105-14-128
549 13 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology
550 Consortium. *Nat Genet* **25**, 25-29 (2000). https://doi.org:10.1038/75556
551 14 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* (2016).
552 15 Team, R. C. R: A Language and Environment for Statistical Computing. (2022).
553 16 van Kessel, J. C. & Hatfull, G. F. Recombineering in Mycobacterium tuberculosis. *Nat*
554 *Methods* **4**, 147-152 (2007). https://doi.org:10.1038/nmeth996

555

556

557