



# Alternative Splicing in Apicomplexan Parasites

Lee M. Yeoh,<sup>a,b</sup> V. Vern Lee,<sup>a</sup> Geoffrey I. McFadden,<sup>b</sup>  Stuart A. Ralph<sup>a</sup>

<sup>a</sup>Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Australia

<sup>b</sup>School of BioSciences, The University of Melbourne, Parkville, Australia

**ABSTRACT** Alternative splicing is a widespread, essential, and complex component of gene regulation. Apicomplexan parasites have long been recognized to produce alternatively spliced transcripts for some genes and can produce multiple protein products that are essential for parasite growth. Recent approaches are now providing more wide-ranging surveys of the extent of alternative splicing; some indicate that alternative splicing is less widespread than in other model eukaryotes, whereas others suggest levels comparable to those of previously studied groups. In many cases, apicomplexan alternative splicing events appear not to generate multiple alternative proteins but instead produce aberrant or noncoding transcripts. Nonetheless, appropriate regulation of alternative splicing is clearly essential in *Plasmodium* and *Toxoplasma* parasites, suggesting a biological role for at least some of the alternative splicing observed. Several studies have now disrupted conserved regulators of alternative splicing and demonstrated lethal effects in apicomplexans. This minireview discusses methods to accurately determine the extent of alternative splicing in Apicomplexa and discuss potential biological roles for this conserved process in a phylum of parasites with compact genomes.

**KEYWORDS** *Plasmodium*, RNA splicing, *Toxoplasma*, apicomplexan parasites, posttranscriptional control mechanisms

Gene regulation at the level of mRNA abundance has been extensively characterized in apicomplexans using glass and silicon microarrays, SAGE analyses, expressed sequence tag (EST) profiling, and RNA-seq projects, but analysis of posttranscriptional regulation has until recently been sparsely interrogated. This is despite experiments indicating that posttranscriptional steps are crucial for determining cellular function and differentiation in Apicomplexa. The importance of posttranscriptional gene regulation is emphasized by the surprisingly poor correlation between raw transcript and protein abundances in apicomplexan parasites. Temporal patterns in peak timing of transcription match quite poorly with appearance of corresponding proteins, even allowing for a lag between transcription and translation, and ranking of abundance by transcript levels correlates only moderately with abundance of the encoded proteins measured by quantitative proteomics (1–3). Some of this discrepancy may be explained by shortcomings in the measurement tools, but this discrepancy also points to considerable posttranscriptional regulation. Several mechanisms involving translational repression of existing transcripts have been identified and regulate transitions between stages (4–7), but these mechanisms appear not to explain all the mismatch between transcript and protein levels within stages.

To fully appreciate how parasites coordinate protein expression, we clearly need a better comprehension of the events regulating expression downstream of transcriptional initiation. Constitutive splicing and alternative splicing clearly play major roles in posttranscriptional control in other eukaryotic groups. In this minireview, we discuss recent advances that paint a fuller picture of these important aspects of posttranscriptional regulation in apicomplexan parasites.

**Citation** Yeoh LM, Lee VV, McFadden GI, Ralph SA. 2019. Alternative splicing in apicomplexan parasites. mBio 10:e02866-18. <https://doi.org/10.1128/mBio.02866-18>.

**Invited Editor** David Sibley, Washington University School of Medicine

**Editor** Danielle A. Garsin, University of Texas Health Science Center at Houston

**Copyright** © 2019 Yeoh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Stuart A. Ralph, [saralph@unimelb.edu.au](mailto:saralph@unimelb.edu.au).

**Published** 19 February 2019

## SPLICING

The best-known form of RNA processing is splicing. Splicing of protein-coding genes removes noncoding introns from within precursor mRNA molecules, turning them into mature mRNAs with a single open reading frame ready for translation. In eukaryotes, protein-coding genes are spliced using a large complex of RNAs and proteins called the spliceosome. Two distinct spliceosome types exist in many eukaryotes: the major, U2-type spliceosome excises most introns, while a minor or U12-type spliceosome removes a small percentage of introns, often those with variant splice sites. While some apicomplexan introns depart from the canonical TG-AG sites, the minor spliceosome appears to be absent in Apicomplexa, and indeed from Alveolata altogether (8). Recent work suggests that components of the apicomplexan U2-type spliceosome have acquired certain divergent features (9). For example, *Plasmodium* spliceosomal RNAs (UsnRNAs) possess unusual 3' poly(A) extensions (10), while several proteins normally involved in snRNA trafficking are apparently absent in apicomplexans (11), and some spliceosomal proteins contain divergent sequence features (12). Nonetheless, from what has been described thus far, the overall assembly, structure, and function of this apparatus closely reflects what is known from model eukaryotes, and we refer readers to a recent review for a survey of the general splicing machinery (9).

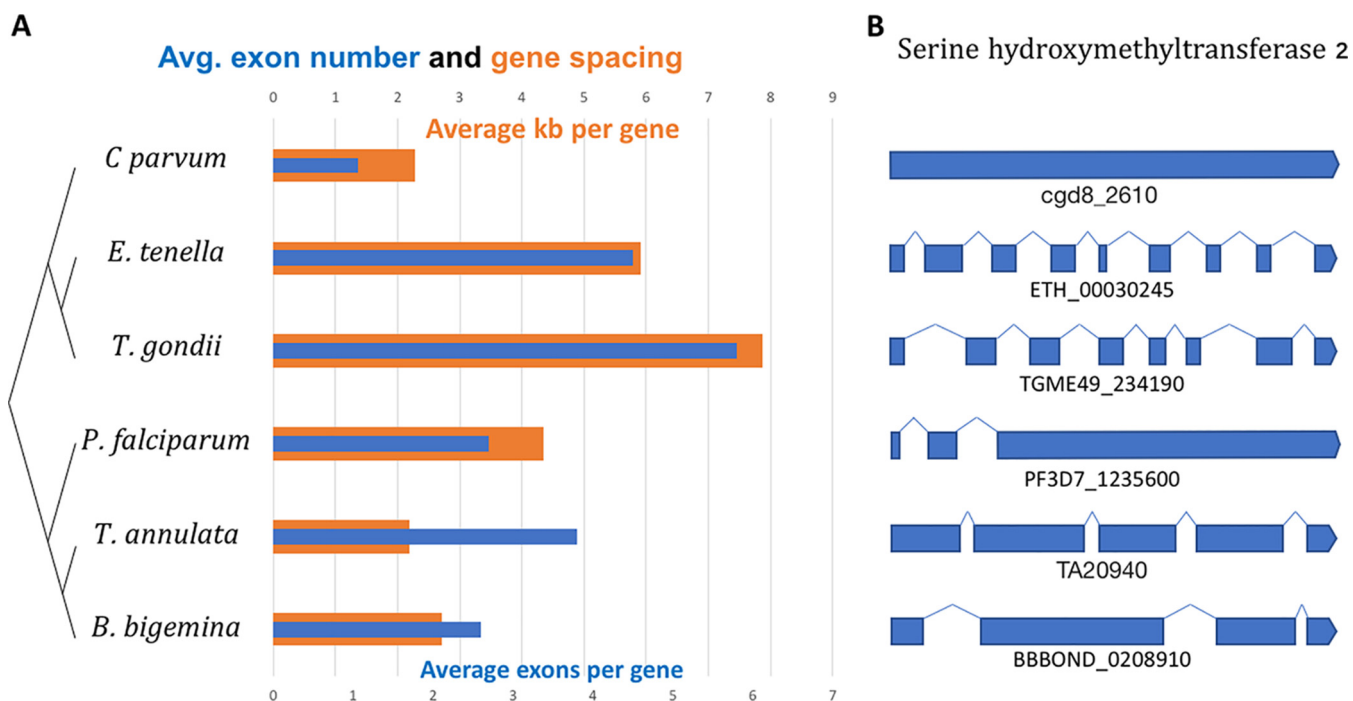
## EXON/INTRON DISTRIBUTION IN APICOMPLEXA

While the machinery for removing introns in Apicomplexa is apparently largely conserved and constant, the number, size, and distribution of introns are strikingly diverse in different apicomplexan genera. The number of genes in Apicomplexa is relatively consistent compared to the highly variable genomes of phyla such as Arthropoda or Angiospermae, with most apicomplexan genera possessing somewhere around 6,000 ( $\pm 40\%$ ) genes. Genome size and gene density within genomes, however, are highly variable; the smallest genome so far sequenced, *Babesia microti* (6.1 to 6.5 Mbp) (13) is 10 to 20 times smaller than some of the coccidian genomes like *Toxoplasma gondii* (65.7 Mbp) and *Sarcocystis neurona* (~127 Mbp) (14). As in other eukaryotes, this genome size variation largely tracks with the number of exons per gene. Whereas some small-genome apicomplexans have hardly any introns (fewer than 5% of *Cryptosporidium* genes are predicted to have an intron [15]), some species with larger genomes have an average of more than five exons per gene. There is generally an inverse relationship between the density of genes in apicomplexan genomes and number of exons—parasites with many genes per kilobase of genome generally have fewer exons (Fig. 1). Conceivably, complex gene structure allows for more intricate RNA processing in some apicomplexans, and more possibilities for gene regulation through alternative splicing, although this possibility remains to be experimentally tested. It is also possible, but unproven, that the compact genomes in this phylum (~1/4 the number of genes, but 1/130 the genome size of humans) generate a requirement for alternative splicing to allow a smaller complement of protein-coding genes.

## ALTERNATIVE SPLICING

**How much alternative splicing is there?** The discovery of mRNA splicing in the late 1970s was simultaneous with the revelation that a single species of pre-mRNA could be spliced differentially, creating multiple, distinct mature mRNAs (16, 17) now known as alternative splicing. More recent analyses have demonstrated that alternative splicing is widespread in metazoans. For example, in humans, Wang and colleagues sequenced more than 400 million 32-bp cDNA fragments from ten different tissues and five mammary cancer cell lines (18). Wang et al. (18) found that 92 to 94% of genes were alternatively spliced, with most of these alternatively spliced variants showing tissue-specific regulation. Indeed, alternative splicing has been frequently linked to tissue specificity in metazoans (19), and it is essential for cell differentiation (20).

Although apicomplexan genomes are less well characterized than model animal genomes, a large number of studies have surveyed alternative splicing in Apicomplexa. The early sequencing of a 13.6-kb contig and associated cDNAs from *Plasmodium*



**FIG 1** Gene structure in Apicomplexa. (A) Gene structure in *Cryptosporidium parvum*, *Eimeria tenella*, *Toxoplasma gondii*, *Plasmodium falciparum*, *Theileria annulata*, and *Babesia bigemina*. Whereas some apicomplexan genera have very few introns, others have many introns in some genes and at least one intron in most genes. Exon number in the phylum tends to track with genome size. (B) Gene structure varies widely within the phylum Apicomplexa, even between closely related genera. Apicomplexan orthologues of a representative gene, serine hydroxymethyltransferase 2, are depicted as one example. The transcripts are similar but not exactly equal lengths, but all are drawn to scale relative to the length of each gene. Gene IDs are listed below each gene.

*berghei* uncovered six genes, two of which exhibited alternative splicing (21). In addition, one of these two genes was specific to gametocytes in both *P. berghei* and *Plasmodium falciparum* (21). More recent larger surveys include expressed sequence tag (EST) projects, targeted sequencing, and transcriptome-wide RNA-seq projects. Medium-throughput collections include a large number of cDNA expressed sequence tag (EST) libraries for a wide variety of genera such as *Plasmodium*, *Babesia*, *Eimeria*, *Toxoplasma*, *Sarcocystis*, and *Cryptosporidium* (22–27) and traditional Sanger sequencing of full-length cDNAs (28–30). Alternative splicing of a few specific genes had been implicated from expressed sequence tag findings (31, 32). However, very few of the EST libraries were explicitly analyzed for alternative splicing. One notable exception identified 42 alternatively spliced genes in *P. falciparum*, with 29 of these further corroborated by PCR (33).

A separate, medium-throughput screen that focused on alternative splicing identified 14 alternatively spliced genes from 88 randomly selected open reading frames (ORFs) from *P. falciparum*, representing 16% of ORFs analyzed (34). This study employed Sanger sequencing of targeted ORFs from schizont and gametocyte stages. While this method had the theoretical advantage of more thorough identification of alternative splicing for each individual ORF, the authors sequenced only two to eight clones per intron. This depth would miss most of the rarer splice isoforms, and the authors therefore contended that the true prevalence of alternatively spliced proteins must be substantially higher.

The past decade has seen transcriptome studies extend beyond EST characterization to more extensive Illumina RNA-seq surveys. However, very few experiments have specifically focused on detecting alternative splicing, and most have used nonspecific algorithms that lack the power to correctly detect alternative splicing. Otto and colleagues sequenced just under 5 million 37- to 54-bp cDNA fragments in total from seven time points in the intraerythrocytic cycle of *P. falciparum* (35). From this, they identified 75 alternatively spliced genes from 5,438 genes analyzed (1.4%). While this

number appears lower than the corresponding values for the human genome, it remains to be seen whether this discrepancy is due to a real biological difference or simply an issue of depth. In addition to the reduced depth of sequencing, the inclusion of limited life cycle stages of unicellular protists could well have left much alternative splicing undetected. While the aforementioned *P. falciparum* transcriptomic survey included different intraerythrocytic stages (35), it is likely that including more diverse life cycle stages, such as those in the liver and mosquito, will identify additional transcript isoforms.

A subsequent study found a greater number of *P. falciparum* genes, 254 or 4.5% of genes analyzed, but this analysis was again restricted to blood stages (36). The amount of alternative splicing detected from similar studies in other *Plasmodium* species has also been quite low, with only 77 such genes detected in *Plasmodium vivax* (37) and 50 genes in *Plasmodium yoelii* (38). In *Toxoplasma gondii*, only 77 genes (0.8% of the genes analyzed) were initially identified as alternatively spliced from sequencing of a single stage, followed by *de novo* transcript assembly, alignment to the genome, and detection of differences to canonical gene models supported by at least two unique reads (39). However, another study using a novel alternative splicing detection method identified 1,914 alternatively spliced genes (22.6%) in the same organism, although even this underestimated the total amount, as it excluded intron retention as a class of alternative splicing (40). The discrepancy between the two studies in *T. gondii* may be due to several factors. First, the former study sequenced 270 million paired-end reads, but these were only 40 bp long. The latter sequenced only 50 million paired-end reads, but these were 100 bp long. Second, there were substantial differences in computational methodology. The former generated *de novo* transcript assemblies, while the latter directly mapped reads and junctions onto the canonical genome from ToxoDB (41). This also resulted in different methods to detect alternative splicing. The first study constructed full transcripts, identifying differences in isoforms at individual genetic loci. The latter focused purely on alternative splicing identification, observing alternative splicing events in the absence of transcript models merely by detection of incongruent bridging reads.

**Stage-specific alternative splicing in apicomplexans.** A unifying feature of all apicomplexan studies of alternative splicing is that most analysis is concentrated on only a few stages, for example the intraerythrocytic asexual stages of *P. falciparum* or tachyzoites of *T. gondii*. In multicellular organisms such as humans, a major hypothesized role of alternative splicing is to produce different protein transcripts in different tissues. In unicellular protists, the analogue to tissue specificity may be stage-specific alternative splicing, with different protein isoforms in different life stages. López-Barragán and colleagues again sequenced the *P. falciparum* transcriptome, attempting to add an additional stage, mosquito-infecting ookinetes (42). They identified 178 new alternatively spliced genes, although the authors identified alternative splicing from each stage independently, rather than observing changes in splicing between multiple stages. Due to contamination with mosquito RNA, they only managed to sequence a relatively small amount of ookinete-specific RNA (0.7 million reads compared to ~5 million reads for RNA-seq analyses of other apicomplexan stages), resulting in identification of only 15 additional alternatively spliced transcripts in this stage. A similar study of *Toxoplasma* also found different alternative splice events in proliferating tachyzoite stages (42 genes with alternative splicing detected using a *de novo* assembly detection method) compared to persistent bradyzoite stages (65 genes with alternative splicing) (43). However, as with the other work described above, this study examined splicing independently at each stage and did not test for statistically differential splicing between stages, which must be a priority for future multistage studies of splicing in Apicomplexa.

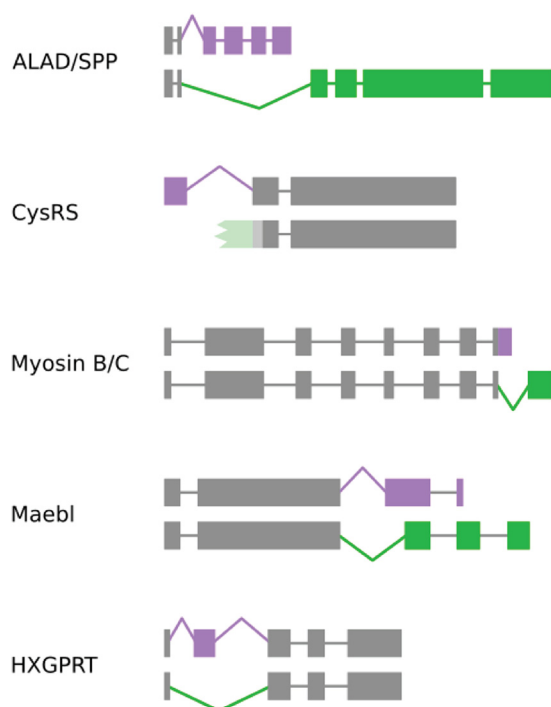
**Limitations in detection of alternative splicing.** Although the advent of Illumina sequencing-based RNA-seq has provided an abundance of sequence reads from which to detect alternative splicing, the short Illumina reads limit the detection of concurrent

alternative splicing events. Some apicomplexan genes undergo multiple alternative splicing events, which only occur simultaneously on specific transcript isoforms (44). Ideally, detection of these tandem events would include whether they occur on the same or different transcripts. However, detection of simultaneous events requires sequencing reads that traverse all alternative splicing events on a single transcript. Newer sequencing technology such as PacBio and Oxford Nanopore can produce longer reads, mitigating this problem, but currently there is a dearth of bioinformatic tools that adequately address these issues.

Bioinformatic limitations in analysis of splicing are often of major concern (see reference 45 for a review of available software). Even when software works well for general data sets, it is often written for human genomes, and hence not optimized for more compact genomes with frequent overlaps of 5' and 3' UTRs such as those found in apicomplexan parasites, which require substantial processing of sequencing data before it is useable (40). A final issue is that sequencing depth correlates with alternative splicing detection. Lower read depth is insufficient to detect the rarest transcripts, but as sequencing depth increases, less abundant isoforms are revealed, resulting in greater detection of alternative splicing (46–48). The depth of whole-transcriptome sequencing in Apicomplexa has generally not been as extensive as in metazoans, which will underestimate the extent of alternative splicing through failure to detect rarer isoforms. Estimates of percentages of genes with alternative splicing can therefore be understood only in the context of sequencing depth. This difficulty of detecting rare splicing events is also reflected in the absence of widespread standards or cutoffs for defining alternative splicing from RNA-seq reads. The most-used database for apicomplexan genomes, EuPathDB, collates transcriptomic experiments from diverse species performed at a range of sequencing depths but generally does not attempt to catalogue alternative splicing events. *Plasmodium falciparum*, the best curated of these parasites, has only sparse annotation (~100 genes with annotated alternative transcripts).

**What is the function of alternative splicing?** The functional significance of alternative splicing has remained contentious since its earliest detection. Are most alternative transcripts the result of a sloppy splicing apparatus, are they important sources of proteomic diversity, or is their primary importance instead in regulating transcript stability and protein levels? Early discoveries of alternative splicing were of a few specific genes, and thus, often associated with specific functional data. Hence, the original hypothesis was that alternative splicing mostly increased the diversity of protein isoforms by producing multiple biologically relevant proteins for each gene locus (49, 50). However, given the large amount of alternative splicing observed after the advent of RNA-seq, the validity of this assertion has been questioned (51). Recent proteomic studies have tried to address the relevance of alternative splicing by quantifying the protein output corresponding to alternative transcripts. For the vast majority of genes, a single protein isoform is most abundant; most other putative protein isoforms are unable to be detected by proteomic analyses (52, 53). Hence, some have argued that the function of alternative splicing is not proteomic diversity (51, 54). Conversely, it is plausible that many alternative protein isoforms are merely too low in abundance to detect (55). Subsequent proteomic analyses also suggest that limitations in peptide cleavage techniques in protein preparation (56) and sensitivity limits of some mass spectrometry detection techniques (57) may explain some of the low detection of proteins from alternative transcripts. Our view is that elucidation of this question likely requires application of proteomic surveys to more diverse samples (between cell types within organisms and between diverse evolutionary groups) as well as improved proteomic preparation, detection, and bioinformatic methodology.

Aside from its role in generating different protein isoforms, alternative splicing is likely to influence protein abundance through regulation of nonsense-mediated decay, and it is feasible that some transcript isoforms that do not code for proteins have as-yet-unknown regulatory functions (58). Such consequences of alternative splicing

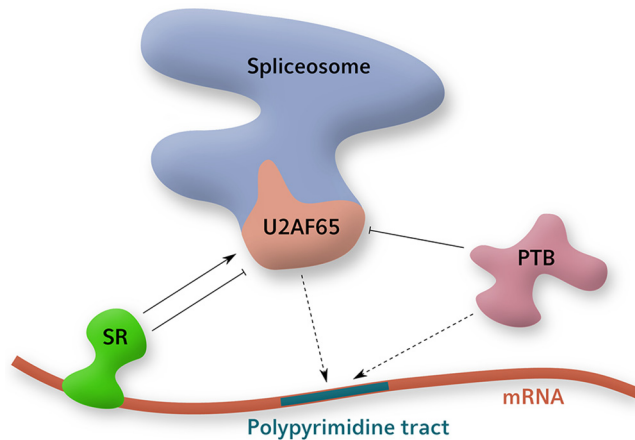


**FIG 2** Examples of alternatively spliced genes in Apicomplexa. Boxes represent exons, and thin lines represent introns. Gray or colored regions are common or variable between isoforms, respectively. The structures are not always drawn to scale to emphasize differences in isoforms. For ALAD/SPP (*P. falciparum*), alternative 3' splicing targets two unrelated proteins to the same location (60). CysRS (*T. gondii* and *P. falciparum*, the latter depicted) shows the opposite of ALAD/SPP. Intron retention (and possibly alternative transcriptional start site) results in one mature protein with two locations, the apicoplast and cytosol (63, 106). Lighter colors denote untranslated regions. Myosin B/C (*T. gondii*) shows similar functional results to CysRS but at the C-terminal end. Intron retention (and possibly alternative transcriptional stop site) results in one mature protein with two locations (61). For Maebl (*P. falciparum*), alternative 3' splicing results in a frameshift and different stop codons, encoding either soluble or membrane-bound variants (62). For HXGPRT (*T. gondii*), exon skipping induces localization in either the cytosol or inner membrane complex (44).

are consistent with the observed disparity between the transcriptome and proteome of apicomplexans (1–3). In addition to conflicts between mRNA and protein abundance, *P. falciparum* also displays some temporal discrepancies between polysome-associated mRNAs and steady-state RNAs (59). Bunnik and colleagues also found alternative splicing patterns specific to polysome-associated and non-polysome-associated pools, suggesting that this phenomenon might act as another level of posttranscriptional regulation (59).

Although the extent of alternative splicing that results in protein differences is yet to be settled on a whole-proteome scale, there are certainly a number of alternatively spliced genes in Apicomplexa that do code for novel and detectable protein isoforms. These include a plethora of splicing changes, such as repositioning of one splice site, intron retention, and exon skipping. The functional outcomes of these changes can also be diverse (Fig. 2). In the case of *Plasmodium* delta-aminolevulinic acid dehydratase (ALAD) and the stromal processing peptidase (SPP), two adjacent genes encode unrelated proteins. These genes share exons coding for an N-terminal targeting domain, which directs both into the apicoplast (60). In other examples, a single genetic locus generates a single mature protein, but in variants that possess different targeting information, the resulting isoforms are directed into multiple locations (44, 61–63). Further, some parasite surface proteins are alternatively spliced, which presumably generates distinct protein epitopes and may play a role in antigenic diversity (64–66). Finally, alternative splicing has been recently reported for the chloroquine resistance transporter (*PF*CRT); though the function of this splicing is yet to be determined (67).





**FIG 3** Alternative splicing factors can regulate spliceosomal components. Part of the spliceosome, U2AF65, binds to mRNA, mediating canonical splicing. SR proteins and PTB interact with neighboring sequences and can repress the binding of this component. SR proteins can also enrich this activity.

Thus, apicomplexan alternative splicing can be a highly specific process, which results in at least some increase to protein diversity in these parasites.

### REGULATION OF ALTERNATIVE SPLICING

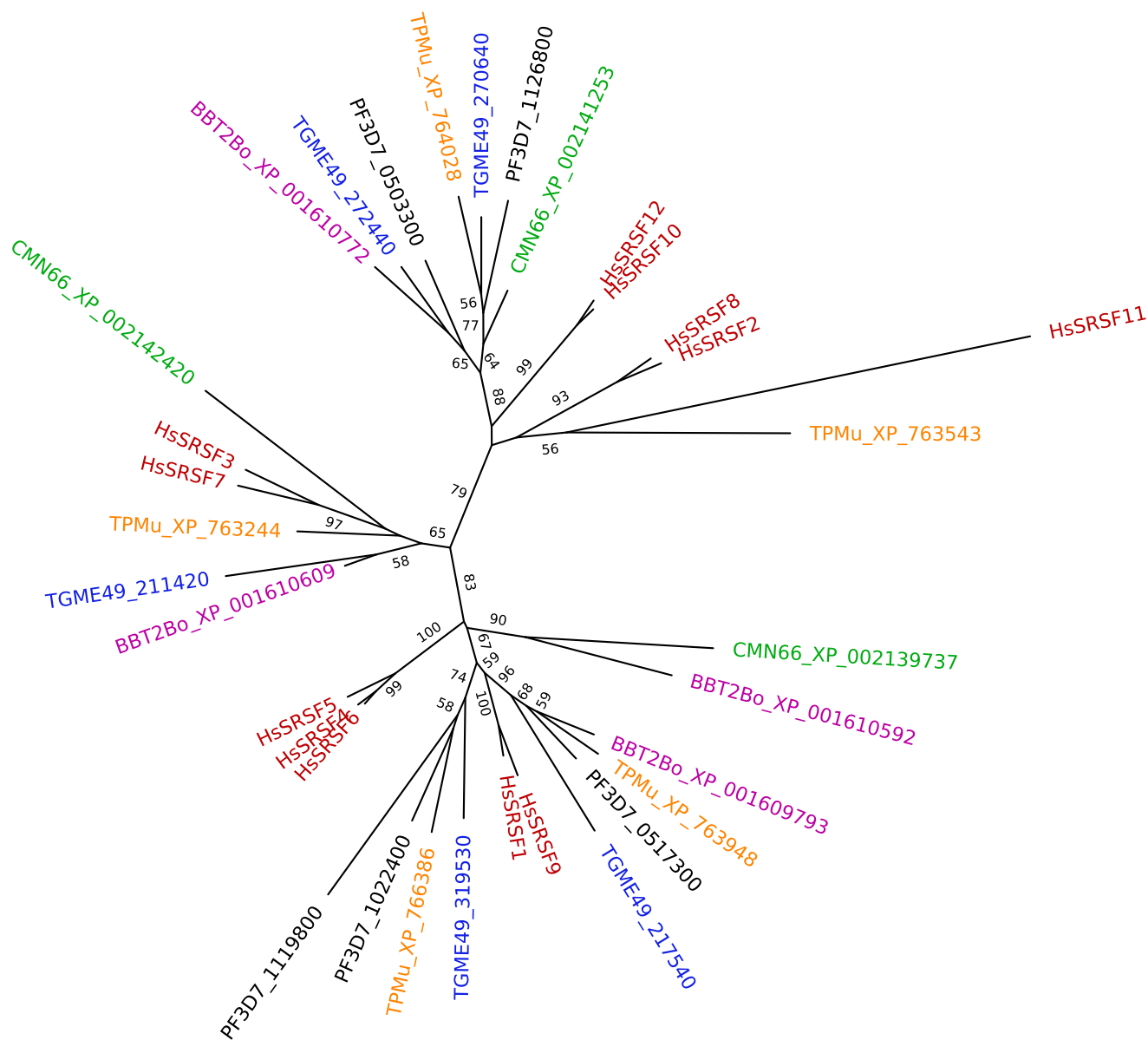
**Mechanism of alternative splicing in Metazoa.** In eukaryotes, numerous regulators can mediate alternative splicing. Some of the most studied alternative splicing factors belong to one of two groups: the serine/arginine-rich (SR) proteins, and the heterogeneous nuclear ribonucleoproteins (hnRNPs) (Fig. 3). The SR protein family includes the alternative splicing factor/splicing factor 2 (ASF/SF2 or SRSF1), with 12 SR proteins identified in humans in total. The hnRNP group includes the polypyrimidine tract-binding protein (PTB).

**SR proteins.** SR proteins contain an N-terminal RNA recognition motif, and a C-terminal arginine/serine-rich (RS) domain, which contains multiple arginine-serine dipeptide repeats (68). SR proteins localize to subnuclear regions called speckles (69); these regions are where splicing factors are assembled and stored. Speckles colocalize with sites of RNA transcription and splicing (70) and are distinct from the nucleolus and condensed chromatin (71). A subset of SR proteins, including both SRSF3 and ASF/SF2, dynamically shuttle between the nucleus and cytoplasm in human cells, with a very minor pool in the latter compartment (72).

The phosphorylated RS domain of SR proteins appears to be sufficient to act as a nuclear localization signal, although in ASF/SF2, synergistic relationships with the other domains confer greater efficiency of nuclear import (73). When phosphorylated, the RS domain binds to two proteins of the importin  $\beta$ /transportin family, transportin-SR and transportin-SR2, which mediate import into the nucleus (74, 75). In some cases, such as SRSF3 (also known as SRp20), the RS domain is sufficient for further targeting to subnuclear speckles, whereas in other cases, including ASF/SF2, this domain is insufficient for speckle localization (73).

SR proteins can both enhance and inhibit splicing, depending on motif location, although the specific mechanism of regulation is poorly understood. However, SR proteins can mediate the efficacy with which spliceosomal components bind to splice sites, in part dependent on the location of the SR protein's binding to RNA (68). SR proteins have additional roles in constitutive splicing, mRNA export, increasing mRNA stability, and regulation of translation (76). Knockdown or overexpression of ASF/SF2 has specifically been used as a means of perturbing alternative splicing, with changes in mRNA isoforms detected by qRT-PCR (77).

**Polypyrimidine tract-binding protein.** The polypyrimidine tract-binding protein (PTB) localizes primarily to the nucleus (excluding the nucleolus), although it also



**FIG 4** A phylogenetic tree inferred for putative apicomplexan and human SR proteins. Numbers denote bootstrap values above 50 for a neighbor-joining distance tree. Sequences from *Plasmodium falciparum* (PF) (black), *Babesia bovis* (BB) (purple), *Toxoplasma gondii* (TG) (blue), *Theileria parva* (TP) (orange), *Cryptosporidium muris* (CM) (green), and *Homo sapiens* (Hs) (red) are indicated. Some apicomplexan SRs group strongly with one or more human SR counterparts, whereas others are clearly paralogous or only weakly allied with a human SR.

shuttles to the cytoplasm in a minor pool that is enhanced during cell stress (78). PTB can also regulate alternative splicing, most commonly repressing splicing by direct competition with the spliceosome. The canonical binding site for PTB is the polypyrimidine tract on introns, in direct competition with the spliceosomal component U2AF65 (79). Both SR proteins and PTB are regulated by phosphorylation, which influences localization, splice site binding, and splicing activity (80–82).

**Mechanism of alternative splicing in Apicomplexa.** We have previously identified *Toxoplasma* and *Plasmodium* homologues to human and *Arabidopsis* SR proteins (40), and here we also present additional putative SR proteins from *Babesia bovis*, *Cryptosporidium muris*, and *Theileria parva*. A phylogenetic tree inferred from these sequences reveals a mixed pattern of evolution with some apicomplexan SR proteins presenting one-to-one orthologous relationships and others showing distant paralogous relationships (Fig. 4). Annotated apicomplexan parasites include a single PTB (PF3D7\_0606500



in *P. falciparum*, TGME49\_090660 in *T. gondii*). There have been several published studies on SR proteins in apicomplexans, but none to our knowledge on PTB orthologues.

The first identification of an apicomplexan SR-protein homologue as a potential alternative splicing factor was in *T. gondii*, when Gubbels and colleagues localized part of the RS-rich domain of TGME49\_319530 (SF2, TgSR1) (83). This localized to a sub-nuclear region, in a single punctum adjacent to the nucleolus, and colocalized with MAb104, the antibody that initially defined membership of the SR protein family in human cells (84). As mentioned above, the human RS domain is sufficient for nuclear localization, although it is not always sufficient for correct subnuclear targeting. Later, we showed the full-length *T. gondii* TgSR3 (TGME49\_211420) localizing to a more diffuse, mottled nuclear pattern (40), that is reminiscent of the nuclear speckles seen for metazoan SR proteins (68).

Fan and colleagues gave the first evidence that apicomplexan SR homologues might have RNA-binding activity from studies of *P. falciparum* (85). SR proteins SFRS4 (PF3D7\_1022400) and PfSR1 (PF3D7\_0517300) were shown to be methylated *in vitro* by *P. falciparum* arginine N-methyltransferase 1, an enzyme that regulates RNA-binding proteins. Similar activity was shown for GBP2 (PF3D7\_1006800) (85), which shows some sequence similarity to canonical SR proteins, but is located on a long branch in the apicomplexan SR phylogeny (40).

Dixit and colleagues localized a *P. falciparum* SR protein using antibodies (86). The article is unclear on which specific SR protein, as it lists two gene IDs: PF3D7\_0503300 (SRSF12) and PF3D7\_0517300 (PfSR1). The SR protein localizes primarily to the nucleus of rings, the nuclear periphery and cytoplasm of mature trophozoites, and mainly the cytoplasm of schizonts and gametocytes. In other eukaryotes, SR proteins are regulated by the action of upstream kinases. These kinases phosphorylate the SR proteins' arginine/serine-rich (RS) domain to modulate both activity and localization. Dixit and colleagues show that the RS domain of their SR protein is phosphorylated *in vitro* by SRPK1 (also known as PFCLK-4, PF3D7\_0302100), a kinase identified by the authors and implicated in RNA splicing (86). Phosphorylation of this putative splicing factor increased RNA binding by 80%, although this protein was not recognized by the aforementioned MAb104 (86). Similarly, PfSR1 was later shown to shuttle between cytoplasm and nucleus, adjacent to nuclear pore complexes (87, 88). The authors further demonstrated that the RS domain was necessary for nuclear localization (87). Ablation of the *P. berghei* SRPK1 orthologue PBANKA\_0401100 caused minor growth defects in asexual blood stages but completely prevented exflagellation of male gametocytes (89). In addition, the SR protein ASF1 (PF3D7\_1119800) was shown to be a substrate *in vitro* of both PFCLK-1 and PFCLK-2, two homologues of a yeast SR protein kinase (90). A later study confirmed localization of SFRS4 and SRSF12, as well as SF-1 (also known as SF1, PF3D7\_1321700) to the nucleus of *P. falciparum* asexual blood stages, primarily in trophozoites (91). No portion was detected in the cytoplasm, although the latter two proteins were also detected in gametocyte nuclei (91). The authors also found that these putative SR proteins were phosphorylated *in vitro* by various members of the four PFCLK kinases (PFCLK-1 to -4) (91).

Hence, it is clear that apicomplexan SR homologues often localize to the nucleus and possibly also to the cytoplasm, are capable of binding to RNA (92), and are, at least in part, regulated by SR kinases (SRPKs/CLKs). There is also a potential link with stage specificity in some cases. The first direct evidence for an apicomplexan SR-protein homologue regulating alternative splicing was demonstrated in *P. falciparum*, when Eshar and colleagues showed that recombinant PfSR1 could regulate alternative splicing in two mammalian reporter systems, both *in vitro* and *in vivo* (87). Overexpression of PfSR1 reduced the growth rate of the asexual blood stage of parasites and changed alternative splicing patterns in each of three *P. falciparum* genes assayed based on qRT-PCR results (87).

In a separate study, overexpression of PfSR1 led to alternative splicing changing in one gene that was tested, and the total level of expression was altered in other genes

(88). While these overall expression changes may arise from the other roles of SR proteins mentioned above (such as RNA stability and export), they may also be caused by pleiotropic effects that arose from the observed growth defects. To determine the targets of *Pf*SR1, Eshar and colleagues later used immunoprecipitation and microarray assays (RIP-Chip) to detect binding of 64 transcripts to this SR protein (88). Two motifs were identified from this pool, GAUGAUGA and GUUGA, both somewhat reminiscent of the human SRSF1 binding site (88). Despite detection of *Pf*SR1 binding to transcripts from only a few genes, the SBM-1 motif was detected in 89% of all *P. falciparum* genes ( $P$  value for motif match  $< 0.001$ ), and the SBM-2 motif in 78% (88). Further, 97% of these motifs were found in exons, and often within genes that contained only a single exon (88). Thus, while these motifs may be required for binding, their near ubiquity in all transcripts in general means that they cannot be predictive factors for *Pf*SR1-mediated splice regulation, at least as detected using this protocol.

As mentioned above, the relative scarcity of alternatively spliced genes in apicomplexan parasites compared to discoveries for model animals might be due to the low depth of high-throughput sequencing (or limited number of stages sampled). However, perturbation of putative alternative splicing factors was a potential way to artificially boost the number of alternatively spliced genes affected by these regulators to a level of easy detection. Hence, we combined pertinent technologies by overexpressing an SR protein in *T. gondii* (TgSR3, TGME49\_211420) followed by RNA-seq (40). Overexpression prevented growth in tachyzoite-stage parasites and was shown to significantly modify alternative splicing in one of three genes assayed by qRT-PCR (40). After analysis of RNA-seq data, we detected changes in alternative splicing in more than 1,200 genes after overexpression of TgSR3 (40).

This stark difference between the scale of splicing perturbation seen with TgSR3 overexpression in *Toxoplasma*, and the number of genes putatively regulated by *P. falciparum* *Pf*SR1, may be caused by several reasons. First, genes in *T. gondii* have more exons than genes in *P. falciparum*, as described above. This may correlate with a greater proclivity to alternative splicing. Second, Eshar and colleagues used a conservative threshold for significance of three standard errors ( $P$  value = 0.0027) (88), whereas we used a  $P$  value of 0.05 (40). Finally, transcripts bind to SR proteins only transiently, whereas the effects on alternative splicing can be detected for the life of that mRNA, resulting in a greater number of genes detected. Further, the presence of *Pf*SR1 motifs in the vast majority of *P. falciparum* genes may suggest that *Pf*SR1 binds to these other genes at different time points.

Similarly to Eshar and colleagues (88), we also observed changes in the total expression level of some genes after SR perturbation. However, this was seen in only 4.6% of the transcripts detected, at 4 h after induction (40). We observed considerable increases at later time points, with little consistency in the genes affected, suggesting that at least some of these changes in gene expression were merely pleiotropic effects. It is therefore possible that many of the observed gene expression changes result not from direct interaction with SR proteins but from downstream effects. An alternative explanation for SR-induced changes in transcript abundance is that the SR-mediated changes in splicing modulate the inclusion of premature stop codons, and the resultant nonsense-mediated decay alters transcript persistence and therefore abundance. This is not readily discernible from SR-induced changes in transcription or from direct SR stabilization of transcripts, though technologies that allow measurements of mRNA dynamics may allow us to address this question better (93, 94). Nonetheless, the small number of transcripts whose abundance is significantly altered, compared to transcripts with changes in alternative splicing, implies that the primary function of the apicomplexan SR proteins is regulation of alternative splicing.

### SPLICING/ALTERNATIVE SPLICING AS A DRUG TARGET

**Kinases regulating splicing and alternative splicing.** Appropriate regulation of alternative splicing is required for survival and proliferation of apicomplexan parasites (40, 87, 89), so compounds that interfere with this process are plausible antiparasitic

drug candidates. There is substantial precedence for chemical modulation of splicing in animal systems, where a large variety of compounds that impact alternative splicing have been reported. The majority of these are compounds that inhibit the SRPK/CLK kinases that control the SR alternative splicing regulators. The complexity of downstream consequences of inhibiting SR kinases defies easy categorization: in some cases, chemical inhibition generates aberrant alternative splicing, whereas in some disease states, inhibition of those kinases restores normal splicing (95). Examples of both scenarios have been proposed as therapeutic drugs. Some SRPK/CLK inhibitors generate aberrant alternative splicing that protects against Duchenne muscular dystrophy by promoting a splice event that skips an inherited deleterious mutation (96), whereas other CLK inhibitors may protect against exon-skipping mutations in Alzheimer's disease (97).

Conceptually, the general simplicity of consensus splicing elements or motifs presents a major challenge in human drug discovery, because disruption of a master splice regulator may have numerous undesired splicing effects. While the CLK inhibitors identified by Sako and colleagues have minimal off-target splicing effects (0.3%) (96), transcriptomic studies are available for very few of the SRPK or CLK inhibitors described in the literature. One cautionary example was the recent termination of a clinical trial (ClinicalTrials.gov identifier NCT02240355 [<https://clinicaltrials.gov/ct2/show/results/NCT02240355>]) of a SMN2 splicing modifier due to concerns with off-target effects (107). However, the scenario for anti-infectives in this space is less challenging— inhibitors with pleiotropic effects on parasite alternative splicing can be plausible drugs as long as they are selective for parasite rather than human proteins. Indeed, SRPKs/CLKs have been validated as potential drug targets in *Plasmodium*. Agarwal and colleagues (90) showed that the *Pf*CLK-1 and *Pf*CLK-2 were refractory to gene disruption. This was independently confirmed by Solyakov and colleagues (98), who conducted a global kinomic and phospho-proteomic study. Kern and colleagues (91) later characterized inhibitors of *Pf*CLKs, which they demonstrated to potently block parasite asexual replication and gametogenesis (91). These compounds were initially identified based on their inhibition of equivalent human kinases, and it is unclear whether there are strong possibilities for selective inhibition of parasite, compared to human, CLKs.

In addition to CLKs, many other potential targets involved in constitutive or alternative splicing machinery have been identified. Following the efforts of Solyakov and colleagues in the global kinomic study, *P. falciparum* casein kinase 1 (*Pf*CK1) was subsequently found to be implicated in the splicing pathway via interatomic experiments (99). *Pf*CK1 is essential (98), but the degree to which splicing disruptions are responsible for this phenotype has not been characterized. Further transcriptomic analyses in parasites with *Pf*CK1 inhibited or knocked down would be of interest. In *T. gondii*, a single point mutation in a previously unidentified protein (*Tg*RRM1) severely affected splicing in a temperature-restrictive mutant (100). These mutants conditionally arrest in the G<sub>1</sub> phase of replication, leading to subsequent death (100).

**Morpholino modulation of splicing.** Given the challenges of identifying compounds that inhibit apicomplexan but not human splice-regulating kinases, morpholino modulation of splicing may present an alternative means of generating specific inhibition. A morpholino oligomer-based approach has been recently developed to target mRNA splicing in *Plasmodium* (101). Morpholino conjugates have been previously used to target mRNAs for degradation or disrupted splicing in metazoans (102), and one morpholino that results in modified alternative splicing has been recently clinically approved for treatment of Duchenne muscular dystrophy (103). Garg and colleagues (101) utilized morpholino conjugates to degrade or inhibit splicing of *P. falciparum* PMT and CRT mRNAs, resulting in strong inhibition of parasite growth. The authors further showed that the morpholino conjugates targeting the chloroquine resistance transporter CRT partially restore chloroquine susceptibility in chloroquine-resistant parasites (101). This approach is of particular interest because splicing of specific genes can be targeted and applied to larger-scale functional analyses. The

specificity of morpholino technology plausibly allows selectivity of therapy that would target parasite splicing processes without affecting human genes.

**Concluding remarks.** Several RNA-seq experiments in apicomplexans detect alternative splicing for relatively few genes (only several percent) compared to well-characterized metazoans (>90%), but this may be confounded by use of lower sequencing depth, poor suitability of the bioinformatic tools employed to analyze these genomes, and sequencing of only a few life cycle stages. Manual, lower-throughput methods detect alternative splicing for a greater proportion (~15 to 20%) of assayed genes, although the selection of such genes may be biased. Our own RNA-seq analyses using purpose-built tools to detect alternative splicing (e.g., DEXSeq [104], Junction Juror [105]) also suggest considerably higher levels of alternative splicing in *Toxoplasma* at least. A research priority will now be to address the extent of alternative splicing between the diverse life stages of these parasites. Only a small proportion of the alternative transcripts detected so far look likely to result in alternative proteins, so proteome diversity does not appear to be the main outcome of alternative splicing in apicomplexans. Alternative splicing may instead be involved in nonsense-mediated decay or transcript turnover, which may partially explain the observed discrepancies between transcript and protein abundance. Nevertheless, perturbation of alternative splicing is detrimental to parasites, making it a worthy drug target to pursue. Future research should help resolve the extent to which these apparently noncoding splice forms are required for parasite growth and differentiation.

## ACKNOWLEDGMENTS

L.M.Y. was supported by an Australian Postgraduate Award. G.I.M. gratefully acknowledges a Program Grant from the National Health and Medical Research Council (Australia), a Discovery Project from the Australian Research Council, and an Australian Laureate Fellowship from the Australian Research Council. S.A.R. is supported by an Australian Research Council Discovery Project grant (DP160100389) and an NHMRC RD Wright Biomedical fellowship (APP1062504).

## REFERENCES

1. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ, Yates JR, Winzeler EA. 2004. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res* 14:2308–2318. <https://doi.org/10.1101/gr.2523904>.
2. Foth BJ, Zhang N, Mok S, Preiser PR, Bozdech Z. 2008. Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites. *Genome Biol* 9:R177. <https://doi.org/10.1186/gb-2008-9-12-r177>.
3. Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z. 2011. Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 10:M110.006411. <https://doi.org/10.1074/mcp.M110.006411>.
4. Mair GR, Lasonder E, Garver LS, Franke-Fayard BM, Carret CK, Wiegant JC, Dirks RW, Dimopoulos G, Janse CJ, Waters AP. 2010. Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog* 6:e1000767. <https://doi.org/10.1371/journal.ppat.1000767>.
5. Mair GR, Braks JA, Garver LS, Wiegant JC, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP. 2006. Regulation of sexual development of *Plasmodium* by translational repression. *Science* 313: 667–669. <https://doi.org/10.1126/science.1125129>.
6. Gomes-Santos CS, Braks J, Prudencio M, Carret C, Gomes AR, Pain A, Feltwell T, Khan S, Waters A, Janse C, Mair GR, Mota MM. 2011. Transition of *Plasmodium* sporozoites into liver stage-like forms is regulated by the RNA binding protein Pumilio. *PLoS Pathog* 7:e1002046. <https://doi.org/10.1371/journal.ppat.1002046>.
7. Miao J, Li J, Fan Q, Li X, Li X, Cui L. 2010. The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite *Plasmodium falciparum*. *J Cell Sci* 123:1039–1049. <https://doi.org/10.1242/jcs.059824>.
8. Turunen JJ, Niemela EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* 4:61–76. <https://doi.org/10.1002/wrna.1141>.
9. Suvorova ES, White MW. 2014. Transcript maturation in apicomplexan parasites. *Curr Opin Microbiol* 20:82–87. <https://doi.org/10.1016/j.mib.2014.05.012>.
10. Eliana C, Javier E, Moises W. 2011. *Plasmodium falciparum* spliceosomal RNAs: 3' and 5' end processing. *Acta Trop* 117:105–108. <https://doi.org/10.1016/j.actatropica.2010.10.008>.
11. Bawankar P, Shaw PJ, Sardana R, Babar PH, Patankar S. 2010. 5' and 3' end modifications of spliceosomal RNAs in *Plasmodium falciparum*. *Mol Biol Rep* 37:2125–2133. <https://doi.org/10.1007/s11033-009-9682-4>.
12. Hossain M, Sharma S, Korde R, Kanodia S, Chugh M, Rawat K, Malhotra P. 2013. Organization of *Plasmodium falciparum* spliceosomal core complex and role of arginine methylation in its assembly. *Malar J* 12:333. <https://doi.org/10.1186/1475-2875-12-333>.
13. Cornillot E, Hadj-Kaddour K, Dassouli A, Noel B, Ranwez V, Vacherie B, Augagneur Y, Brès V, Duclos A, Randazzo S, Carcy B, Debierre-Grockiego F, Delbecq S, Moubri-Ménage K, Shams-Eldin H, Usmani-Brown S, Bringaud F, Wincker P, Vivarès CP, Schwarz RT, Schettters TP, Krause PJ, Gorenflot A, Berry V, Barbe V, Ben Mamoun C. 2012. Sequencing of the smallest Apicomplexan genome from the human pathogen *Babesia microti*. *Nucleic Acids Res* 40:9102–9114. <https://doi.org/10.1093/nar/gks700>.
14. Blazejewski T, Nursimulu N, Pszeny V, Dangoudoubiyam S, Namasiyayam S, Chiasson MA, Chessman K, Tonkin M, Swapna LS, Hung SS, Bridgers J, Ricklefs SM, Boulanger MJ, Dubey JP, Porcella SF, Kissinger JC, Howe DK, Grigg ME, Parkinson J. 2015. Systems-based analysis of the *Sarcocystis neurona* genome identifies pathways that contribute to a heteroxenous life cycle. *mBio* 6:e02445-14. <https://doi.org/10.1128/mBio.02445-14>.



15. Heiges M, Wang H, Robinson E, Aurrecochea C, Gao X, Kaluskar N, Rhodes P, Wang S, He C-Z, Su Y, Miller J, Kraemer E, Kissinger JC. 2006. CryptoDB: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids Res* 34:D419–D422. <https://doi.org/10.1093/nar/gkj078>.
16. Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74:3171–3175.
17. Chow LT, Gelinis RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12:1–8.
18. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–476. <https://doi.org/10.1038/nature07509>.
19. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413–1415. <https://doi.org/10.1038/ng.259>.
20. Fiszbein A, Giono LE, Quaglino A, Berardino BG, Sigaut L, von Bilderling C, Schor IE, Steinberg JHE, Rossi M, Pietrasanta LI, Caramelo JJ, Srebrow A, Kornblihtt AR. 2016. Alternative splicing of G9a regulates neuronal differentiation. *Cell Rep* 14:2797–2808. <https://doi.org/10.1016/j.celrep.2016.02.063>.
21. van Lin LH, Pace T, Janse CJ, Birago C, Ramesar J, Picci L, Ponzi M, Waters AP. 2001. Interspecies conservation of gene order and intron-exon structure in a genomic locus of high gene density and complexity in *Plasmodium*. *Nucleic Acids Res* 29:2059–2068. <https://doi.org/10.1093/nar/29.10.2059>.
22. Chakrabarti D, Reddy GR, Dame JB, Almira EC, Laipis PJ, Ferl RJ, Yang TP, Rowe TC, Schuster SM. 1994. Analysis of expressed sequence tags from *Plasmodium falciparum*. *Mol Biochem Parasitol* 66:97–104.
23. Wan KL, Blackwell JM, Ajioka JW. 1996. *Toxoplasma gondii* expressed sequence tags: insight into tachyzoite gene expression. *Mol Biochem Parasitol* 75:179–186.
24. de Vries E, Corton C, Harris B, Cornelissen AWCA, Berriman M. 2006. Expressed sequence tag (EST) analysis of the erythrocytic stages of *Babesia bovis*. *Vet Parasitol* 138:61–74. <https://doi.org/10.1016/j.vetpar.2006.01.040>.
25. Strong WB, Nelson RG. 2000. Preliminary profile of the *Cryptosporidium parvum* genome: an expressed sequence tag and genome survey sequence analysis. *Mol Biochem Parasitol* 107:1–32.
26. Howe DK. 2001. Initiation of a *Sarcocystis neurona* expressed sequence tag (EST) sequencing project: a preliminary report. *Vet Parasitol* 95:233–239.
27. Ng ST, Sanusi Jangi M, Shirley MW, Tomley FM, Wan KL. 2002. Comparative EST analyses provide insights into gene expression in two asexual developmental stages of *Eimeria tenella*. *Exp Parasitol* 101:168–173.
28. Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S. 2007. Comparisite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. *Nucleic Acids Res* 35:D431–D438. <https://doi.org/10.1093/nar/gkl1039>.
29. Wakaguri H, Suzuki Y, Katayama T, Kawashima S, Kibukawa E, Hiranuka K, Sasaki M, Sugano S, Watanabe J. 2009. Full-Malaria/Parasites and Full-Arthropods: databases of full-length cDNAs of parasites and arthropods, update 2009. *Nucleic Acids Res* 37:D520–D525. <https://doi.org/10.1093/nar/gkn856>.
30. Yamagishi J, Wakaguri H, Yokoyama N, Yamashita R, Suzuki Y, Xuan X, Igarashi I. 2014. The *Babesia bovis* gene and promoter model: an update from full-length EST analysis. *BMC Genomics* 15:678. <https://doi.org/10.1186/1471-2164-15-678>.
31. Cui L, Fan Q, Hu Y, Karamycheva SA, Quackenbush J, Khuntirat B, Sattabongkot J, Carlton JM. 2005. Gene discovery in *Plasmodium vivax* through sequencing of ESTs from mixed blood stages. *Mol Biochem Parasitol* 144:1–9. <https://doi.org/10.1016/j.molbiopara.2005.05.016>.
32. Kooij TWA, Franke-Fayard B, Renz J, Kroeze H, van Dooren MW, Ramesar J, Augustijn KD, Janse CJ, Waters AP. 2005. *Plasmodium berghei* alpha-tubulin II: a role in both male gamete formation and asexual blood stages. *Mol Biochem Parasitol* 144:16–26. <https://doi.org/10.1016/j.molbiopara.2005.07.003>.
33. Lu F, Jiang H, Ding J, Mu J, Valenzuela JG, Ribeiro JMC, Su X-Z. 2007. cDNA sequences reveal considerable gene prediction inaccuracy in the *Plasmodium falciparum* genome. *BMC Genomics* 8:255. <https://doi.org/10.1186/1471-2164-8-255>.
34. Iriko H, Jin L, Kaneko O, Takeo S, Han E-T, Tachibana M, Otsuki H, Torii M, Tsuboi T. 2009. A small-scale systematic analysis of alternative splicing in *Plasmodium falciparum*. *Parasitol Int* 58:196–199. <https://doi.org/10.1016/j.parint.2009.02.002>.
35. Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C, Linás M. 2010. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol Microbiol* 76:12–24. <https://doi.org/10.1111/j.1365-2958.2009.07026.x>.
36. Sorber K, Dimon MT, DeRisi JL. 2011. RNA-Seq analysis of splicing in *Plasmodium falciparum* uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. *Nucleic Acids Res* 39:3820–3835. <https://doi.org/10.1093/nar/gkq1223>.
37. Zhu L, Mok S, Imwong M, Jaidee A, Russell B, Nosten F, Day NP, White NJ, Preiser PR, Bozdech Z. 2016. New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. *Sci Rep* 6:20498. <https://doi.org/10.1038/srep20498>.
38. Li J, Cai B, Qi Y, Zhao W, Liu J, Xu R, Pang Q, Tao Z, Hong L, Liu S, Leerkes M, Quiñones M, Su X-Z. 2016. UTR introns, antisense RNA and differentially spliced transcripts between *Plasmodium yoelii* subspecies. *Malar J* 15:30. <https://doi.org/10.1186/s12936-015-1081-9>.
39. Hassan MA, Melo MB, Haas B, Jensen KDC, Saeij JJP. 2012. De novo reconstruction of the *Toxoplasma gondii* transcriptome improves on the current genome annotation and reveals alternatively spliced transcripts and putative long non-coding RNAs. *BMC Genomics* 13:696. <https://doi.org/10.1186/1471-2164-13-696>.
40. Yeoh LM, Goodman CD, Hall NE, van Dooren GG, McFadden GI, Ralph SA. 2015. A serine-arginine-rich (SR) splicing factor modulates alternative splicing of over a thousand genes in *Toxoplasma gondii*. *Nucleic Acids Res* 43:4661–4675. <https://doi.org/10.1093/nar/gkv311>.
41. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ, Jr, Wang H, Brunk BP. 2007. ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res* 36:D553–D556. <https://doi.org/10.1093/nar/gkm981>.
42. López-Barragán MJ, Lemieux J, Quiñones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su X-Z. 2011. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* 12:587. <https://doi.org/10.1186/1471-2164-12-587>.
43. Chen LF, Han XL, Li FX, Yao YY, Fang JP, Liu XJ, Li XC, Wu K, Liu M, Chen XG. 2018. Comparative studies of *Toxoplasma gondii* transcriptomes: insights into stage conversion based on gene expression profiling and alternative splicing. *Parasit Vectors* 11:402. <https://doi.org/10.1186/s13071-018-2983-5>.
44. Chaudhary K, Donald RGK, Nishi M, Carter D, Ullman B, Roos DS. 2005. Differential localization of alternatively spliced hypoxanthine-xanthine-guanine phosphoribosyltransferase isoforms in *Toxoplasma gondii*. *J Biol Chem* 280:22053–22059. <https://doi.org/10.1074/jbc.M503178200>.
45. Alamancos GP, Agirre E, Eyraes E. 2014. Methods to study splicing from high-throughput RNA sequencing data. *Methods Mol Biol* 1126:357–397. [https://doi.org/10.1007/978-1-62703-980-2\\_26](https://doi.org/10.1007/978-1-62703-980-2_26).
46. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 15:121–132. <https://doi.org/10.1038/nrg3642>.
47. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13. <https://doi.org/10.1186/s13059-016-0881-8>.
48. Liu Y, Ferguson JF, Xue C, Silverman IM, Gregory B, Reilly MP, Li M. 2013. Evaluating the impact of sequencing depth on transcriptome profiling in human adipose. *PLoS One* 8:e66883. <https://doi.org/10.1371/journal.pone.0066883>.
49. Breitbart RE, Andreadis A, Nadal-Ginard B. 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* 56:467–495. <https://doi.org/10.1146/annurev.bi.56.070187.002343>.
50. Andreadis A, Gallego ME, Nadal-Ginard B. 1987. Generation of protein isoform diversity by alternative splicing: mechanistic and biological implications. *Annu Rev Cell Biol* 3:207–242. <https://doi.org/10.1146/annurev.cb.03.110187.001231>.
51. Tress ML, Abascal F, Valencia A. 2017. Alternative splicing may not be the key to proteome complexity. *Trends Biochem Sci* 42:98–110. <https://doi.org/10.1016/j.tibs.2016.08.008>.

52. Ezkurdia I, Rodriguez JM, Carrillo-de Santa Pau E, Vázquez J, Valencia A, Tress ML. 2015. Most highly expressed protein-coding genes have a single dominant isoform. *J Proteome Res* 14:1880–1887. <https://doi.org/10.1021/pr501286b>.
53. Abascal F, Ezkurdia I, Rodriguez-Rivas J, Rodriguez JM, del Pozo A, Vázquez J, Valencia A, Tress ML. 2015. Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput Biol* 11:e1004325. <https://doi.org/10.1371/journal.pcbi.1004325>.
54. Tress ML, Abascal F, Valencia A. 2017. Most alternative isoforms are not functionally important. *Trends Biochem Sci* 42:408–410. <https://doi.org/10.1016/j.tibs.2017.04.002>.
55. Blencowe BJ. 2017. The relationship between alternative splicing and proteomic complexity. *Trends Biochem Sci* 42:407–408. <https://doi.org/10.1016/j.tibs.2017.04.001>.
56. Wang X, Codreanu SG, Wen B, Li K, Chambers MC, Liebler DC, Zhang B. 2018. Detection of proteome diversity resulted from alternative splicing is limited by trypsin cleavage specificity. *Mol Cell Proteomics* 17:422–430. <https://doi.org/10.1074/mcp.RA117.000155>.
57. Liu Y, Gonzalez-Porta M, Santos S, Brazma A, Marioni JC, Aebersold R, Venkitesan AR, Wickramasinghe VO. 2017. Impact of alternative splicing on the human proteome. *Cell Rep* 20:1229–1241. <https://doi.org/10.1016/j.celrep.2017.07.025>.
58. Lareau LF, Brenner SE. 2015. Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* 32:1072–1079. <https://doi.org/10.1093/molbev/msv002>.
59. Bunnik EM, Chung D-WD, Hamilton M, Ponts N, Saraf A, Prudhomme J, Flensburg L, Le Roch KG. 2013. Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome Biol* 14:R128. <https://doi.org/10.1186/gb-2013-14-11-r128>.
60. van Dooren GG, Su V, D'Ombrian MC, McFadden GI. 2002. Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J Biol Chem* 277:23612–23619. <https://doi.org/10.1074/jbc.M201748200>.
61. Delbac F, Sängler A, Neuhaus EM, Stratmann R, Ajioka JW, Torsell C, Herm-Götz A, Tomavo S, Soldati T, Soldati D. 2001. *Toxoplasma gondii* myosins B/C: one gene, two tails, two localizations, and a role in parasite division. *J Cell Biol* 155:613–623. <https://doi.org/10.1083/jcb.200012116>.
62. Singh N, Preiser P, Réna L, Balu B, Barnwell J, Blair P, Jarra W, Voza T, Landau I, Adams JH. 2004. Conservation and developmental control of alternative splicing in *maeb1* among malaria parasites. *J Mol Biol* 343:589–599. <https://doi.org/10.1016/j.jmb.2004.08.047>.
63. Pham JS, Sakaguchi R, Yeoh LM, De Silva NS, McFadden GI, Hou Y-M, Ralph SA. 2014. A dual-targeted aminoacyl-tRNA synthetase in *Plasmodium falciparum* charges cytosolic and apicoplast tRNACys. *Biochem J* 458:513–523. <https://doi.org/10.1042/BJ20131451>.
64. Sutherland CJ. 2001. Stevor transcripts from *Plasmodium falciparum* gametocytes encode truncated polypeptides. *Mol Biochem Parasitol* 113:331–335.
65. Blythe JE, Suretheran T, Preiser PR. 2004. STEVOR—a multifunctional protein? *Mol Biochem Parasitol* 134:11–15.
66. Fonager J, Cunningham D, Jarra W, Koernig S, Henneman AA, Langhorne J, Preiser P. 2007. Transcription and alternative splicing in the yir multigene family of the malaria parasite *Plasmodium yoelii*: identification of motifs suggesting epigenetic and post-transcriptional control of RNA expression. *Mol Biochem Parasitol* 156:1–11. <https://doi.org/10.1016/j.molbiopara.2007.06.006>.
67. Gadalla NB, Malmberg M, Adam I, Oguike MC, Beshir K, Elzaki S-E, Mukhtar I, Gadalla AA, Warhurst DC, Ngasala B, Mårtensson A, El-Sayed BB, Gil JP, Sutherland CJ. 2015. Alternatively spliced transcripts and novel pseudogenes of the *Plasmodium falciparum* resistance-associated locus pfcr1 detected in East African malaria patients. *J Antimicrob Chemother* 70:116–123. <https://doi.org/10.1093/jac/dku358>.
68. Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol* 10:242. <https://doi.org/10.1186/gb-2009-10-10-242>.
69. Spector DL. 1993. Macromolecular domains within the cell nucleus. *Annu Rev Cell Biol* 9:265–315. <https://doi.org/10.1146/annurev.cb.09.110193.001405>.
70. Spector DL. 1996. Nuclear organization and gene expression. *Exp Cell Res* 229:189–197. <https://doi.org/10.1006/excr.1996.0358>.
71. Roth MB, Murphy C, Gall JG. 1990. A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J Cell Biol* 111:2217–2223.
72. Cáceres JF, Sreaton GR, Krainer AR. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev* 12:55–66.
73. Cáceres JF, Misteli T, Sreaton GR, Spector DL, Krainer AR. 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138:225–238.
74. Kataoka N, Bachorik JL, Dreyfuss G. 1999. Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol* 145:1145–1152.
75. Lai MC, Lin RI, Huang SY, Tsai CW, Tarn WY. 2000. A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem* 275:7950–7957.
76. Huang Y, Steitz JA. 2005. SRproteins along a messenger's journey. *Mol Cell* 17:613. <https://doi.org/10.1016/j.molcel.2005.02.020>.
77. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* 14:185–193. <https://doi.org/10.1038/nsmb1209>.
78. Sawicka K, Bushell M, Spriggs KA, Willis AE. 2008. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans* 36:641–647. <https://doi.org/10.1042/BST0360641>.
79. Saullière J, Sureau A, Expert-Bezançon A, Marie J. 2006. The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol Cell Biol* 26:8755–8769. <https://doi.org/10.1128/MCB.00893-06>.
80. Mermoud JE, Cohen P, Lamond AI. 1992. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res* 20:5263–5269.
81. Mermoud JE, Cohen PT, Lamond AI. 1994. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J* 13:5679–5688.
82. Marinescu V, Loomis PA, Ehmann S, Beales M, Potashkin JA. 2007. Regulation of retention of FosB intron 4 by PTB. *PLoS One* 2:e828. <https://doi.org/10.1371/journal.pone.0000828>.
83. Gubbels M-J, Wieffer M, Striepen B. 2004. Fluorescent protein tagging in *Toxoplasma gondii*: identification of a novel inner membrane complex component conserved among Apicomplexa. *Mol Biochem Parasitol* 137:99–110. <https://doi.org/10.1016/j.molbiopara.2004.05.007>.
84. Roth MB, Zahler AM, Stolk JA. 1991. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol* 115:587–596.
85. Fan Q, Miao J, Cui L, Cui L. 2009. Characterization of PRMT1 from *Plasmodium falciparum*. *Biochem J* 421:107–118. <https://doi.org/10.1042/BJ20090185>.
86. Dixit A, Singh PK, Sharma GP, Malhotra P, Sharma P. 2010. PFSRP1, a novel splicing-related kinase from *Plasmodium falciparum*. *J Biol Chem* 285:38315–38323. <https://doi.org/10.1074/jbc.M110.119255>.
87. Eshar S, Allemand E, Sebag A, Glaser F, Muchardt C, Mandel-Gutfreund Y, Karni R, Dzikowski R. 2012. A novel *Plasmodium falciparum* SR protein is an alternative splicing factor required for the parasites' proliferation in human erythrocytes. *Nucleic Acids Res* 40:9903–9916. <https://doi.org/10.1093/nar/gks735>.
88. Eshar S, Altenhofen L, Rabner A, Ross P, Fastman Y, Mandel-Gutfreund Y, Karni R, Llinás M, Dzikowski R. 2015. PFSR1 controls alternative splicing and steady-state RNA levels in *Plasmodium falciparum* through preferential recognition of specific RNA motifs. *Mol Microbiol* 96:1283–1297. <https://doi.org/10.1111/mmi.13007>.
89. Tewari R, Straschil U, Bateman A, Böhme U, Cherevach I, Gong P, Pain A, Billker O. 2010. The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* 8:377–387. <https://doi.org/10.1016/j.chom.2010.09.006>.
90. Agarwal S, Kern S, Halbert J, Przyborski JM, Baumeister S, Dandekar T, Doerig C, Pradel G. 2011. Two nucleus-localized CDK-like kinases with crucial roles for malaria parasite erythrocytic replication are involved in phosphorylation of splicing factor. *J Cell Biochem* 112:1295–1310. <https://doi.org/10.1002/jcb.23034>.
91. Kern S, Agarwal S, Huber K, Gehring AP, Strödke B, Wirth CC, Brühl T, Abodo LO, Dandekar T, Doerig C, Fischer R, Tobin AB, Alam MM, Bracher F, Pradel G. 2014. Inhibition of the SR protein-phosphorylating CLK kinases of *Plasmodium falciparum* impairs blood stage replication



- and malaria transmission. PLoS One 9:e105732. <https://doi.org/10.1371/journal.pone.0105732>.
92. Ganguly AK, Verma G, Bhavesh NS. 28 November 2018. The N-terminal RNA recognition motif of PFSR1 confers semi-specificity for pyrimidines during RNA recognition. J Mol Biol <https://doi.org/10.1016/j.jmb.2018.11.020>.
  93. Painter HJ, Carrasquilla M, Llinas M. 2017. Capturing in vivo RNA transcriptional dynamics from the malaria parasite *Plasmodium falciparum*. Genome Res 27:1074–1086. <https://doi.org/10.1101/gr.217356.116>.
  94. Shock JL, Fischer KF, DeRisi JL. 2007. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. Genome Biol 8:R134. <https://doi.org/10.1186/gb-2007-8-7-r134>.
  95. Ohe K, Hagiwara M. 2015. Modulation of alternative splicing with chemical compounds in new therapeutics for human diseases. ACS Chem Biol 10:914–924. <https://doi.org/10.1021/cb500697f>.
  96. Sako Y, Ninomiya K, Okuno Y, Toyomoto M, Nishida A, Koike Y, Ohe K, Kii I, Yoshida S, Hashimoto N, Hosoya T, Matsuo M, Hagiwara M. 2017. Development of an orally available inhibitor of CLK1 for skipping a mutated dystrophin exon in Duchenne muscular dystrophy. Sci Rep 7:46126. <https://doi.org/10.1038/srep46126>.
  97. Jain P, Karthikeyan C, Moorthy NS, Waiker DK, Jain AK, Trivedi P. 2014. Human CDC2-like kinase 1 (CLK1): a novel target for Alzheimer's disease. Curr Drug Targets 15:539–550.
  98. Solyakov L, Halbert J, Alam MM, Semblat J-P, Dorin-Semblat D, Reininger L, Bottrill AR, Mistry S, Abdi A, Fennell C, Holland Z, Demarta C, Bouza Y, Sicard A, Nivez M-P, Eschenlauer S, Lama T, Thomas DC, Sharma P, Agarwal S, Kern S, Pradel G, Graciotti M, Tobin AB, Doerig C. 2011. Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*. Nat Commun 2:565. <https://doi.org/10.1038/ncomms1558>.
  99. Dorin-Semblat D, Quashie N, Halbert J, Sicard A, Doerig C, Peat E, Ranford-Cartwright L, Doerig C. 2007. Functional characterization of both MAP kinases of the human malaria parasite *Plasmodium falciparum* by reverse genetics. Mol Microbiol 65:1170–1180. <https://doi.org/10.1111/j.1365-2958.2007.05859.x>.
  100. Suvorova ES, Croken M, Kratzer S, Ting LM, Conde de Felipe M, Balu B, Markillie ML, Weiss LM, Kim K, White MW. 2013. Discovery of a splicing regulator required for cell cycle progression. PLoS Genet 9:e1003305. <https://doi.org/10.1371/journal.pgen.1003305>.
  101. Garg A, Wesolowski D, Alonso D, Deitsch KW, Ben Mamoun C, Altman S. 2015. Targeting protein translation, RNA splicing, and degradation by morpholino-based conjugates in *Plasmodium falciparum*. Proc Natl Acad Sci U S A 112:11935–11940. <https://doi.org/10.1073/pnas.1515864112>.
  102. Morcos PA. 2007. Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. Biochem Biophys Res Commun 358:521–527. <https://doi.org/10.1016/j.bbrc.2007.04.172>.
  103. Lim KR, Maruyama R, Yokota T. 2017. Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des Devel Ther 11:533–545. <https://doi.org/10.2147/DDDT.S97635>.
  104. Anders S, Reyes A, Huber W. 2012. Detecting differential usage of exons from RNA-seq data. Genome Res 22:2008–2017. <https://doi.org/10.1101/gr.133744.111>.
  105. Yeoh LM, Goodman CD, Mollard V, McFadden GI, Ralph SA. 2017. Comparative transcriptomics of female and male gametocytes in *Plasmodium berghei* and the evolution of sex in alveolates. BMC Genomics 18:734. <https://doi.org/10.1186/s12864-017-4100-0>.
  106. Pino P, Aeby E, Foth BJ, Sheiner L, Soldati T, Schneider A, Soldati-Favre D. 2010. Mitochondrial translation in absence of local tRNA aminoacylation and methionyl tRNA Met formylation in Apicomplexa. Mol Microbiol 76:706–718. <https://doi.org/10.1111/j.1365-2958.2010.07128.x>.
  107. Bates DO, Morris JC, Oltean S, Donaldson LF. 2017. Pharmacology of modulators of alternative splicing. Pharmacol Rev 69:63–79. <https://doi.org/10.1124/pr.115.011239>.



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Yeoh, LM; Lee, VV; McFadden, GI; Ralph, SA

**Title:**

Alternative Splicing in Apicomplexan Parasites

**Date:**

2019-01-01

**Citation:**

Yeoh, LM; Lee, VV; McFadden, GI; Ralph, SA, Alternative Splicing in Apicomplexan Parasites, MBIO, 2019, 10 (1)

**Persistent Link:**

<http://hdl.handle.net/11343/221990>

**File Description:**

Published version