

25 **Abstract**

26 Filarial nematodes can cause debilitating diseases in humans. They have complicated
27 life cycles involving an insect vector and mammalian hosts, and they go through a
28 number of developmental molts. While whole genome sequences of parasitic worms are
29 now available, very little is known about transcription factor (TF) binding sites and their
30 cognate transcription factors that play a role in regulating development. To address this
31 gap, we developed a novel motif prediction pipeline, Emotif Alpha, that integrates ten
32 different motif discovery algorithms, multiple statistical tests, and a comparative analysis
33 of conserved elements between the filarial worms *Brugia malayi* and *Onchocerca*
34 *volvulus*, and the free-living nematode *Caenorhabditis elegans*. We identified stage-
35 specific TF binding motifs in *B. malayi*, with a particular focus on those potentially
36 involved in the L3-L4 molt, a stage important for the establishment of infection in the
37 mammalian host. Using an *in vitro* molting system, we tested and validated three of
38 these motifs demonstrating the accuracy of the motif prediction pipeline.

39

40

41 **Introduction**

42 *Brugia malayi* is a mosquito-borne filarial nematode and one of the causative agents of
43 lymphatic filariasis, commonly known as elephantiasis. Currently, 856 million people in
44 52 countries require preventative chemotherapy to stop the spread of the disease
45 (Gordon et al. 2018). Transmission occurs when the mosquito vector introduces
46 infective third-stage larvae (L3) during their blood meal. The larvae then migrate to the
47 lymphatic vessels where they molt twice and develop into adults. Over their lifespan
48 adult females produce millions of microfilariae (immature larvae) that circulate in the
49 blood, allowing for continued transmission. Chronic lymphatic filariasis can cause
50 permanent and disfiguring damage, characterized by lymphoedema (tissue swelling)
51 and elephantiasis (tissue thickening) of the lower limbs, and hydrocele (scrotal swelling).

52

53 Over the past decade, a few parasitic nematode genomes have been sequenced,
54 including *B. malayi* (Ghedini et al., 2007), *Loa loa* (Desjardins et al. 2013), and
55 *Onchocerca volvulus* (Cotton et al. 2016). Transcriptomic experiments have helped
56 quantify differentially expressed genes and their biological implications (Bennuru et al.,
57 2016; Choi et al., 2011; Grote et al, 2017; Kariuki, Hearne, & Beerntsen, 2010; Li, Wang,
58 Rush, Mitreva, & Weil, 2012). However, little is known about how these genes are
59 regulated through cis-regulatory motifs. Motifs that have been characterized in *B. malayi*
60 and that are available in the CIS-BP database (Weirauch et al. 2014) are purely
61 bioinformatic predictions based on transcription factor binding site (TFBS) homology. *De*
62 *novo* DNA motif discovery is an effective bioinformatic method for studying
63 transcriptional gene regulation (Dieterich & Sommer, 2008), and a number of motif

64 discovery methods and tools currently exist. These include expectation-maximization
65 methods, such as MEME (Bailey et al. 2006) and Improbizer (Ao et al. 2004); Gibbs
66 sampling methods, such as BioProspector (Liu et al. 2001) and MotifSampler (Thijs et al.
67 2002); k-mer enumeration methods such as Weeder (Pavesi, et al. 2004), DME (Smith
68 et al. 2005), and DECOD (Huggins et al. 2011); ensemble methods such as W-
69 ChIPMotifs (Jin et al. 2009), and GimmeMotifs (Heeringen & Veenstra, 2011); and deep
70 learning methods such as DanQ (Quang & Xie, 2016) and DeepFinder (Lee et al. 2018).
71 Based on the input types, motif discovery approaches can also be classified as either
72 generative or discriminative. Generative motif discovery models use pre-defined
73 background models (e.g., the Hidden Markov Model), while discriminative motif
74 discovery models need to explicitly specify a set of background sequences. In this study,
75 we developed Emotif Alpha that integrates a number of the current methods based on
76 the aforementioned models and filters the motifs using a Z-test, random forest feature
77 importance, and sequence homology.

78

79 Gene promoter regions play a crucial role in gene regulation yet remain largely
80 uncharacterized in *B. malayi*. Among the very few promoters that have been previously
81 described and validated in *B. malayi* is that of Heat Shock Protein 70 (HSP70) (Shu et
82 al. 2003). A previous study showed that while the regulatory domains of the HSP70
83 promoter were similar to other eukaryotes, the core promoter domains appeared to be
84 distinct (Higazi et al., 2005). And nothing is known about motifs regulating
85 developmentally expressed genes in *B. malayi*. There is thus a need for systematic
86 identification, annotation, and experimental validation of *B. malayi* promoter motifs

87 associated with gene regulation to better characterize filaria gene expression patterns.
88 To better understand how promoter elements regulate stage-specific gene expression,
89 we performed differential gene expression analysis of the L3 to L4 molt, the first
90 developmental step important for the establishment of infection in the mammalian host,
91 and motif discovery using the Emotif Alpha pipeline. Several promoter motifs appeared
92 to be associated with the regulation of the L3 to L4 molt. Our results provide an initial
93 overview of the putative regulatory mechanisms in the filariae that could be targeted
94 using novel intervention strategies for control.

95

96 **Results**

97 **Stage-specific expression of serpins, peptidases, cysteine protease inhibitors,** 98 **and structural constituents of the cuticle during the L3 to L4 molt**

99 Since the L3 to L4 molt is of particular interest because it corresponds to the life cycle
100 stage when infective larvae establish infection, we focused in this study on identifying
101 genes that are differentially expressed during this unique process. We used RNA-seq to
102 profile transcription at different time points during the molt, collecting samples from the
103 infective L3 (from mosquitoes), L3 Day 6, and L3 Day 9 worms recovered from infected
104 gerbils (NCBI PRJNA557263). We combined this transcriptome data with previously
105 published L4 data (Grote et al. 2017) that corresponds to Day 14 post infection of
106 gerbils (**Table 1**). In total, 2.36 billion reads were generated, with 1.38 billion reads
107 mapping to the *B. malayi* genome. Each biological replicate received an average of 272
108 million reads, with an average of 173 million reads that were successfully mapped

109 (Table 1). Of the 11,841 *B. malayi* gene models, 87.6% were expressed in at least one
110 stage of the L3 to L4 molt (Fig. 1). The molting expression data shows unique stage-
111 specific profiles for each stage of the molt with significant differential expression
112 between days.

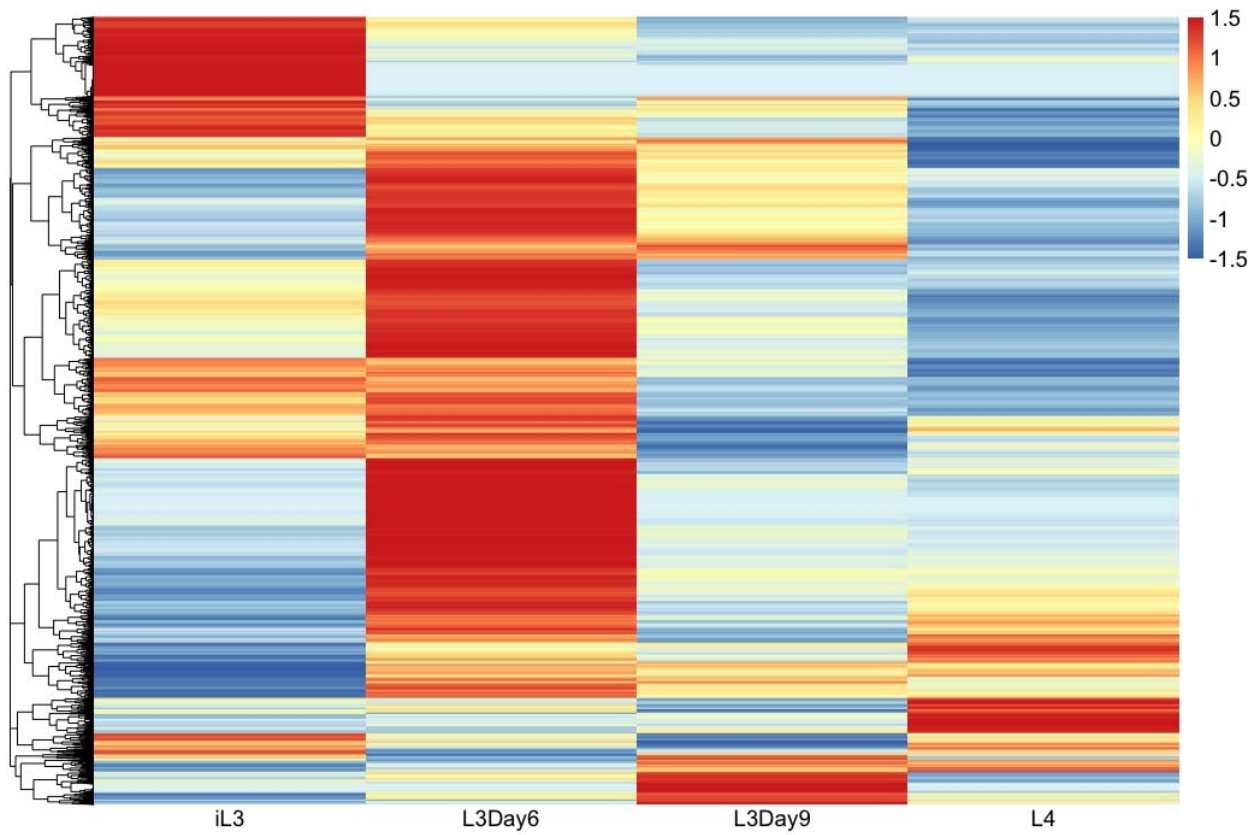
113

Library	Total reads (millions)	Stage Total (millions)	Total Mapped Reads (millions)	Stage Total Mapped Reads (millions)	% <i>B. malayi</i> Genes Expressed
L3a	294	763	266	639	84
L3b	223		204		
L3c	246		169		
L3D6a	508	971	131	477	82.3
L3D6b	243		152		
L3D6c	220		194		
L3D9a	300	440	170	266	81.2
L3D9b	140		96		
L4a	104	195	47	96	72.4
L4b	90		49		

114

115 **Table 1:** RNA-Seq summary. The table shows the total reads sequenced and mapped
116 in each biological replicate at each developmental stage, L3 to L4; lower case a, b, and
117 c refer to separate biological replicates.

118



119 **Figure 1: Expression of *Brugia malayi* genes during the L3 to L4 molt.** Expression is in
120 FPKMs and is Z-scale normalized by row prior to clustering. High expression is indicated by red
121 and low expression by blue. Time-points included infective L3 larvae (iL3), L3 larvae at Day 6 of
122 molting (L3D6), L3 larvae at Day 9 of molting (L3D9), and L4 larvae. Biological replicates have
123 been combined.

124

125 We determined differentially expressed genes during the L3 to L4 molt using both
126 DESeq (Anders & Huber, 2010) and EdgeR (Robinson et al. 2010) to perform pairwise
127 comparisons between the four samples. To get a high-confidence list of differentially
128 expressed genes, we used the consensus of the two algorithms. For the purposes of
129 this study, we focused on the genes that were up-regulated at each stage of molting, as

130 compared to the other stages, and did a gene-annotation enrichment analysis for each
131 stage. We found that up-regulated genes in iL3 larvae, as compared to L3D6, L3D9,
132 and L4, were enriched for annotations involving cysteine-type peptidase activity as well
133 as serpin domains and serpin family proteins. Cysteine-type peptidases are essential for
134 molting in *B. malayi* (Guiliano D.B. et al. 2004, Lustigman S. et al. 2004) and serpins are
135 serine protease inhibitors that have previously been shown to be involved in
136 immunomodulation and host immune evasion during infection (Zang X, et al. 2001). We
137 identified five different cysteine-type peptidases and two cysteine-type endopeptidase
138 inhibitors that were upregulated in iL3 larvae. By day 6 of molting, structural constituents
139 of the cuticle, including collagen (the main component of the cuticle) were enriched in
140 the up-regulated gene sets. We also see the up-regulation of several metalloproteases.
141 At day 9, genes involved in signaling were enriched among the up-regulated genes, as
142 were several different metalloproteases. At day 14 (L4 larvae), we again see an
143 enrichment of structural constituents of the cuticle. Similarly to those enriched in L3 day
144 6 larvae, they are all mostly orthologs of *C. elegans* col (COLlagen) genes, which are
145 themselves orthologs of human MARCO genes (macrophage receptor with collagenous
146 structure). The set of structural constituents enriched at day 14 is, however, a
147 completely unique set of collagen genes as compared to the genes observed at day 6.
148 These stage-specific enrichments reflect the order of peptidases and structural
149 constituents necessary for the building of a new L4 cuticle, the separation of the old L3
150 cuticle from the developing L4 cuticle, and the shedding of the old L3 cuticle.

152 **Identification of 12 motifs associated with transcription factor binding that are**
153 **enriched in the L3 to L4 molt**

154 To better understand the regulatory program of *B. malayi* during the L3 to L4 molt, we
155 analyzed statistically over-represented DNA motifs in regions upstream of genes that
156 were upregulated during molting. To do so, we developed a motif identification pipeline
157 called Emotif Alpha (**Fig. 2**). First, we used the transcriptome data from the different
158 stages of the L3 to L4 molt to generate lists of genes up-regulated at each stage of the
159 molt using pair-wise comparisons. We then did a motif discovery analysis on each gene
160 set using a combination of three motif discovery tools: GimmeMotifs, DME, and DECOD.
161 GimmeMotifs is an ensemble of generative motif discovery tools—including Homer
162 (Heinz et al. 2010), AMD (Shi et al. 2011), BioProspector, MDmodule (Conlon et al.
163 2003), MEME, Weeder, GADEM (Li et al. 2009), and Improbizer—while DME and
164 DECOD are discriminative motif discovery tools. We did a discriminative motif discovery
165 analysis by randomly selecting background promoter region sets from all *B. malayi*
166 genes, excluding the differentially expressed genes. These background sets are three
167 times larger than the foreground sets. We selected motif lengths between 6- and 15-mer.
168 In total, we identified 20,025 motifs.

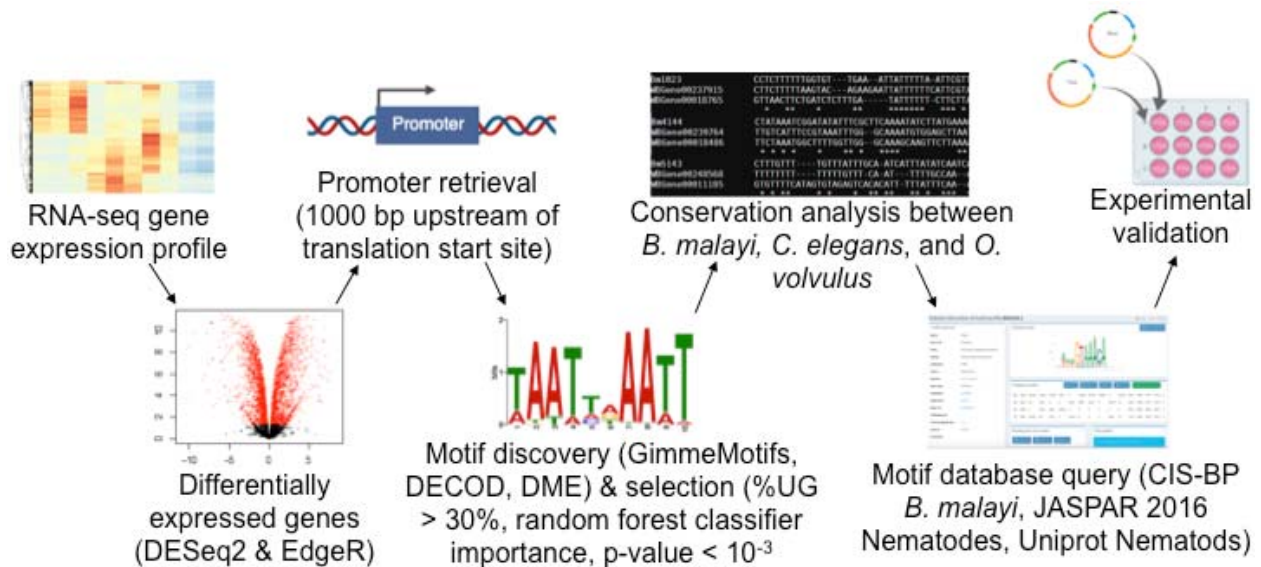


Figure 2: Workflow of promoter motif identification. The six main steps for motif

discovery were: 1) generation of an RNA-seq profile, 2) determination of up-regulated genes for every pairwise comparison using DESeq2 (FDR<0.01) and EdgeR (P-value<0.01), 3) promoter retrieval: 1000 bp upstream of the translation start site, 4) ensemble motif discovery using GimmeMotifs (Homer, AMD, BioProspector, MDmodule, MEME, Weeder, GADEM, and Improbizer), DECOD, DME, and selection of enriched motifs: %UG > 30% and random forest classifier feature importance and over-representation p-value < 10⁻³, 5) TFBS conservation analysis between *B. malayi*, *C. elegans*, and *O. volvulus*. 6) motif database query: CIS-BP *B. malayi*, JASPAR 2016 Nematodes, Uniprot Nematodes. Finally, a subset of those identified motifs were experimentally validated.

169

170 To select statistically significant motifs, we first assessed the motifs by a random forest
171 classifier using scikit-learn (Varoquaux et al. 2015). The random forest algorithm uses
172 bootstrap sampling and constructs a decision tree for each sub-sample. To evaluate the
173 motifs, we used both Gini impurity (Gordon et al.1984) and information gain (Quinlan et
174 al. 1993) criteria and retained the union of the resulting top 40 motifs. We further filtered
175 the motifs by foreground coverage (i.e. UG%), removing motifs occurring in less than
176 30% of the genes up-regulated at that stage of molting. We then used a Z-test to

177 compare the frequency of a motif in the up-regulated genes with the expected frequency
178 in the background promoters. Using a significance level (p-value) cutoff of 10^{-3} , we
179 selected 395 motifs.








180




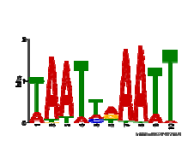

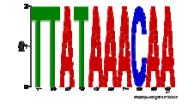
181 We retrieved a collection of 163 known nematode transcription factor binding sites
182 (TFBSs) from the MEME suite (<http://meme-suite.org/>), searching the motif databases
183 JASPAR CORE 2016 nematodes (Mathelier et al. 2016), CIS-BP *Brugia malayi*
184 (Weirauch et al. 2014), and Uniprobe worm (Newburger et al. 2009). We matched the
185 remaining motifs to known TFBSs with TOMTOM (Gupta et al. 2007). If two motifs were
186 matched to the same binding site and they were discovered from the same gene list, we
187 considered them to be redundant and kept the one with the lowest over-representation
188 p-value. This step narrowed our list down to 27 motifs that had matches to known
189 binding sites.

190

191 We next performed a conservation analysis amongst nematodes using an adaptation of
192 a published method (Roy et al. 2013). We retrieved orthologous gene information
193 among *B. malayi*, *C. elegans*, and *O. volvulus* from Wormbase ParaSite Biomart (Howe
194 et al. 2017). We extracted up to 1Kb upstream from the translation start sites for *B.*
195 *malayi* genes, assuming these regions would contain the promoter. We performed
196 multiple sequence alignments using CLUSTALW2 (Larkin et al. 2007), and defined a
197 motif as conserved if it occurred at the same position in the orthologous promoter region
198 alignment of either *C. elegans* or *O. volvulus*. This step resulted in 12 remaining motifs
199 (**Table 2**) that were (1) enriched (p-value < 10^{-3}) and (2) conserved in either *C. elegans*

200 or *O. volvulus*. The frequency of motif occurrence in the putative promoter regions of
 201 up-regulated genes ranges from 33% to 94%. The fold enrichment, representing the
 202 ratio between motif frequencies in the up-regulated gene promoters and background
 203 promoters, ranges from 1.28 to 2.19.

Name	Logo	%UG ^a	Ratio ^b	P-value	Matched known motif	transcription factor (TF)	Function	Gene pair of conserved sites
M1.1 <L4,L3>		64	1.63	1.4*10 ⁻⁴	MA0928.1 (<i>C. elegans</i>)	zfh-2	Involved in hermaphrodite genitalia development, locomotion, nematode larval development and receptor-mediated endocytosis	[Bm4560, OVOC6906], [Bm856, OVOC3639]
M1.2 <L3D9,L3>		84.1	1.31	7.1*10 ⁻⁵				[Bm856, OVOC3639], [Bm17348, OVOC8896], [Bm17988, OVOC3292]
M1.3 <L3D6,L3>		94.4	1.28	7.6*10 ⁻⁶				[Bm2559, C34C6.3], [Bm2821, OVOC10446], [Bm17348, OVOC8896], [Bm3341, OVOC10396], [Bm4257, OVOC2553]
M1.4 <L3D9,L3D6>		84.7	1.55	9.6*10 ⁻⁸				[Bm7179, OVOC394], [Bm2270, OVOC2391]
M2 <L3D6,L4>		93.3	1.31	2.9*10 ⁻⁵	MA0927.1 (<i>C. elegans</i>)	vab-7	Required for DB motorneuron identity and posterior DB axonal polarity	[Bm4904, OVOC2123]
M3.1 <L3D9,L3D6>		80.6	1.33	3.0*10 ⁻⁴	MA0542.1 (<i>C. elegans</i>)	elt-3	Controlling hypodermal cell differentiation	[Bm2802, OVOC9504]
M3.2 <L3D9,L3>		81.8	1.32	8.3*10 ⁻⁵				[Bm10655, C28A5.3, OVOC9600]

M4.1 <L4,L3D6>		34.8	2.19	4.9×10^{-4}	MA0537.1 (<i>C. elegans</i>)	blimp-1	Loss of blmp-1 activity via deletion mutation has been reported to result in small, dumpy animals with abnormal fat content	[Bm2802, OVOC9504], [Bm6190, OVOC827]
M4.2 <L4,L3>		56	1.71	1.6×10^{-4}				[Bm7019, OVOC7405], [Bm6190, OVOC827]
M4.3 <L3D9,L3D6>		38.9	2.01	1.0×10^{-5}				[Bm2802, OVOC9504], [Bm7179, OVOC394]
M5* <L3,L3D6>		93.3	1.23	2.9×10^{-4}	M5348_1.02 (CIS-BP <i>Brugia</i> inferred)	Bm8528	Retinal homeobox protein Rx3	[Bm1938, OVOC2080], [Bm8228, C27D6.4], [Bm1559, OVOC7210], [Bm4184, OVOC3386]
M6* <L3,L3D6>		33.3	2.03	3.0×10^{-5}	M5221_1.02 (CIS-BP <i>Brugia</i> inferred)	Bm4429	Involved in regulation of transcription, DNA-templated and steroid hormone mediated signaling pathway	[Bm1938, OVOC2080], [Bm6642, Y48B6A.12]
M7* <L3,L3D9>		54.5	1.19	6.1×10^{-2}	M0739_1.02 (CIS-BP inferred)	Bm3608	Involved in cell growth, proliferation, differentiation, and longevity.	[Bm4184, OVOC3386]

204 **Table 2:** Table of enriched promoter motifs over the L3-L4 molt. Motifs were found to be
205 enriched (p -value $< 10^{-3}$) in the upstream elements of up-regulated genes between different
206 stages of molting and to be conserved (in either *C. elegans* or *O. volvulus*). ^aFrequency of a
207 motif in up-regulated gene promoters. ^bRelative frequency of a motif in up-regulated gene
208 promoters vs. background promoters. *These two motifs have been validated experimentally.
209 Note that M7 was included in the experimental validation because it passed 4 out of 5 filters,
210 including foreground coverage filter, random forest filter, known motif filter and conservation
211 filter. However, it was not included in the 12 reported motifs due to its non-significant p -value.

212

213 The 12 selected motifs matched known binding sites for 6 transcription factors in *C.*
214 *elegans* (**Table 2**), all of which are involved in development, aging, and/ or movement.
215 Motifs M1.1, M1.2, M1.3 and M1.4 matched a zinc-finger protein, zfh-2, which is
216 involved in hermaphrodite genitalia development, locomotion, nematode larval
217 development and receptor-mediated endocytosis. Motif M2 matched vab-7, which is
218 associated with DB motor neuron identity and posterior DB axonal polarity. Motifs M3.1
219 and M3.2 matched elt-3, which is related to aging (Budovskaya et al. 2008). Motifs
220 M4.1, M4.2, and M4.3 matched blmp-1. M5 matched a homeobox protein, Bm8528, and
221 M6 matched a nuclear receptor Bm4429. The 12 motifs are conserved in either *C.*
222 *elegans* or *O. volvulus* (**Table 2**, last column). Moreover, the occurrence of M3.2 in the
223 Bm10655 promoter region is conserved in orthologs in both *C. elegans* (promoter of
224 C28A5.3) and *O. volvulus* (promoter of OVOC9600).

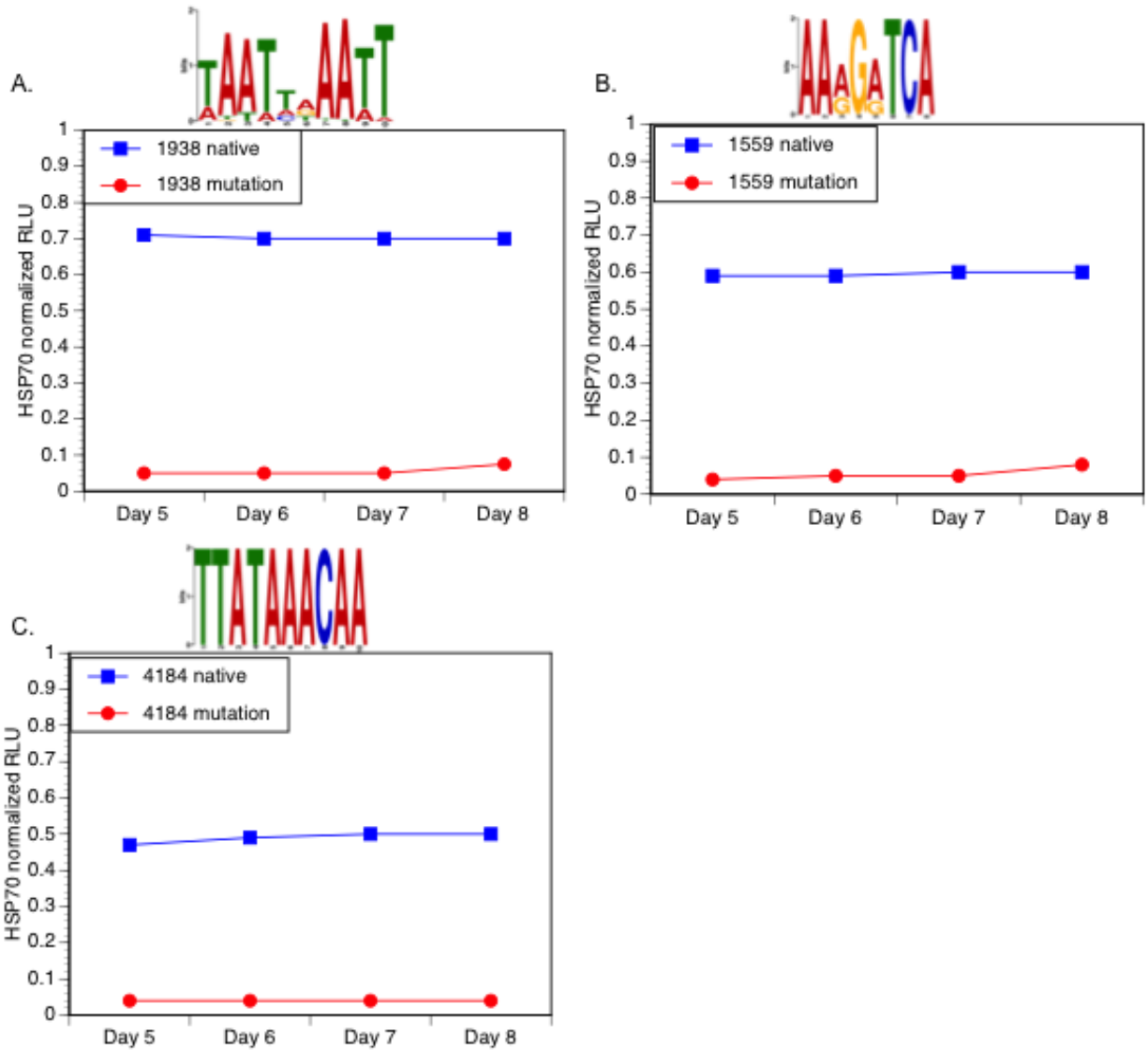
225

226 The motif analysis reveals how some of the differential expression of different proteases
227 may be orchestrated during the L3 to L4 molt. Motif M1.4 is found in the promoter region
228 of Bm2270 a metalloprotease significantly up-regulated in L3D9 worms. Bm2270 is an
229 ortholog of nas-37 in *C. elegans* and has been shown to be involved in collagen and
230 cuticulin-based cuticle development and ecdysis. Motif M5 is found in the promoter
231 region of Bm1938 and is predicted to encode a serpin. Bm1938 is one of the serpins
232 that was found to be significantly up-regulated in the iL3 larvae.

233

234 **L3 stage-specific transcription factor binding motifs can be validated *in vitro***

235 Three of the motifs (M5, M6, and M7) were chosen for validation based on their
236 enrichment in the promoters of genes up-regulated in the mid to late stages of the L3 to
237 L4 molt. Three separate genes, each containing one chosen motif, were tested. The 1
238 kbp upstream region of each gene was amplified from *B. malayi* genomic DNA and
239 cloned upstream of the firefly luciferase reporter gene in the expression vector pGL3
240 Basic (Shu et al. 2003). *B. malayi* L3 were then transfected with the constructs in a co-
241 culture system as previously described (Liu et al. 2018). The parasites were induced to
242 molt *in vitro* and then assayed for luciferase activity. The number of relative light units
243 (RLUs) observed were normalized to those obtained from parasites transfected in
244 parallel with a construct consisting of the *B. malayi* HSP70 promoter driving the
245 expression of the firefly luciferase reporter (Shu et al. 2018). The experiment was
246 performed with both the native promoter and a mutant promoter where the nucleotides
247 of the motif had been randomly shuffled. All of the native promoters produced significant
248 amounts of reporter luciferase activity in the molting parasites (ranging from 40%- 70%
249 of the activity produced by the HSP70 construct transfected positive controls; **Fig. 3**).
250 However, when the putative motifs were mutated, the activity in all the promoters tested
251 decreased by 80-90% (**Fig. 3**).



252 **Figure 3: Promoter motif validation.** A) Promoter motif validation in L3 worms that were
253 molting in vitro using the native promoter of Bm1559 and a mutated motif M5 version of the
254 same promoter. B) Promoter motif validation using the native promoter of Bm1938 and a
255 mutated motif M6 version of the promoter. C) Promoter motif validation using the native
256 promoter of Bm4184 with a mutated motif M7 version of the promoter. In each panel luciferase
257 activity obtained from the constructs is normalized against parasites transfected with a construct
258 containing the Bm HSP70 promoter driving the expression of the firefly luciferase reporter.

260 **Discussion**

261 Parasitic nematodes such as *B. malayi* maintain a complicated lifecycle involving both
262 an insect and a mammalian host, and undergo a number of molts. The L3 to L4 molt
263 that occurs immediately upon infection of the mammalian host is of particular interest as
264 it marks the establishment of infection and thus represents an attractive point for drug
265 intervention. Little is known, however, about how *B. malayi* regulates the transitions
266 between these stages. Prior to our study, nothing was known about promoter motifs that
267 regulate developmentally expressed genes in *B. malayi*. Because stage transitions rely
268 on precise transcriptional control through the interaction of transcription factors and their
269 binding sites, we set out to characterize the potential transcription factor binding motifs
270 of these parasites and to identify motifs that contribute to stage-specific expression of
271 genes involved in early worm development in the mammalian host.

272

273 We found a number of enriched motifs and were able to define both conserved motifs
274 across molting as well as stage-specific motifs. While some of the motifs we identified
275 are conserved in other nematodes, such as *C. elegans*, a number of motifs represent
276 novel binding sites potentially reflecting the differences in development and the parasitic
277 lifestyle. It is known that molting is regulated by an ecdysone-like response system
278 (Barker et al. 1991; Mhashilkar et al. 2016; Mhashilkar et al. 2016; Warbrick et al. 1993,
279 Lui et al, 2012). Two of the identified motifs and their cognate TFs appear to be related
280 to the ecdysone response. For example, *zfh-2*, the transcription factor predicted to bind
281 four of our identified motifs, is a common cofactor implicated in ecdysone signaling in *D.*
282 *melanogaster* (Davis et al. 2011). *Blimp-1*, the transcription factor predicted to bind

283 three of our identified motifs, is an ecdysone-inducible repressor that is essential for the
284 prepupal development in *Drosophila* (Akagi and Ueda 2011). Validation results suggest
285 that our pipeline is able to identify biologically-relevant motifs involved in molting. This
286 analysis provided biological insight into the development of the parasite as well as the
287 identification of novel drug targets.

288

289 Future work needs to be done to expand this analysis across the lifecycle of the
290 nematode at different stages of development, in its different hosts (i.e. human vs.
291 mosquito). While transcriptomic data from these stages exists and can be used to
292 predict motifs, validation at other stages *in vivo* will prove more difficult. However, with
293 recent innovation in filarial transgenics, it is now possible to create stable transgenic
294 parasite lines that will allow functional testing *in vivo* of defined promoter motifs at all life
295 stages of the parasite (Liu et al. 2018).

296

297 **Materials and Methods**

298 **Transcriptomic study design**

299 All parasites were obtained from FR3 (Filariasis Research Reagent Resource Center;
300 BEI Resources, Manassas, VA, USA), where they were isolated from infected gerbils
301 (*Meriones unguiculatus*) or mosquitoes (*Aedes aegypti*). Worms were flash-frozen and
302 shipped to the New York Blood Center for processing. For transcriptomic sequencing,
303 infective third-stage larvae (iL3) were recovered from mosquitoes and mammalian stage
304 larvae were recovered from gerbils at 6 and 9 days post infection (dpi). At 6 dpi, larvae

305 are typically undergoing the molt from L3 to L4, while by 9 dpi the molt is complete
306 (Mutafchiev et al. 2014). Data was combined with previously published stages 14 dpi
307 (L4) (Grote et al. 2017).

308

309 **RNA isolation, library preparation and sequencing**

310 Total RNA was prepared from *B. malayi* worms as previously described (Grote et al.
311 2017). RNA was prepared from 3 biological replicates of infective L3 (iL3; 2000 larvae
312 each), 3 replicates of 6 dpi larvae (1500 each) and 2 replicates of 9 dpi larvae (1300
313 each). *B. malayi* worms were homogenized in Trizol (ThermoFisher) using a hand-held
314 pestle in 1.5mL tubes containing the worms. Total RNA was extracted by organic
315 extraction using Trizol and the PureLink RNA mini kit (ThermoFisher) and after being
316 treated with DNaseI (New England Biolabs). Ribosomal RNA (rRNA) depletion was
317 performed using Terminator (Epicentre), a 5'-phosphate-dependent exonuclease that
318 degrades transcripts with a 5' monophosphate. Libraries were prepared using the
319 NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs) according to
320 manufacturer instructions. Library quality was assessed using a D1000 ScreenTape
321 Assay (Agilent) prior to sequencing. Library concentrations were assessed using the
322 qPCR library quantification protocol (KAPA biosystems). Libraries were sequenced on
323 the Illumina NextSeq500 platform with 150bp paired-end reads. To minimize the
324 confounding effects of lane-to-lane variation, libraries were multiplexed and sequenced
325 with technical replicates on multiple lanes. Each biological replicate received an
326 average of 135 million mapped reads (PRJNA557263).

327

328 **Sequencing alignment and expression analysis**

329 Read quality was assessed using FastQC (Babraham Bioinformatics). Sequence reads
330 from each sample were analyzed with the Tuxedo suite of tools (Kim et al. 2013;
331 Trapnell et al. 2013; Trapnell et al. 2010). Reads were mapped with Tophat2's Bowtie2-
332 very-sensitive algorithm to the annotated *B. malayi* genome assembly (WormBase.org).
333 The resulting BAM files were then used with HtSeq to obtain raw read counts.
334 Differential gene expression analysis was performed using both DESeq and EdgeR,
335 and the overlapping genes (FDR < 0.01 & P-value < 0.01) were retained. Up-regulated
336 genes were characterized for each pairwise comparison between L3, L3 Day 6 (L3D6),
337 L3 day 9 (L3D9), and L4 worms. For example, <L3,L4> refers to the up-regulated genes
338 in L3 compared to L4. Two pairs of comparisons, <L3D9,L4> and <L4,L3D9>, were
339 dropped due to the limited number of up-regulated genes (≤ 3), possible due to the
340 L3D9 being actually younger L4. The up-regulated gene lists were filtered using log2
341 fold change (logFC) with the following thresholds: $|\logFC| = 7$ for <L3D6,L3>,
342 <L3D6,L3D9>, <L3D6,L4>; $|\logFC| = 4$ for <L3D9,L3>, <L3,L3D6>, <L3,L3D9>,
343 <L3,L4>, <L4,L3>; $|\logFC| = 2.5$ for <L3D9,L3D6>, <L4,L3D6>. The reason for varying
344 the threshold was that the number of up-regulated genes in each list varied significantly;
345 for motif discovery tools to search efficiently, the number of sequences were limited to
346 less than one hundred. In total, 10 up-regulated gene lists were used for motif discovery
347 **(Table S1).**

348

349 Potential promoter sequences were retrieved from WormBase ParaSite Biomart (Howe
350 et al. 2016) web interface, capturing the 1000bp upstream of the translation start site for
351 each gene.

352

353 **The Emotif Alpha pipeline for regulatory motif identification**

354 The Emotif Alpha pipeline (freely available at: https://github.com/YichaoOU/Emotif_Alpha)
355 was developed to automate motif discovery analysis for the 10 up-regulated gene lists.
356 This pipeline was written in python and was applied to perform all aforementioned motif
357 analyses. The motif discovery step used multiple tools and was run in parallel at the
358 Ohio SuperComputer Center. Motif length search was from 6 to 14. Motif scanning was
359 done using FIMO (Grant et al. 2011) with a default p-value threshold of 10^{-4} . We
360 implemented 5 different motif filters. (1) Foreground coverage (i.e., UG%) was defined
361 as the proportion of up-regulated gene promoters containing the given motif. We set a
362 minimal foreground coverage at 30%. (2) Motifs were then filtered by a random forest
363 classifier. The union of the top 40 motifs that resulted from either Gini impurity or
364 information gain criterion was retained. (3) Motif enrichment p-value was calculated
365 using Z-test and the cutoff was 10^{-3} . (4) Known motif filter was performed using
366 TOMTOM and a collection of 163 known nematode TFBSs. The motif similarity p-value
367 threshold was 10^{-4} . (5) Conservation analysis was performed using a method described
368 in (Roy et al. 2013). Only conserved motifs were kept.

369

370 ***In vitro* validation of promoter transcription motifs**

371 The putative TF motifs M5, M6, and M7 were chosen for validation based on their
372 enrichment in the promoters of genes upregulated in the mid to late stages of the L3 to
373 L4 molt. Three different genes, each containing one of the chosen motifs, were used for
374 the validation assay. As previously described in Shu et al. (Shu et al. 2003), we
375 amplified the 1 kbp region upstream of each gene from *B. malayi* genomic DNA and
376 cloned upstream of the firefly luciferase reporter gene in the expression vector pGL3
377 Basic. We then transfected *B. malayi* L3 larvae with the constructs in a co-culture
378 system as previously described (Liu et al. 2018). The parasites were induced to molt by
379 the addition of ascorbic acid on day 5, and parasites were assayed for luciferase activity
380 on days 5-8, as by day 9 the molting was complete. We normalized the number of RLUs
381 observed to those obtained from parasites transfected in parallel with a construct
382 consisting of the *B. malayi* HSP70 promoter driving the expression of the firefly
383 luciferase reporter (Shu et al. 2018) to control for accumulation of the firefly luciferase
384 over time during the duration of the experiment. We did the experiment with both
385 the native promoter and a mutant promoter where the nucleotides of the motif had
386 been randomly shuffled (**Table S2**).

387

388 **Funding**

389 This work was supported by grant US National Institute of Allergy and
390 Infectious Diseases (NIAID) grant number R21 AI135172-01A1 to TRU, R56 AI101372
391 to EG, TRU, and SL, and R21 AI126466 to EG. LW received funding from the *The Ohio*
392 *University GERB Program*. AG received funding from the T32 Ruth L. Kirschstein

393 Institutional National Research Service Award (T32AI007180) and the F31 Ruth L.

394 Kirschstein Pre-doctoral Individual NRSA (F31AI131527).

395

396

397

References

- 398 Akagi, K., Ueda, H. (2011). Regulatory Mechanisms of Ecdysone-Inducible Blimp-1
399 Encoding a Transcriptional Repressor that is Important for the Prepupal
400 Development in *Drosophila*. *Development, Growth & Differentiation*, 53 (5): 697–
401 703.
- 402 Anders, S., & Huber, W. (2010). Differential Expression Analysis for Sequence Count
403 Data. *Genome Biology*, 11(R106).
- 404 Ao, W., Gaudet, J., Kent, W. J., Muttumu, S., & Mango, S. E. (2004). Environmentally
405 Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*,
406 305(5691), 1743–1746. <https://doi.org/10.1126/science.1102216>
- 407 Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: Discovering and
408 Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Research*, 34(suppl_2),
409 W369–W373. <https://doi.org/10.1093/nar/gkl198>
- 410 Barker, G. C., J. G. Mercer, H. H. Rees, & R. E. Howells. (1991). The Effect of
411 Ecdysteroids on the Microfilarial Production of *Brugia pahangi* and the Control of
412 Meiotic Reinitiation in the Oocytes of *Dirofilaria immitis*. *Parasitology Research*, 77
413 (1): 65–71
- 414 Bennuru, S., Cotton, J.A., Ribeiro, J.M., Grote, A., Harsha, B., Holroyd, N., Mhashilkar,
415 A., Molina, D.M., Randall, A.Z., Shandling, A.D., Unnasch, T.R., Ghedin, E.,
416 Berriman, M., Lustigman, S., Nutman, T.B. (2016). Stage-Specific Transcriptome
417 and Proteome Analyses of the Filarial Parasite *Onchocerca volvulus* and Its
418 *Wolbachia* Endosymbiont. *mBio*, 7(6). PMC5137501.

- 419 Breiman, L., Friedman, J., Stone, C., Olshen, R.A. Classification and Regression Trees.
420 Taylor & Francis, Jan 1, 1984.
- 421 Budovskaya, Y.V., Wu, K., Southworth, L., Jiang, M., Tedesco, P., Johnson, T., & Kim,
422 S. (2008). An Elt-3/elt-5/elt-6 GATA Transcription Circuit Guides Aging in *C.*
423 *Elegans*. *Cell*, 134 (2): 291–303.
- 424 Choi, Y.-J., Ghedin, E., Berriman, M., McQuillan, J., Holroyd, N., Mayhew, G. F., ...
425 Michalski, M. L. (2011). A Deep Sequencing Approach to Comparatively Analyze
426 the Transcriptome of Lifecycle Stages of the Filarial Worm, *Brugia malayi*. *PLOS*
427 *Neglected Tropical Diseases*, 5(12), 1–10.
428 <https://doi.org/10.1371/journal.pntd.0001409>
- 429 Conlon, E.M., Liu, X.S., Lieb, J., & Liu, J. (2003). Integrating Regulatory Motif Discovery
430 and Genome-Wide Expression Analysis. *Proceedings of the National Academy of*
431 *Sciences of the United States of America*, 100 (6): 3339–44.
- 432 Cotton, J. A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., Doyle, S., et al.
433 (2016). The Genome of *Onchocerca volvulus*, Agent of River Blindness. *Nature*
434 *Microbiology*, 2 (2). <https://doi.org/10.1038/nmicrobiol.2016.216>.
- 435 Das, M. K., & Dai, H.-K. (2007). A Survey of DNA Motif Finding Algorithms. *BMC*
436 *Bioinformatics*, 8(7), S21. <https://doi.org/10.1186/1471-2105-8-S7-S21>
- 437 Davis, M.B., SanGil, I., Berry, G., Olayokun, R., & Neves, L. (2011). Identification of
438 Common and Cell Type Specific LXXLL Motif EcR Cofactors Using a Bioinformatics
439 Refined Candidate RNAi Screen in *Drosophila Melanogaster* Cell Lines. *BMC*
440 *Developmental Biology*, 11 (November): 66.

- 441 Desjardins, C. A., et al. (2013). Genomics of *Loa loa*, a Wolbachia-free Filarial Parasite
442 of Humans. *Nature Genetics*, 45(5): 495-500.
- 443 Dieterich, C., & Sommer, R.J. (2008). A Caenorhabditis Motif Compendium for Studying
444 Transcriptional Gene Regulation. *BMC Genomics*, 9(1), 30.
445 <https://doi.org/10.1186/1471-2164-9-30>
- 446 Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya,
447 A., et al. (2005). The Wolbachia Genome of *Brugia Malayi*: Endosymbiont Evolution
448 within a Human Pathogenic Nematode. *PLoS Biology*. 3 (4): e121.
- 449 Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., ... Scott, A. L. (2007).
450 Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science*, 317(5845),
451 1756–1760. <https://doi.org/10.1126/science.1145406>
- 452 Gordon, A. D., Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984).
453 Classification and Regression Trees. *Biometrics*. <https://doi.org/10.2307/2530946>.
- 454 Gordon, C.A., Jones, M.K., McManus, D.P. (2018) The history of Bancroftian Lymphatic
455 Filariasis in Australasia and Oceania: Is there a Threat of Re-occurrence in Mainland
456 Australia? *Tropical Medicine and Infectious Disease*, 3(2), 58.
- 457 Grant C.E., Bailey T.L., Noble W.S. (2011). FIMO: Scanning for Occurrences of a Given
458 Motif. *Bioinformatics*, 27(7):1017–1018. doi:10.1093/bioinformatics/btr064
- 459 Grote, A., Lustigman, S., Ghedin, E. (2017) Lessons from the genomes and
460 transcriptomes of filarial nematodes. *Molecular and Biochemical Parasitology*,
461 215:23-29. doi: 10.1016/j.molbiopara.2017.01.004.

- 462 Grote, A., Voronin, D., Ding, T., Twaddle, A., Unnasch, T., Lustigman, S., & Ghedin, E.
463 (2017). Defining *Brugia Malayi* and *Wolbachia* Symbiosis by Stage-Specific Dual
464 RNA-Seq. *PLoS Neglected Tropical Diseases*, 11 (3): e0005357.
- 465 Guiliano, D.B., Hong, X., McKerrow, J.H., Blaxter, M.L., Oksov, Y., Liu, J., Ghedin, E.,
466 Lustigman, S. (2004). A Gene Family of Cathepsin L-like Proteases of Filarial
467 Nematodes are Associated with Larval Molting and Cuticle and Eggshell
468 Remodeling. *Mol Biochem Parasitol*, 136: 227-42.
- 469 Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., & Noble, W. (2007). Quantifying
470 Similarity between Motifs. *Genome Biology*, 8 (2): R24.
- 471 Heeringen, S. J. Van, & Veenstra, G. J. C. (2011). GimmeMotifs: a De Novo Motif
472 Prediction Pipeline for ChIP-sequencing Experiments. *Bioinformatics*, 27(2), 270–
473 271. <https://doi.org/10.1093/bioinformatics/btq636>
- 474 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y., Laslo, P., Cheng, J., Murre, C.,
475 Singh, H., & Glass, C. (2010). Simple Combinations of Lineage-Determining
476 Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage
477 and B Cell Identities. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- 478 Higazi, T. B., DeOliveira, A., Katholi, C. R., Shu, L., Barchue, J., Lisanby, M., &
479 Unnasch, T. R. (2005). Identification of Elements Essential for Transcription in
480 *Brugia malayi* Promoters. *Journal of Molecular Biology*, 353(1), 1–13.
481 <https://doi.org/https://doi.org/10.1016/j.jmb.2005.08.014>
- 482 Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase
483 ParaSite – a Comprehensive Resource for Helminth Genomics. *Molecular and*

- 484 *Biochemical Parasitology*.
- 485 <https://doi.org/https://doi.org/10.1016/j.molbiopara.2016.11.005>
- 486 Huggins, P., Zhong, S., Shiff, I., Beckerman, R., Laptenko, O., Prives, C., ... Bar-
487 Joseph, Z. (2011). DECOD: Fast and Accurate Discriminative DNA Motif Finding.
488 *Bioinformatics*, 27(17), 2361–2367. <https://doi.org/10.1093/bioinformatics/btr412>
- 489 Jin, V. X., Apostolos, J., Nagisetty, N. S. V. R., & Farnham, P. J. (2009). W-ChIPMotifs:
490 a Web Application Tool for De Novo Motif Discovery from ChIP-Based High-
491 Throughput Data. *Bioinformatics*, 25(23), 3191–3193.
- 492 Kariuki, M. M., Hearne, L. B., & Beerntsen, B. T. (2010). Differential Transcript
493 Expression between the Microfilariae of the Filarail Nematodes, *Brugia malayi* and
494 *B. pahangi*. *BMC Genomics* 11(1), 225. <https://doi.org/10.1186/1471-2164-11-225>
- 495 Kim, D., Pertea, D., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. (2013).
496 TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions,
497 Deletions and Gene Fusions. *Genome Biology*, 14 (4): R36.
- 498 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam,
499 H., Valentin, F., et al. (2007). Clustal W and Clustal X Version 2.0.
500 *Bioinformatics*, 23 (21): 2947–48.
- 501 Lee, N. K., Azizan, F. L., Wong, Y. S., & Omar, N. (2018). DeepFinder: An Integration of
502 Feature-Based and Deep Learning Approach for DNA Motif Discovery.
503 *Biotechnology & Biotechnological Equipment*, 1–10.
- 504 Li, B.-W., Wang, Z., Rush, A. C., Mitreva, M., & Weil, G. J. (2012). Transcription
505 Profiling Reveals Stage- and Function-Dependent Expression Patterns in the

- 506 Filarial Nematode *Brugia malayi*. *BMC Genomics*, 13(1), 184.
507 <https://doi.org/10.1186/1471-2164-13-184>
- 508 Li, Leping. (2009). GADEM: A Genetic Algorithm Guided Formation of Spaced Dyads
509 Coupled with an EM Algorithm for Motif Discovery. *Journal of Computational*
510 *Biology: A Journal of Computational Molecular Cell Biology*, 16 (2): 317–29.
- 511 Liu, X., Brutlag, D. L., & Liu, J. S. (2001). BioProspector: Discovering Conserved DNA
512 Motifs in Upstream Regulatory Regions of Co-expressed Genes. *Pacific*
513 *Symposium on Biocomputing*, 127–138.
- 514 Liu C, Enright T, Tzertzinis G, Unnasch TR. (2012) Identification of Genes Containing
515 Ecdysone Response Elements in the Genome of *Brugia malayi*. *Mol Biochem Parasitol*,
516 186:38-43; PMID: PMC 3501679.
- 517 Liu C., Mhashilkar, A.S., Chabanon, J., Xu, S., Lustigman, S., Adams, J.H., et al. (2018)
518 Development of a Toolkit for *piggyBac*-Mediated Integrative Transfection of the
519 Human Filarial Parasite *Brugia malayi*. *PLoS Negl Trop Dis*, 12(5): e0006509.
520 <https://doi.org/10.1371/journal.pntd.0006509>
- 521 Lustigman S, Zhang J, Liu J, Oksov Y, Hashmi S. (2004). RNA Interference Targeting
522 Cathepsin L and Z-like Cysteine Proteases of *Onchocerca volvulus* Confirmed their
523 Essential Function during L3 Molting. *Mol Biochem Parasitol*, 138: 165-70.
- 524 Mathelier, A., Fornes, O., Arenillas, D., Chen, C., Denay, G., Lee, J., Shi, W., et al.
525 (2016). JASPAR 2016: A Major Expansion and Update of the Open-Access
526 Database of Transcription Factor Binding Profiles. *Nucleic Acids Research*, 44 (D1):
527 D110–15.

- 528 Mhashilkar, A.S., Adapa, S.R., Jiang, R.H., Williams, S.A., Zaky, W., Slatko, B.E., Luck,
529 A.N., Moorhead, A.R., Unnasch, T.R. (2016) Phenotypic and Molecular Analysis of
530 the Effect of 20-Hydroxyecdysone on the Human Filarial Parasite *Brugia malayi*. *Int*
531 *J Parasitol.*, 46(5-6):333-41. doi: 10.1016/j.ijpara.2016.01.005.
- 532 Mhashilkar, A.S., Vankayala, S.L., Lui, C., Kearns, F., Mehrotra, P., Tzertzinis, G., Palli,
533 S.R., Woodcock, H.L., Unnasch, T.R. (2016) Identification of Ecdysone Hormone
534 Receptor Agonists as a Therapeutic Approach for Treating Filarial Infections. *PLoS*
535 *Neglected Tropical Disease*, 10(6):e0004772. doi: 10.1371/journal.pntd.0004772.
- 536 Mutafchiev, Y., Bain, O., Williams, Z., McCall, J.W., & Michalski, M.L. (2014).
537 Intraperitoneal Development of the Filarial Nematode *Brugia Malayi* in the
538 Mongolian Jird (*Meriones Unguiculatus*). *Parasitology Research*, 113 (5): 1827–35.
- 539 Newburger, D. E., & Bulyk, M. L. (2009). UniPROBE: An Online Database of Protein
540 Binding Microarray Data on Protein-DNA Interactions. *Nucleic Acids Research*.
541 <https://doi.org/10.1093/nar/gkn660>.
- 542 Pavesi, G., Mereghetti, P., Mauri, G., & Pesole, G. (2004). Weeder Web: Discovery of
543 Transcription Tactor Binding Sites in a Set of Sequences from Co-regulated Genes.
544 *Nucleic Acids Research*, 32(suppl_2), W199–W203.
545 <https://doi.org/10.1093/nar/gkh465>
- 546 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,
547 M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A.,
548 Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Louppe, G. (2012). Scikit-
549 learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12.

- 550 Quang, D., & Xie, X. (2016). DanQ: a Hybrid Convolutional and Recurrent Deep Neural
551 Network for Quantifying the Function of DNA Dequences. *Nucleic Acids Research*,
552 44 (11), e107. <https://doi.org/10.1093/nar/gkw226>
- 553 Quinlan, J. R. (1993). Constructing Decision Trees. *C4.5*. [https://doi.org/10.1016/b978-
554 0-08-050058-4.50007-3](https://doi.org/10.1016/b978-0-08-050058-4.50007-3)
- 555 Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor
556 Package for Differential Expression Analysis of Digital Gene Expression Data.
557 *Bioinformatics*, 26. <https://doi.org/10.1093/bioinformatics/btp616>
- 558 Roy, S., Wapinski, I., Pfiffner, J., French, C., Socha, A., Konieczka, J., Habib, N., Kellis,
559 Thompson, M., & Regev, A. (2013). Arboretum: Reconstruction and Analysis of the
560 Evolutionary History of Condition-Specific Transcriptional Modules. *Genome
561 Research*, 23 (6): 1039–50.
- 562 Shi, J., Yang, W., Chen, M., Du, Y., Zhang, J., & Wang, K. (2011). AMD, an Automated
563 Motif Discovery Tool Using Stepwise Refinement of Gapped Consensuses. *PLoS
564 One*, 6 (9): e24576.
- 565 Shu, L., Katholi, C.R., Higazi, T., & Unnasch, T.R. (2003). Analysis of the *Brugia Malayi*
566 HSP70 Promoter Using a Homologous Transient Transfection System. *Molecular
567 and Biochemical Parasitology*, 128 (1): 67–75.
- 568 Smith, A. D., Sumazin, P., & Zhang, M. Q. (2005). Identifying tissue-selective
569 transcription factor binding sites in vertebrate promoters. *Proceedings of the
570 National Academy of Sciences of the United States of America*.
571 <https://doi.org/10.1073/pnas.0406123102>

- 572 Thijs, G., Marchal, K., Lescot, M., Rombauts, S., Moor, B. D., Rouze, P., & Moreau, Y.
573 (2002). A Gibbs Sampling Method to Detect Overrepresented Motifs in the
574 Upstream Regions of Co-expressed Genes. *Journal of Computational Biology*, 9.
575 <https://doi.org/10.1089/10665270252935566>
- 576 Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., & Pachter, L.
577 (2013). Differential Analysis of Gene Regulation at Transcript Resolution with RNA-
578 Seq. *Nature Biotechnology*, 31 (1): 46–53.
- 579 Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J.,
580 Salzberg, S.L., Wold, B.J., & Pachter, L. (2010). Transcript Assembly and
581 Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform
582 Switching during Cell Differentiation. *Nature Biotechnology*, 28 (5): 511–15.
- 583 Varoquaux, G., Buitinck, L., Louppe, G., Grisel, L., Pedregosa, F., & Mueller, A. (2015).
584 Scikit-Learn. GetMobile: Mobile Computing and Communications.
585 <https://doi.org/10.1145/2786984.2786995>.
- 586 Warbrick, E.V., Barker, G.C., Rees, H.H., Howells, R.E. (1993). The Effect of
587 Invertebrate Hormones and Potential Hormone Inhibitors on the Third Larval Molt
588 of the Filarial Nematode, *Dirofilaria immitis*, in vitro. *Parasitology*, 107: 459–463.
- 589 Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P.,
590 Najafabadi, H.S., et al. (2014). Determination and Inference of Eukaryotic
591 Transcription Factor Sequence Specificity. *Cell*, 158 (6): 1431–43.

- 592 Xu, S., Lui, C., Tzertzinis, G., Ghedin E., Evans, C., Kaplan, R., Unnasch T. (2011). *In*
593 *Vivo* Transfection of Developmentally Competent *Brugia malayi* Infective Larvae.
594 *International Journal for Parasitology*, 41(3-4): 355-362.
- 595 Zang X., Maizels R.M. (2001). Serine Proteinase Inhibitors from Nematodes and the
596 Arms Race between Host and Pathogen. *Trends Biochemical Science*, 26: 191-
597 197.
- 598