BMC Genomics

CrossMark

Comparative ribosome profiling uncovers a dominant role for translational control in *Toxoplasma gondii*

Musa A. Hassan[1,2*], Juan J. Vasquez[3,8], Chew Guo-Liang[4], Markus Meissner[5,6] and T. Nicolai Siegel[3,6,7]

Abstract

Background: The lytic cycle of the protozoan parasite *Toxoplasma gondii,* which involves a brief sojourn in the extracellular space, is characterized by defined transcriptional profiles. For an obligate intracellular parasite that is shielded from the cytosolic host immune factors by a parasitophorous vacuole, the brief entry into the extracellular space is likely to exert enormous stress. Due to its role in cellular stress response, we hypothesize that translational control plays an important role in regulating gene expression in *Toxoplasma* during the lytic cycle. Unlike transcriptional profiles, insights into genome-wide translational profiles of *Toxoplasma gondii* are lacking.

Methods: We have performed genome-wide ribosome profiling, coupled with high throughput RNA sequencing, in intracellular and extracellular *Toxoplasma gondii* parasites to investigate translational control during the lytic cycle.

Results: Although differences in transcript abundance were mostly mirrored at the translational level, we observed significant differences in the abundance of ribosome footprints between the two parasite stages. Furthermore, our data suggest that mRNA translation in the parasite is potentially regulated by mRNA secondary structure and upstream open reading frames.

Conclusion: We show that most of the *Toxoplasma* genes that are dysregulated during the lytic cycle are translationally regulated.

Keywords: Ribosome profiling, RNA-sequencing, Translation efficiency, *Toxoplasma gondii*, Apicomplexan

Background

All living organisms are constantly exposed to a variety of biological stress, which may include limiting nutrient availability, oxidative stress, temperature shock, DNA damage and infection. Consequently, organisms show remarkable regulatory plasticity that allows them to thrive under different, sometimes harsh, environmental conditions [1, 2]. Historically, due to the relative ease of obtaining global transcript abundance estimates, most studies quantify fluctuations in mRNA abundance to gain insights into organismal response to stress [3, 4]. However, the catalogue of expressed genes and proteins is modulated at various steps, including mRNA splicing, export, stability,

translation, and protein degradation [5]. Consequently, transcript abundance rarely mirrors cellular protein levels [6]. Although the relative contribution of each of these steps in the gene-expression pathway is equivocal, mRNA translation is known to play a significant role in modulating cellular protein levels [7, 8]. Indeed, translational control of gene expression is known to provide opportunities for controlling spatial and temporal protein distribution [9]. Furthermore, because most eukaryotic mRNAs can be detected in cells at least 2 h after expression [10], compared to de novo transcription, regulating the translation of the available mRNAs provides a mechanism to rapidly adjust cellular protein levels in response to drastic changes in the environment or developmental stages [11, 12]. In fact, most translationally regulated mRNAs are known to encode proteins that regulate important cellular processes such as stress response, development, and synaptic transmission [7].

* Correspondence: musa.hassan@roslin.ed.ac.uk
[1]Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, UK
[2]The Centre for Tropical Livestock Genetics and Health, The Roslin Institute, University of Edinburgh, Edinburgh, UK
Full list of author information is available at the end of the article

Hassan *et al. BMC Genomics* (2017) 18:961

Page 2 of 12

Toxoplasma gondii is an obligate intracellular apicomplexan that infects virtually all warm-blooded vertebrates. In the definitive feline host, *Toxoplasma* undergoes sexual recombination, but reverts to asexual reproduction in the intermediate host, which includes humans. Asexual reproduction in *Toxoplasma* is characterized by the rapidly dividing tachyzoite stage. However, in response to host-derived stress factors, such as immune response, the rapidly dividing tachyzoites convert to the semi-dormant encysted bradyzoites and establish lifelong chronic infections in the central nervous system and muscle tissues of the host [13, 14]. Establishment of chronic infection is important for the re-entry of the parasite into the definitive host, and for horizontal transmission within intermediate hosts, through the predation and consumption of food products from chronically infected hosts, respectively [15]. The tachyzoite-to-bradyzoite conversion reportedly mirrors a stress response [2] and does not only involve significant changes in the parasite physiology and morphology, but also is accompanied by altered gene expression profiles [16]. During the lytic cycle the parasite invades a host cell, replicates, and then lyses out of the host cell before infecting a new host cell. This process temporarily exposes the parasite to the extracellular milieu. The extracellular viability of the parasite is reported to decrease dramatically between 6 and 12 h after egress [17], indicating the level of biological stress induced on the parasite by host factors. Indeed, transcriptional data on most *Toxoplasma* strains have revealed stage-specific expression of several genes, such as surface antigens, stress response genes, virulence genes, and metabolic enzymes [13, 18, 19]. Consequently, regulating transcript abundance, and by extension their protein products, is key in regulating *Toxoplasma* developmental stages and intercellular transmission.

Translational regulation of gene expression has emerged as a key factor in the biology of apicomplexan parasites [20–23]. In *Plasmodium,* translational regulation is reported to modulate stage conversion and host-parasite interactions [20, 21]. For example, while *Pb2* transcripts, a surface antigen, can be detected in *Plasmodium berghei* female gametocytes, the translation of *Pb2* mRNA occurs only when the parasite is in the mosquito gut [24]. In *Toxoplasma*, genetic perturbation of the eukaryotic elongation factor 2 alpha (eIF2α), an important component of the translation initiation complex, affects extracellular viability. Phosphorylated eIF2α is essential for transferring the initial methionyl tRNA (Met-tRNAi) to the 40S pre-initiation complex [25]. However, when phosphorylated at a regulatory serine (serine-51), eIF2α is unable to activate Met-tRNAi and global translation is diminished [25]. *Toxoplasma* parasites expressing eIF2 (TgIF2α) with a mutation on the

regulatory serine (serine-71) are reported to exhibit decreased extracellular viability [26]. Lower expression levels of *eIF4,* another translation initiation factor, has been observed in bradyzoites and attenuated *Toxoplasma* strains [2, 23, 24, 26]. Finally, the endoplasmic reticulum (ER) stress response in *Toxoplasma* is characterized by preferential translation of a subset of genes, including the transcriptional regulator *AP2* [23, 27]. This is particularly important since the integrity of the parasite ER is pivotal for the proper folding of essential proteins required for parasite invasion, immune evasion and the establishment of chronic infection [24]. Thus, it is plausible that the viability, pathogenesis, and transmission of *Toxoplasma* are dependent on its ability to recognize and translationally respond to host-derived stress.
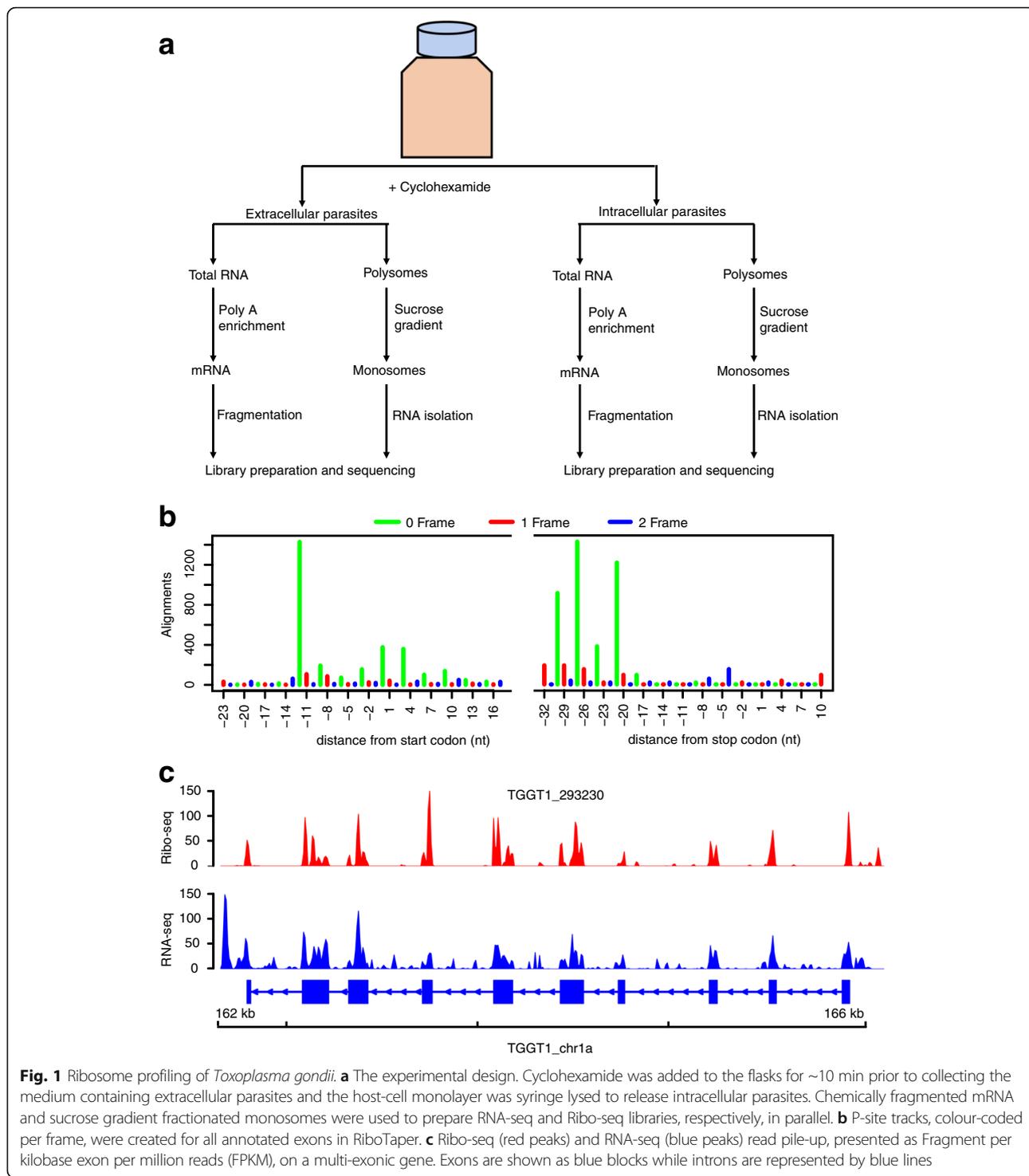
Genome-wide insights on translational control and the underlying molecular factors that regulate mRNA translation in *Toxoplasma* are largely unknown. Here, we access the translational landscape in *Toxoplasma gondii* and determine its impact on intercellular parasite transmission. To do this, we have used ribosome profiling to capture genome-wide translational profiles of intracellular and extracellular *Toxoplasma* parasites infecting human foreskin fibroblasts. Our data reveal a putative role for translational control in regulating parasite gene expression during the lytic cycle. Additionally, our data revealed variable translational efficiency of several dysregulated *Toxoplasma* mRNAs, such as the mRNAs encoding dense granules, which are known to be spatially secreted during the lytic cycle. Finally, our data suggest that that mRNA secondary structure, putatively affect mRNA translation in *Toxoplasma*. These results not only provide greater insights into *Toxoplasma* gene regulation, but also provide a resource and template for elucidating the function of translational control in *Toxoplasma* biology. Finally, the ribosome footprints, will provide an additional resource for annotating *Toxoplasma* transcript features.

Results
Generation of mRNA profiles and ribosome footprints in intracellular and extracellular *Toxoplasma*
To investigate genome-wide transcriptional and translational status in *Toxoplasma* during the lytic cycle, we performed RNA sequencing (RNA-seq) and ribosome profiling on two biological replicates of extracellular and intracellular parasites as previously described [4] (the experimental layout is depicted in Fig. 1a).

The basic concept of ribosome profiling is that actively translated mRNAs are protected from ribonucleases by the decoding ribosomes. However, other classes of RNA-binding proteins can protect mRNA from nucleases. Therefore, the presence of sequencing reads derived from nuclease-resistant RNA fragments does not

Hassan *et al. BMC Genomics* (2017) 18:961

Page 3 of 12

Fig. 1 Ribosome profiling of *Toxoplasma gondii*. **a** The experimental design. Cyclohexamide was added to the flasks for ~10 min prior to collecting the medium containing extracellular parasites and the host-cell monolayer was syringe lysed to release intracellular parasites. Chemically fragmented mRNA and sucrose gradient fractionated monosomes were used to prepare RNA-seq and Ribo-seq libraries, respectively, in parallel. **b** P-site tracks, colour-coded per frame, were created for all annotated exons in RiboTaper. **c** Ribo-seq (red peaks) and RNA-seq (blue peaks) read pile-up, presented as Fragment per kilobase exon per million reads (FPKM), on a multi-exonic gene. Exons are shown as blue blocks while introns are represented by blue lines

necessarily infer active translation. Since ribosomes decode mRNA by reading 3-nucleotides (3-nt) at a time, 3-nt periodicity on ribosome footprints is often used to distinguish ribosome protected RNA from other classes of nuclease resistant RNAs [3, 28–30]. Therefore, to increase coverage, we pooled ribosome-protected RNA

footprints from the two biological replicates for each sample, and used sub-codon resolution to call high confidence translated open reading frames (ORFs) in canonical *Toxoplasma* coding sequences. To do this, we used RiboTaper, a ribosome profiling analysis program that defines the peptidyl-site (P-site; the second tRNA entry

Hassan *et al. BMC Genomics* (2017) 18:961

Page 4 of 12

site linked to the growing polypeptide chain) of ribosome-protected RNA sequencing (Ribo-seq) reads mapping over annotated transcripts [3]. Henceforth, unless otherwise stated, all analyses on extracellular or intracellular samples are based on pooled Ribo-seq or RNA-seq data. Because 3-nt periodicity often vary between different Ribo-seq read lengths, we performed sub-codon resolution on 25–30-nt reads, which is within the range of 80S ribosome-protected RNA lengths [5]. We observed a strong 3-nt periodicity in 29-nt footprints, with up to 12-nt upstream of the AUG start site covered by ribosome footprints (12-nt offset) (Fig. 1b). Similar offsets were obtained using Riboprofiling [31], a Bioconductor package for processing Ribo-seq data (Additional file 1: Figure S1). Unlike RNA-seq reads that sometimes contain reads aligning to intronic regions, ribosome footprints mapped predominantly to annotated *Toxoplasma* protein coding regions (Fig. 1c). Therefore, the ribosome footprints in the current experiment are mostly derived from ribosome-protected nuclease-resistant mRNA fragments and can be used to accurately quantify translation in *Toxoplasma*.

Ribosome profiling confirms translation of annotated CDSes and identifies novel translated ORFs in *Toxoplasma*

RNA-seq alone cannot distinguish translated from non-translated transcripts. Additionally, it is not clear whether some annotated non-coding RNAs contain translated small open reading frames. These problems are exacerbated in non-model organisms, such as *Toxoplasma*, with incompletely annotated genomes. Because Ribo-seq captures ribosome-engaged mRNAs, it is often used to not only estimate the translation efficiencies of annotated coding regions, but also to identify novel translated ORFs. Consequently, we used RiboTaper, as previously described [3], to identify translated ORFs based on 3-nt periodicity and P-site positions in the expressed *Toxoplasma* genes. Because the current annotation of *Toxoplasma* gene structures (ToxoDB.org; GT1 v28 [32]) is incomplete, and RiboTaper classifies ORFs based on known coding regions, we initially used RNA-seq reads (~500 million paired-end reads from this and a parallel study [33]) to update GT1 gene structures. To do this, we performed genome-guided transcript assembly using Trinity [34], followed by transcript structure resolution using the Program to Assemble Spliced Alignments (PASA) [35], as previously described [36]. Subsequently, we updated the structures of 6442 transcripts, mostly due to the addition of 5′ and 3′ UTRs (mean lengths of 435-nt and 508-nt, respectively) (Fig. 2a). Next, we used RiboTaper and identified 4224 ORFs in 4195 genes based on the updated transcript structures. Noteworthy, the identification of ORFs in RiboTaper is based on codon resolution on the Ribo-seq

reads that map to annotated transcript features rather than the simple presence of ribosome footprints. Thus, the number of translated ORFs identified by this approach may be lower than the actual number of genes with ribosome footprints. Besides canonical ORFs, we identified 172 novel ORFs, mainly due to the alternative splicing of annotated transcripts (Fig. 2b), PASA-updated new transcripts structures (Fig. 2c), or novel transcripts (Fig. 2d). Therefore, by using ribosome footprints, we not only provide a greater resolution of the canonical ORFs to include alternative isoforms, but also identify novel ORFs in *Toxoplasma* (GT1).

Steady-state mRNA and translation efficiency in intracellular and extracellular parasites

We sought to evaluate global translational divergence in intracellular versus extracellular type I *Toxoplasma* parasites. Initially, we used HTSeq [37] to obtain raw read counts from the uniquely mapped RNA-seq and Ribo-seq reads, which were then normalized (Normalized Read Counts, NRC) in DESeq2 [38] to adjust for variation in sequencing depths across samples. Even though some genes were lowly expressed, approximately 7065 genes (83% of the ~8460 genes annotated in the GT1 genome (v28) were expressed (average RNA-seq NRC > 5 across samples) (Additional file 2 A). Of the expressed genes, 6508 had ribosome footprints (average Ribo-seq NRC > 5 across samples), suggesting that 557 transcripts are non-coding or poorly translated in our parasite populations (Additional file 2 B). Interestingly, 274/557 (> 50% of the potentially non-coding or poorly translated genes) had an average RNA-seq NRC > 10 (mean NRC = 41.18; SEM = ±3.48), suggesting that these genes are expressed above background levels (which we arbitrarily set at NRC < 5) but are either translationally repressed in these parasite populations or non-coding. The protein products for most of these 274 genes are annotated in ToxoDB as "hypothetical", but also included the KRUF proteins, which are encoded from a highly expanded gene family in the GT1 strain [39]. Also included in the 274 poorly translated or non-coding genes was the *Toxoplasma* translation initiation factor 2 (TgIF2K-C), which is required for the parasites' response to intracellular glutamine starvation in human cells [40]. These 274 transcripts were functionally enriched for, among others, "cell adhesion" (Bonferroni *P* value = *3.58e-4*) and "microtubule motor activity" (Bonferroni *P* value = *9.96e-4*). Although they are included in the current GT1 genome annotation, 27 of the 274 genes did not have any corresponding proteomic data in ToxoDB [32], suggesting that they are non-coding. On the other hand, 83 transcripts exhibited low abundance with an average RNA-seq NRC < 5 (mean NRC = 3.14; SEM = ±0.14), but had

Hassan *et al. BMC Genomics* (2017) 18:961

Page 5 of 12
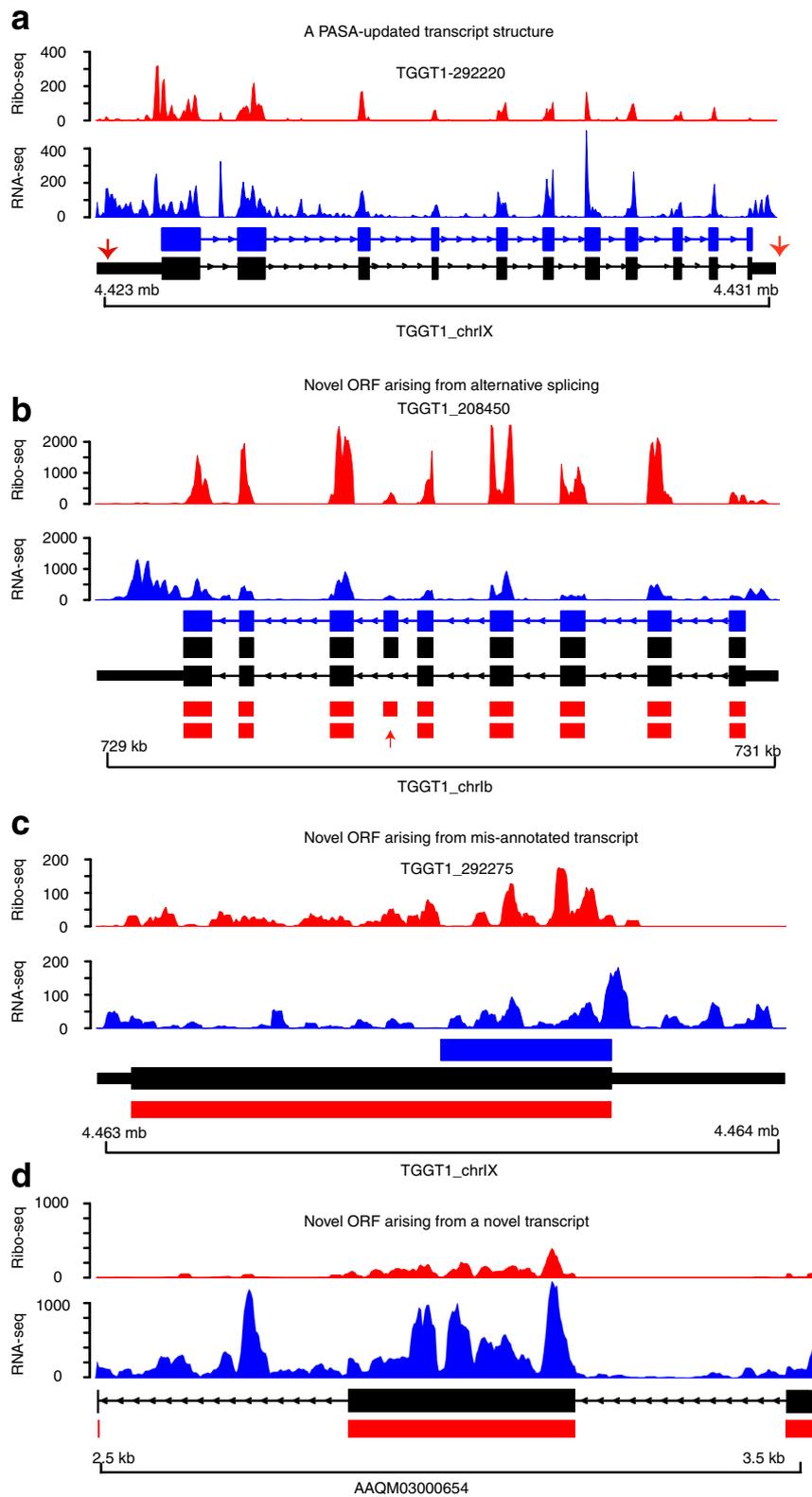
Fig. 2 (See legend on next page.)

Hassan *et al. BMC Genomics* (2017) 18:961

Page 6 of 12

(See figure on previous page.)

Fig. 2 Ribosome profiling reveals novel and annotated open reading frames. **a** Shown are examples of PASA-updated *Toxoplasma* transcript structure annotations (Black) and the corresponding current ToxoDB transcript annotation (Blue). A) The PASA-updated transcript is due to the addition of 5′/3′ UTRs (red arrows). Prediction of ORFs based on the PASA-updated transcript structures in RiboTaper identified canonical and novel ORFs. The novel ORFs were mainly due to; **b** alternative splicing of annotated transcripts (skipped exon, red arrow), **c** potentially mis-annotated transcripts structures, and **d** novel transcripts. The RiboTaper-predicted ORFs in B-D are presented as red blocks. In all the figures, Ribo-seq and RNA-seq read coverage on each transcript is shown in Fragment per kilobase exon per million reads (FPKM)

average Ribo-seq NRC > 5 (mean NRC = 13.17; SEM = ±3.52) (Additional file 2 C). These 87 genes included the SAG-related sequence (SRS) gene family that are implicated in *Toxoplasma* virulence in mice [41].

Next, we compared differences in mRNA abundance and ribosome occupancy between the intracellular and extracellular parasites. Using a Benjamini-Hochberg False Discovery Rate (FDR) ≤ 10%, we identified three classes of differentially regulated genes: 1) 891 genes that varied both at the level of transcript abundance and ribosome occupancy i.e. concordant (RNA + RIBO) (Additional file 2 D), 2) 645 genes that varied only at the levels of mRNA abundance (RNA-ONLY) (Additional file 2 E), and 3) 1324 genes that varied only at the level of ribosome occupancy (RIBO-ONLY) (Fig. 3a and Additional file 2 F), indicating that many of the genes that are dysregulated in *Toxoplasma* during the lytic cycle are regulated at the translational level. To determine the overall contribution of translation in regulating gene expression during *Toxoplasma's* lytic cycle, we used a standardized major-axis estimation

(SME) [42] analysis to calculate the slopes of fold changes in RNA-seq or Ribo-seq NRCs between intracellular and extracellular parasites. Unlike the RIBO + RNA transcripts, where the slope approached 1 (slope = 1.15), indicating the co-occurrence of changes in transcript abundance and ribosome occupancy, the slope for RIBO-ONLY transcripts (slope = 2.87) was significantly (*P* value < 2.22e-16) greater than 1 (Fig. 3b), confirming that many differences in gene expression between the intracellular and extracellular parasites occur at the translation level.

Next, we calculated differences in translation efficiency (TE) for each expressed transcript between extracellular and intracellular parasites using Ribodiff [43]. At a Benjamini-Hochberg FDR ≤ 10%, we identified differential TE in 834 genes in intracellular versus extracellular parasites (Additional file 2 G). Because of the potential variation in mRNA between intracellular and extracellular parasites, which in the absence of a spike-in control during RNA sequencing may skew the data, we complemented the ribodiff protocol by ranking the genes based
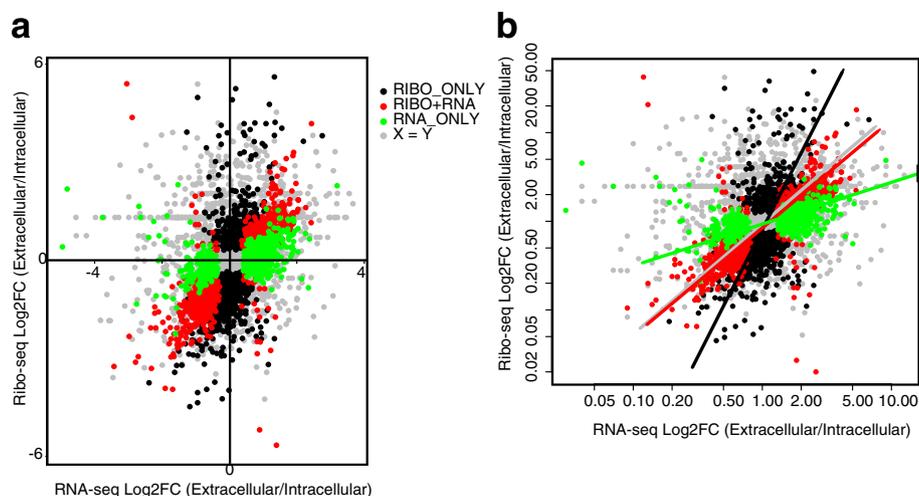
Fig. 3 Translation regulation is predominant in *Toxoplasma gondii*. **a** Although most transcripts showed no differential regulation (grey dots *n = 4332*) between the intracellular and extracellular parasites, several dysregulated genes were significantly (Benjamini–Hochberg *FDR ≤ 0.1*) regulated at the level of mRNA abundance only (green dots, RNA-ONLY, *n = 650*), ribosome occupancy only (black dots, RIBO-ONLY; *n = 1339*), and at the level of both mRNA abundance and ribosome occupancy (red dots, RIBO + RNA, *n = 899*). Differences in translational efficiency in intracellular versus extracellular parasites were calculated in Ribodiff on transcripts with average RNA-seq DESeq2 normalized read counts ≥ 5. **b** The slope for red spots approached 1 (x = y), indicating equal fold changes in gene regulation at the level of mRNA abundance and ribosome occupancy between the intracellular and extracellular parasites (standardized major-axis estimation; SMA). Significant (*P < 2.22e-16*) divergence of black and red slopes demonstrate differences in gene regulation at the level of ribosome occupancy (RIBO-ONLY) and mRNA abundance (RNA-ONLY), respectively

Hassan *et al. BMC Genomics* (2017) 18:961

Page 7 of 12

on the z-scores of TE in each parasite population. We considered genes with RNA-seq NRC ≥ 5 (7065 genes) and at least two standard deviations above or below the mean TE in each population as translationally up- or down- regulated, respectively. By this metric, 868 genes were translationally down-regulated while 119 genes were up-regulated in intracellular parasites. On the other hand, 1004 and 236 genes were down- and up-regulated, respectively, in extracellular parasites. Of the dysregulated genes, 344 and 556 genes were exclusively dysregulated in intracellular and extracellular parasites, respectively (not deviating from the mean or dysregulated in the opposite directions in the two populations, e.g. up-regulated in intracellular but down-regulated in extracellular parasites). The "sporozoite development protein (TGGT1_257010)" and "BT1 family protein (TGGT1_236020)" genes were the most down- (z-score = −5.0; Log_2TE = −6.03) and up-regulated (z-score = 4.79; Log_2TE = 4.15), respectively, in intracellular parasites. In extracellular parasites, "the transporter, major facilitator family protein (TGGT1_266870)" and "CMGC kinase, CDK family (TGGT1_253580)" genes were the most down- (z-score = −5.85; Log_2TE = 6.43) and up-regulated (z-score = 6.13; Log_2TE = 5.55), respectively. Although dense granules are secreted by intracellular parasites [44], the translation efficiency for genes encoding these proteins (GRA1, GRA4, and GRA7) was up-regulated in extracellular parasites. Additionally, the translation of genes encoding the alveolin domain-containing inner membrane complex (IMC) proteins (IMC1, IMC4, IMC6, and IMC10), which are required during intracellular *Toxoplasma* cell division [45], were up-regulated in extracellular parasites.

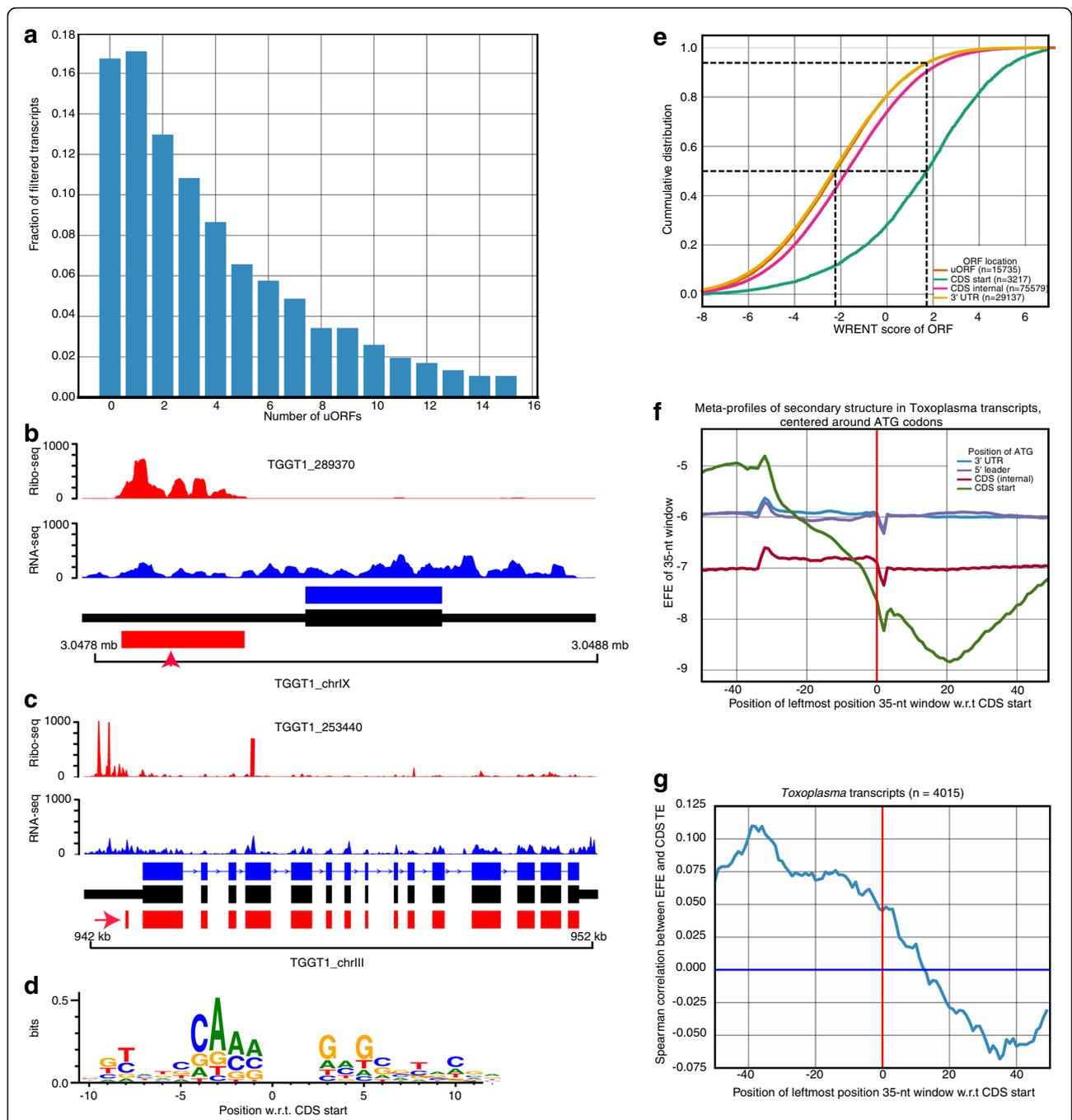### Most *Toxoplasma* transcripts contain open reading frames (ORFs) at their 5′ untranslated regions

Besides translation at canonical protein coding sequences (CDSes), ribosome profiling can reveal novel coding sequences, including coding sequences at the 5′ and 3′ untranslated regions (upstream and downstream ORFs, uORFs and dORFs, respectively) [3, 28]. Because the prevalence and translation regulatory potential of uORFs is largely unknown in *Toxoplasma*, we used a support vector classifier [29] to identify translated uORFs. Based on the presence of a start and an in-frame downstream stop codon, we observed a high prevalence of uORFs in *Toxoplasma*, with some transcripts having > 4 non-overlapping uORFs (Fig. 4a). From 4577 transcripts with annotated 5′ UTRs of lengths ≥ 20-nt, we identified uORFs in 3348 (73%). Similar abundance of uORFs has also been reported in *Plasmodium falciparum* [21]. We filtered the transcripts further to 2770 (translated uORFs) based on the presence of ribosome footprints, 3-nt periodicity on Ribo-seq reads, and a minimum level of expression of the cognate transcript (Fragment per kilobase exon per million reads; FPKM ≥ 1).

In other eukaryotes, uORFs are not only prevalent, but also regulate translation of cognate downstream CDSes [20, 28]. Consistent with the reported uORF-mediated repression of translation at canonical CDSes [46–48], we observed individual examples of highly-translated uORFs upstream of their cognate lowly-translated CDSes (Fig. 4b-c). Because, sequence and mRNA secondary structure can modulate translation [49–51], we performed linear regression with these features against translation efficiency in *Toxoplasma*, as previously described [28]. Briefly, we used annotated *Toxoplasma* CDSes lacking uORFs as a training set to define the sequence motif that promotes translation initiation (initiation context), by weighting the contribution of position-specific scoring matrix (PSSM) to translation efficiency of individual transcripts (Fig. 4d). Next, we used the PSSM to score initiation sequences in individual transcripts that contain uORFs (weighted relative entropy, WRENT) (See Methods). Relative to canonical CDSes, WRENT scores at uORFs were generally unfavourable to translation initiation (Fig. 4e). We then calculated the secondary structure ensemble free energy [52], using the ViennaRNA package [53], in a 35-nt sliding window across entire transcripts to evaluate the effect of mRNA secondary structure on translation. Unlike humans and mice [28], *Toxoplasma* transcripts exhibited an unstable secondary structure before the CDS start codon and a more stable secondary structure after the CDS start codon (Fig. 4f). Moreover, the stability of the secondary structure at these regions correlates with translation efficiency of the transcripts (Fig. 4g, and Additional file 3: Figure S2 and Additional file 4: Table S1). Thus, most uORFs in *Toxoplasma* are not efficiently translated and mRNA secondary structure putatively regulate translation efficiency in *Toxoplasma*.

Discussion

During the lytic cycle, *Toxoplasma* frequently transitions between an intracellular and extracellular niche, that is characterized by a variety of molecular changes in the parasite, including distinct transcriptional profiles [54]. Although components of the translation initiation complex, such as eIF2α, reportedly modulate stress response, extracellular survival and, virulence in *Toxoplasma* [23, 26], global translational changes during *Toxoplasma* lytic cycle are largely unknown. Here, we used ribosome profiling to reveal that translational regulation of gene expression is prominent during *Toxoplasmas*' lytic cycle. Additionally, our data suggest mRNA secondary structure potentially regulate translation in *Toxoplasma*. Even though most of the genes expressed during the lytic cycle are known to exhibit a

Hassan *et al. BMC Genomics* (2017) 18:961

Page 8 of 12

Fig. 4 Open reading frames in 5′ leader sequences (uORFs) are abundant in *Toxoplasma*. **a** Based on the presence of a start and a downstream in-frame stop codon, some *Toxoplasma* transcripts contain more than 14 non-overlapping uORFs. **b-c** Individual *Toxoplasma* transcripts with an upstream translated uORFs (red arrow) exhibit reduced ribosome occupancy at the cognate canonical protein coding sequences. Ribo-seq (red peaks) and RNA-seq (blue peaks) read pile-up on ToxDB (blue) and PASA-updated (black) transcript structures. **d** Weighted position-specific scoring matrices (PSSMs) around (±10 nucleotides) annotated ATG of transcripts lacking uORFs. Vertical axis represents weighted relative entropy (WRENT). **e** Cumulative WRENT scores around ATG start codons along the coding transcripts. Dotted lines indicate median WRENT scores of uORF and CDS, and the proportion of uORFs (~90%) with WRENT scores less than the median CDS WRENT score. **f** Predicted secondary structure ensemble free energies (EFEs) in a sliding 35-nt window on 5′ UTRs, CDSes, and 3′ UTRs. A more negative EFE indicates more stable secondary structures. **g** The secondary structure EFE correlates with translation efficiency in CDSes. See also Additional file 3: Figure S2 and Additional file 4: Table S1

Hassan *et al. BMC Genomics* (2017) 18:961

Page 9 of 12

cyclic expression pattern coinciding with the different cell cycle stages [55], we show that the expression and translation of most of these genes are not temporally or spatially synchronized during the lytic cycle. However, since parasite replication and egress is not synchronized among individual parasites, it is impossible to decipher the translational changes that occur as the parasite adapts to the extracellular microenvironment. Thus, it is not clear whether the differences in translation efficiencies between intracellular and extracellular parasites observed in this study are maintained throughout the lytic cycle. With single-cell or time course analysis of the *Toxoplasma* "translatome", we may be able to show fluctuations in translation as the parasites egress or re-infect host cells.

Interestingly, *Toxoplasma* transcripts exhibited less stable RNA secondary structure before the ATG start site. Similar reduction in RNA secondary structure have been reported in zebrafish [28]. In contrast to CDSes, this switch from unstable to stable secondary structure around the initiation codon was not observed in uORFs. This distinction in the initiation context of uORFs and CDSes, in terms of both sequence and secondary structure suggests that these two features are important for start site selection in *Toxoplasma*. uORFs have been shown to be prevalent and regulatory in a variety of organisms, including Apicomplexans [21, 56]. However, their prevalence and regulatory potential in *Toxoplasma* is largely unknown. We show that, while uORFs are prevalent in *Toxoplasma*, their translation is not favoured, probably due to selection at their initiation contexts (sequence and secondary structure). Nevertheless, we observed individual cases where high translation at an uORF correlates with weak translation at a cognate downstream CDS, which raises interesting questions that are worthy of further investigations. For example, is the translation of uORF unfavourable at all the developmental stages? How is the translation of uORFs regulated in *Toxoplasma*? Additionally, the mechanisms that regulate translation efficiency in *Toxoplasma*, which are equivocal, are worthy of further investigation. High ribosome occupancy may not be related to high rates of translation but rather ribosome pausing [57], which can be caused by long stretches of rare codons, high mRNA secondary structure, or interactions of the growing polypeptide chain with the ribosome [58, 59]. Overall, it is worthy investigating the role of translational control in modulating *Toxoplasma* strain differences in virulence, adaption to variable host genetic background or host cell activation status.

Conclusion

The results presented in this work reveals key aspects of translational control in *Toxoplasma gondii* during the lytic cycle. We show that many dysregulated genes are translationally regulated during intercellular parasite transmission and that uORFs are prevalent, although not translationally favoured in *Toxoplasma gondii*. We anticipate that this work will be the basis for future research on translational regulation in the different development stages of the parasite and host cell microenvironments.

Methods

Parasite culture, ribosome isolation and, sequencing libraries

Toxoplasma gondii was maintained in the laboratory by serial passage on human foreskin fibroblasts (HFFs), according to standard procedures [36]. For ribosome profiling, HFF monolayers in T175 flasks were infected with a high inoculum of a type I (RH) *Toxoplasma* strain. After 2 h of infection, the cell culture medium was removed, the monolayer rinsed with ice cold Phosphate saline buffer (PBS) to remove any extracellular parasites, fresh cell culture medium added, and the parasites let to replicate and lyse for ~18 h. 10 mins before harvest, cyclohexamide (100 μg/ml) was added to the cell culture. Cell culture supernatant, containing lysed out extracellular parasites, was harvested and passed through 5 μm filters to remove HFFs. The remaining HFF monolayer, containing intracellular parasites, was rinsed with PBS to remove any extracellular parasites, scrapped, syringe lysed using 27G needles, and passed through 5 μm filters. The parasites were pelleted by centrifugation at 1700 × g, 4 °C for 7 min. The parasite pellets (intracellular and extracellular) were washed with polysome buffer and processed for ribosome profiling, as previously described [4].

Pre-processing of Ribo-seq and RNA-seq data

Ribo-seq and RNA-seq reads were stripped from adapter sequences and aligned to the *GT1 Toxoplasma* genome (v28) using the split-aware aligner STAR [60], allowing up to 4 mismatches and discarding reads shorter than 20 nt. P-site locations and read length off-sets were inferred from the Ribo-seq data as previously described [3]. Normalized read counts (NRC) values for Ribo-seq and RNA-seq data were calculated in DESeq2 [61] based on counts data generated using HTseq [37]. P-site positions, RNA-seq coverage, RNA-site positions for different de novo assembled gene structures were created in RiboTaper [3]. All the raw and processed data can be obtained from NCBI using the are GEOarchive accession number GSE99395.

Exon level annotation and ORF identification

First, GT1 transcripts were reconstructed *de novo* in Trinity [34] and PASA [62] guided by the GT1 genome (ToxoDB v28) [32]. Next, we used RiboTaper to identify

Hassan *et al. BMC Genomics* (2017) 18:961

Page 10 of 12

ORFs as previously described [3]. Briefly, we used the annotated canonical coding sequences (CCDS) in ToxoDB to distinguish de novo assembled exons that; 1) overlap annotated exons in ToxoDB (CCDS), 2) do not overlap any exons inside CCDS-containing genes (non-CCDS) and, 3) overlap non-CCDS containing genes or do not overlap any annotated gene (non-CCDS). The non-CCDS included novel 5′/3′ UTRs, alternatively spliced exons, and novel exons. Next ORFs were defined based on the presence of an AUG start codon and an in-frame stop codon, after training the pipeline with 1000 CCDS from ToxoDB and random shuffling of the P-sites. Next, every transcript with a pair of consecutive start-stop codons (ORFs) was tested for 3-nt periodicity ($P \leq 0.05$) and all ORFs with less than 50% of in-frame P-sites discarded (Refer to [3] for a detailed description of the RiboTaper pipeline). Translation initiation context, RNA-secondary structure, upstream open reading frames (uORFs) repressiveness and, uORF positional frequencies and biases were identified and modelled as previously described [28].

Additional files

Additional file 1: Figure S1. 3-nucleotide periodicity of ribo-seq reads as predicted in the Bioconductor package, Riboprofiling for all reads and 26–30-nt reads. The numbers in red show read offsets from the AUG start site. (AI 145 kb)

Additional file 2: Dataset. A) Average DESeq2 normalized RNA-seq and Ribo-seq read counts for all the transcripts tested for differential expression and ribosome occupancy. B) Genes exhibiting low ribosome occupancy but above background mRNA abundance. C) Genes exhibiting low mRNA abundance but above background ribosome occupancy. D) Genes that are differentially regulated both at the level of mRNA abundance and ribosome occupancy (RNA + RIBO). E) Genes that are differentially regulated only at the level of mRNA abundance (RNA-Only). F) Genes that are differentially regulated only at the level of ribosome occupancy (RIBO-Only). G) Genes that exhibit significant differences in translation efficiency between in intercellular and extracellular parasites. (XLS 4356 kb)

Additional file 3: Figure S2. A scatter-plot of the correlations between secondary EFE and CDS TE in all filtered transcripts. (AI 7943 kb)

Additional file 4: Table S1. Individual correlations of sequence features with CDS TE for various transcript sets. (DOC 36 kb)

Abbreviations
CCDS: Canonical coding DNA sequences; dORF: Downstream Open Reading Frame; FDR: False Discovery Rate; HFF: Human foreskin fibroblasts; NRC: Normalized Read Counts; ORF: Open reading frame; PASA: Program to Assemble Spliced Alignments; PBS: Phosphate Buffered Saline; PSSM: Position-specific scoring matrix; SME: Standardized major-axis estimation; TE: Translation efficiency; uORF: Upstream open reading frame; UTR: Untranslated region; WRENT: Weighted relative entropy

Availability of data and materials
All the raw and processed data referenced in this manuscript are available in the GEOarchive under accession number GSE99395 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99395).

Authors' contributions
MAH, TNS and MM conceived and designed the experiment. MAH and JJV performed the experiments and revised the manuscript. MAH and CGL analysed, interpreted data and revised the manuscript. MAH drafted the original manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, UK. [2]The Centre for Tropical Livestock Genetics and Health, The Roslin Institute, University of Edinburgh, Edinburgh, UK. [3]Research Centre for Infectious Diseases, University of Wuerzburg, Wuerzburg 97080, Germany. [4]Computational Biology Program, Basic Sciences and Public Health Sciences Division, Fred Hutchinson Cancer Research Centre, Seattle, WA 98105, USA. [5]Wellcome Centre for Molecular Parasitology, University of Glasgow, Glasgow, UK. [6]Department of Veterinary Sciences, Experimental Parasitology, Ludwig-Maximilians-Universität, München, 80802 Munich, Germany. [7]Biomedical Center Munich, Physiological Chemistry, Ludwig-Maximilians-Universität München, Planegg-Martinsried 82152, Germany. [8]Present address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

References
1. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc. 2012;7(8):1534–50.
2. Narasimhan J, Joyce BR, Naguleswaran A, Smith AT, Livingston MR, Dixon SE, Coppens I, Wek RC, Sullivan WJ Jr. Translation regulation by eukaryotic initiation factor-2 kinases in the development of latent cysts in *Toxoplasma gondii*. J Biol Chem. 2008;283(24):16591–601.
3. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U. Detecting actively translated open reading frames in ribosome profiling data. Nat Methods. 2016;13(2):165–70.
4. Vasquez JJ, Hon CC, Vanselow JT, Schlosser A, Siegel TN. Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res. 2014;42(6):3623–37.
5. Steitz JA. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature. 1969;224(5223):957–64.
6. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13(4):227–32.
7. Loya CM, Van Vactor D, Fulga TA. Understanding neuronal connectivity through the post-transcriptional toolkit. Genes Dev. 2010;24(7):625–35.
8. Kong J, Lasko P. Translational control in cellular and developmental processes. Nat Rev Genet. 2012;13(6):383–94.

Hassan *et al. BMC Genomics* (2017) 18:961

Page 11 of 12

9. Besse F, Ephrussi A. Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat Rev Mol Cell Biol. 2008;9(12):971–80.

10. Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MS. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. 2009;16(1):45–58.

11. Anderson P, Kedersha N. RNA granules. J Cell Biol. 2006;172(6):803–8.

12. Brant-Zawadzki PB, Schmid DI, Jiang H, Weyrich AS, Zimmerman GA, Kraiss LW. Translational control in endothelial cells. J Vasc Surg. 2007;45(Suppl A):A8–14.

13. Cleary MD, Singh U, Blader IJ, Brewer JL, Boothroyd JC. *Toxoplasma gondii* asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. Eukaryot Cell. 2002;1(3):329–40.

14. Kim S-K, Karasov A, Boothroyd JC. Bradyzoite-specific surface antigen SRS9 plays a role in maintaining *Toxoplasma gondii* persistence in the brain and in host control of parasite replication in the intestine. Infect Immun. 2007;75(4):1626–34.

15. Tenter AM, Heckeroth AR, Weiss LM. *Toxoplasma gondii*: from animals to humans. Int J Parasitol. 2000;30(12–13):1217–58.

16. White MW, Radke JR, Radke JB. Toxoplasma development - turn the switch on or off? Cell Microbiol. 2014;16(4):466–72.

17. Khan A, Behnke MS, Dunay IR, White MW, Sibley LD. Phenotypic and gene expression changes among clonal type I strains of *Toxoplasma gondii*. Eukaryot Cell. 2009;8(12):1828–36.

18. Walker R, Gissot M, Huot L, Alayi TD, Hot D, Marot G, Schaeffer-Reiss C, Van Dorsselaer A, Kim K, Tomavo S. Toxoplasma transcription factor TgAP2XI-5 regulates the expression of genes involved in parasite virulence and host invasion. J Biol Chem. 2013;288(43):31127–38.

19. Blader IJ, Manger ID, Boothroyd JC. Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*-infected human cells. J Biol Chem. 2001;276(26):24223–31.

20. Bunnik EM, Chung DW, Hamilton M, Ponts N, Saraf A, Prudhomme J, Florens L, Le Roch KG. Polysome profiling reveals translational control of gene expression in the human malaria parasite Plasmodium falciparum. Genome Biol. 2013;14(11):R128.

21. Caro F, Ahyong V, Betegon M, DeRisi JL. Genome-wide regulatory dynamics of translation in the plasmodium falciparum asexual blood stages. elife. 2014;3

22. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010;11(2):113–27.

23. Joyce BR, Tampaki Z, Kim K, Wek RC, Sullivan WJ Jr. The unfolded protein response in the protozoan parasite *Toxoplasma gondii* features translational and transcriptional control. Eukaryot Cell. 2013;12(7):979–89.

24. Zhang M, Joyce BR, Sullivan WJ Jr, Nussenzweig V. Translational control in Plasmodium and toxoplasma parasites. Eukaryot Cell. 2013;12(2):161–7.

25. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell. 2009;136(4):731–45.

26. Joyce BR, Queener SF, Wek RC, Sullivan WJ Jr. Phosphorylation of eukaryotic initiation factor-2{alpha} promotes the extracellular survival of obligate intracellular parasite *Toxoplasma gondii*. Proc Natl Acad Sci U S A. 2010;107(40):17200–5.

27. Radke JB, Lucas O, De Silva EK, Ma Y, Sullivan WJ Jr, Weiss LM, Llinas M, White MW. ApiAP2 transcription factor restricts development of the Toxoplasma tissue cyst. Proc Natl Acad Sci U S A. 2013;110(17):6871–6.

28. Chew GL, Pauli A, Schier AF. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. Nat Commun. 2016;7:11663.

29. Ji Z, Song R, Regev A, Struhl K. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. elife. 2015;4:e08890.

30. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, et al. A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol Cell. 2015;60(5):816–27.

31. Popa A, Lebrigand K, Paquet A, Nottet N, Robbe-Sermesant K, Waldmann R, Barbry P. RiboProfiling: a bioconductor package for standard Ribo-seq pipeline processing. F1000Res. 2016;5:1309.

32. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, et al. ToxoDB: an integrated *Toxoplasma gondii* database resource. Nucleic Acids Res. 2008;36(suppl 1):D553–6.

33. Gras S, Jackson A, Woods S, Pall G, Whitelaw J, Leung JM, Ward GE, Roberts CW, Meissner M. Parasites lacking the micronemal protein MIC2 are deficient in surface attachment and host cell egress, but remain virulent in vivo. Wellcome Open Res. 2017;2:32.

34. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al. *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010;28(5):503–10.

35. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66.

36. Hassan MA, Melo MB, Haas B, Jensen KD, Saeij JP. 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. 2012;13:696.

37. Anders S, Pyl PT, Huber W. HTSeq–a python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

38. Anders S, Huber W. Differential expression analysis for sequencing count data. Genome Biol. 2010;11:R106.

39. Reid AJ, Vermont SJ, Cotton JA, Harris D, Hill-Cawthorne GA, Konen-Waisman S, Latham SM, Mourier T, Norton R, Quail MA, et al. Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and Neospora caninum: Coccidia differing in host range and transmission strategy. PLoS Pathog. 2012;8(3):e1002567.

40. Konrad C, Wek RC, Sullivan WJ Jr. A GCN2-like eukaryotic initiation factor 2 kinase increases the viability of extracellular *Toxoplasma gondii* parasites. Eukaryot Cell. 2011;10(11):1403–12.

41. Wasmuth JD, Pszenny V, Haile S, Jansen EM, Gast AT, Sher A, Boyle JP, Boulanger MJ, Parkinson J, Grigg ME. Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence. MBio. 2012;3(6)

42. Wanton D, Duursma R, Falster D, Taskinen S. smatr 3 - an R package for estimation and inference about allometric lines. Methods Ecol Evol. 2012;3:257–9.

43. Zhong Y, Karaletsos T, Drewe P, Sreedharan VT, Kuo D, Singh K, Wendel HG, Ratsch G. RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. Bioinformatics. 2017;33(1):139–41.

44. Carruthers VB, Sibley LD. Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. Eur J Cell Biol. 1997;73(2):114–23.

45. Dubey R, Harrison B, Dangoudoubiyam S, Bandini G, Cheng K, Kosber A, Agop-Nersesian C, Howe DK, Samuelson J, Ferguson DJP, et al. Differential roles for inner membrane complex proteins across *Toxoplasma gondii* and Sarcocystis neurona development. mSphere. 2017;2(5)

46. Barbosa C, Peixeiro I, Romao L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 2013;9(8):e1003529.

47. Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdiscip Rev RNA. 2014;5(6):765–78.

48. Somers J, Poyry T, Willis AE. A perspective on mammalian upstream open reading frame function. Int J Biochem Cell Biol. 2013;45(8):1690–700.

49. Hinnebusch AG. Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol Mol Biol Rev. 2011;75(3):434–67. first page of table of contents

50. Kozak M. Pushing the limits of the scanning mechanism for initiation of translation. Gene. 2002;299(1–2):1–34.

51. Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A. uORFdb–a comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res. 2014;42(Database issue):D60–7.

52. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851–61.

53. Lorenz R, Bernhart SH, Honer Z, Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.

54. Szatanek T, Anderson-White BR, Faugno-Fusci DM, White M, Saeij JP, Gubbels MJ. Cactin is essential for G1 progression in *Toxoplasma gondii*. Mol Microbiol. 2012;84(3):566–77.

55. Behnke MS, Wootton JC, Lehmann MM, Radke JB, Lucas O, Nawas J, Sibley LD, White MW. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of *Toxoplasma gondii*. PLoS One. 2010;5(8):e12354.

56. Amulic B, Salanti A, Lavstsen T, Nielsen MA, Deitsch KW. An upstream open reading frame controls translation of var2csa, a gene implicated in placental malaria. PLoS Pathog. 2009;5(1):e1000256.

57. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011;147(4):789–802.

58. Dimitrova LN, Kuroha K, Tatematsu T, Inada T. Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. J Biol Chem. 2009;284(16):10343–52.

Hassan *et al. BMC Genomics* (2017) 18:961

Page 12 of 12

59. Doma MK, Parker R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature. 2006;440(7083):561–4.
60. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
61. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
62. Haas B, Zeng Q, Pearson DM, Cuomo AC, Wortman JR. Approaches to fungal genome annotation. Mycology. 2011;2(3):118–41.