1
2 # T-box transcription factor eomesodermin/Tbr2 in Atlantic cod (*Gadus morhua*
3 # L.): Molecular characterization, promoter structure and function analysis
4
5 Heng Chi[1,2,3*], Kristian Gillebo Sørmo[1], Jing Diao[2], Roy Ambli Dalmo[1*]
6
7 1. Norwegian College of Fishery Science, Faculty of Biosciences, Fisheries and Economics, UiT
8 the Arctic University of Norway, N-9037 Tromsø, Norway
9 2. Shandong Key Laboratory of Disease Control in Mariculture, Marine Biology Institute of
10 Shandong Province, 266104 Qingdao, China
11 3. Key Laboratory of Experimental Marine Biology, Chinese Academy of Sciences, Institute of
12 Oceanology, 266071 Qingdao, China
13
14
15 *Correspondence: Roy Ambli Dalmo, E-mail: roy.dalmo@uit.no, Phone: +47 77644482
16 Heng Chi, E-mail: chiheng@qdio.ac.cn
17
18
19 **Keywords:** Eomesodermin; Atlantic cod; Molecular cloning; Promoter structure; Function
20 analysis
21

22 **Abstract:** Eomesodermin (Eomes) is a member of T-box transcription factor family and plays an
23 important role in the regulation of a wide variety of developmental processes and immune
24 response in animals. Here we report cloning and characterization of the full-length cDNA of
25 Atlantic cod Eomes (GmEomes), which possesses a TBOX_3 domain similar to its counterpart in
26 mammals. The regulated expression was observed in head kidney and spleen in response to live
27 *Vibrio anguillarum* infection *in vivo*, and spleen leukocytes *in vitro* after PMA and poly I:C
28 stimulation. Furthermore, we determined a 694 bp sequence, upstream of the transcriptional start
29 site (TSS), to contain a number of sequence motifs that matched known transcription
30 factor-binding sites. Activities of the presumptive regulatory gene were assessed by transfecting
31 different 5'-deletion constructs in CHSE-214 cells. The results showed that the basal promoters
32 and positive transcriptional regulator activities of GmEomes were dependent by sequences located
33 from -694 to -376 bp upstream of TSS. Furthermore, we found that some Eomes binding sites
34 were present in the 5'-flanking regions of the cod IFNγ gene predicted by bioinformatics.
35 However, Co-transfection of eomesodermin overexpression plasmids with INFγ reporter vector
36 into CHSE-214 cells determined that Atlantic cod eomesodermin played a minor role in INFγ.
37

38 **1. Introduction**

39 T-box genes, a highly conserved family of transcription factors having a common DNA-binding
40 sequence (the T-box), are important in the regulation of a wide variety of developmental processes
41 and immune response in animals [1-4]. Eomesodermin (Eomes) and T-BRAIN1 together with
42 T-box expressed in T-cells (T-bet/ TBX21) all belong to TBR1 (T-box brain protein 1) sub-family
43 of T-box genes in mammals [5-8]. Eomes has been shown to be involved in mesoderm formation
44 in most vertebrates and in trophoblast differentiation in mammals [6, 8-11]. Important roles of
45 Eomes during early development have also been reported in fish [12-15]. In the mammalian
46 immune system, Eomes is highly expressed in $CD8^+$ T cells and functions redundantly with T-bet
47 in the induction of IFN-γ production to prevent $CD8^+$ T cells from differentiating into other T cell
48 subtypes [16-18].

49 Atlantic cod (*Gadus morhua* L.) is an economically and ecologically important species in the
50 northern Atlantic [19]. Recently, genomic analysis of Atlantic cod reveals an immune system that
51 differs significantly from that in other vertebrates and most fish species. The major
52 histocompatibility complex (MHC) II，CD4 and invariant chain (Ii) have been lost during
53 evolution. However, there is an expanded number of MHCI genes and a unique composition of the
54 toll-like receptor family. These compensatory changes in both adaptive and innate immunity
55 suggest that cod may not be more susceptible to infectious disease than most other vertebrates [20],
56 indicating a high relevance to study the importance of innate immunity and mechanisms leading to
57 T-cell dependent cytotoxic responses. A comprehensive understanding of the molecular pathways
58 involved in physiological and immunological responses of Atlantic cod may help to overcome the
59 challenges in health management (e.g. viral diseases) [20,21]. While MHCI, present in all
60 nucleated cells, has long been thought to be exclusively involved in the presentation of
61 endogenous antigens, this classical view has been progressively replaced by a more complete
62 understanding of the cell biology of antigen-presenting cells. Indeed, exogenous antigens can also
63 be processed by the proteasome and loaded on MHCI molecules by an alternative pathway called
64 cross-presentation, which ultimately activates $CD8^+$ T-cells [22]. The presence of such a high
65 number of MHCI loci in cod has therefore led to the hypothesis that different subsets of $CD8^+$ T
66 cells have been generated to compensate the absence of $CD4^+$ T-cells [23].

67 Eomes is an important transcription factor during the immune response of $CD8^+$ T-cells, and has
68 been identified in some teleost species such as Atlantic salmon, rainbow trout, ginbuna crucian
69 carp and zebrafish [24-26], while studies of this transcription factor in Atlantic cod still are
70 lacking. The main objective of this study was to characterise the expression pattern of Eomes in
71 healthy and *V. anguillarum* infected cod, the promoter structure analysis and ability of GmEomes
72 to activate the expression of IFN-γ.
73

74 **2. Materials and methods**

75 *2.1. Cloning and sequencing of GmEomes cDNA*

76 A partial cod EST sequence similar to vertebrate Eomes was identified based on nucleotide and
77 amino acid sequence homology to annotated zebrafish and Atlantic salmon T-bet and Eomes
78 sequences deposited in GenBank using the BLAST software
79 (http://www.ncbi.nlm.nih.gov/BLAST). Two Atlantic cod Eomes ESTs (GenBank accession no:
80 ES786771.1 and ES786391.1) were retrieved. Internal primers were designed from the cod EST
81 sequences and PCR products were obtained from the cDNA library of the stimulated spleen tissue
82 and sequenced. Total RNA (1 μg) isolated from Atlantic cod spleen (~30 mg), using TRIZOL®
83 Reagent (Invitrogen), was used as a template and reverse transcribed to cDNA for RACE (rapid
84 amplification of cDNA ends) with a SMART RACE cDNA Amplification kit (Clontech) according
85 to the manufacturer's instruction and as described previously [27]. The cDNA sequence and
86 deduced amino acid sequence of Atlantic cod Eomes sequences were further analyzed using
87 BLAST and the ExPASy Molecular Biology server (http://us.expasy.org) and Pfamp [28]. Amino
88 acid identity and similarity analysis were done with the Matrix Global Alignment Tool (MatGAT)
89 program v 2.0 using default parameters [29].
90

91 *2.2. Phylogenetic analysis*

92 Multiple sequence alignment was created using CLUSTALW, while MEGA version 4.1 was used

93 to assess the similarities among the aligned sequences (www.ebi.ac.uk/clustalw/) [30]. A

94 phylogenetic tree, based on the deduced amino acid sequences, was constructed using the

95 neighbor-joining (NJ) algorithm, and the reliability of the branching was tested using bootstrap

96 re-samplings with 1 000 pseudo-replicates.

97

98 *2.3. Isolation of 5'-flanking region of the GmEomes gene by genome walking*

99 The 5 ˊ-flanking region of the GmEomes gene was isolated using the Universal GenomeWalker Kit

100 (Clontech). Four GenomeWalker libraries were constructed according to the manufacturer's

101 instruction. For each genome walker experiment, two adjacent reverse primers (GmEomes gwrv1

102 and GmEomes gwrv2) were designed near 5'-UTR region of the target gene (Table 1), and used

103 in two PCRs in combination with the forward adaptor primers AP1 and AP2 (Clontech) for each

104 library. The resulting PCR products from four different libraries were cloned in TOPO vector

105 (Invitrogen), sequenced and analyzed as described above.

106 In order to verify this new sequence, a forward primer (GmEomesEcoRIfw1) was designed within

107 this new sequence and used with a reverse primer (GmEomesSacIIrv) designed within the

108 transcribed region of the Eomes gene (Table 1). PCR from the Atlantic cod genomic DNA was

109 performed, and the products obtained were cloned and sequenced. Identification of transcription

110 factor binding motifs was predicted with TRANSFAC® (Biobase International) and MatInspector

111 version 6.2 [31,32].

112

113 *2.4. Construction of GmEomes reporter gene plasmids*

114 Deletion constructs with successive removal of the 5'-region were generated by PCR using the
115 forward primers GmEomesEcoRIfw1, GmEomesEcoRIfw2, and GmEomesEcoRIfw3 having
116 recognition sequences for restriction endonuclease EcoRI, while the reverse primer
117 GmEomesSacIIrv (Table 1) had a SacII restriction site to generate the constructs p(-694/+23)Luc,
118 p(-376/+23)Luc, and p(−216/+23)Luc respectively. The promoterless pMet Luciferase Reporter
119 (Clontech) was used as reporter plasmids for cloning.

120 Both the PCR products of different 5'-deletion constructs and the basic reporter vectors were
121 digested with their respective restriction enzymes (New England) and ligated (T4 DNA ligase) to
122 generate the above-mentioned constructs for each basic reporter vector (pMet Luciferase) in
123 parallel. All plasmid DNA constructs were isolated using Endo-free Plasmid Mini Kit (Qiagen) to
124 have high quality plasmid for transfection. All plasmid constructs were verified by restriction map
125 analysis and DNA sequencing.

126

127 *2.5. Cell culture, transfection and reporter activity assay*

128 Chinook salmon embryonic cells (CHSE-214) were seeded in a flask (Nunc) containing L-15
129 medium (Invitrogen), penicillin (60 μg ml^{-1}), streptomycin (100 μg ml^{-1}), 1% non-essential amino
130 acid (NEAA, Gibco) and 8% fetal calf serum (FCS) at 20°C in an incubator for one week. Cells
131 were washed twice in 10 ml phosphate buffered saline (PBS), and then 1.5 ml trypsin (1.25%) was
132 added. Cells were re-suspended in L-15 medium (8% FCS, 1% NEAA, without antibiotics) and
133 the cell number was adjusted to give a proper density. Cells were washed twice with PBS and
134 re-suspended in buffer R (Invitrogen) and divided into tubes (10μl per tube). For each tube, $2x10^{5}$

135 cells and plasmids {300 ng luciferase vector and 50 ng pSEAP2 control vector (Clontech)} were
136 mixed and transfection was performed using electroporation by the Neon Transfection System
137 (Invitrogen): 10 μl volume using D1 program (voltage 1100, pulse width 30, pulse no. 2) in line
138 with the protocol supplied by the manufacturer. The pSEAP2-control vector for normalizing
139 transfection efficiency was included in all assays to standardize protein expression levels.
140 The transfected CHSE-214 cells were seeded in 24 well plates (Nunc) at a density of $2x10^5$ cells
141 $well^{-1}$. 12 h after transfection, the medium was removed and replaced by fresh medium. The
142 analysis for Metridia luciferase activity using Ready-To-Glow™ Secreted Luciferase Reporter
143 System (Clontech) and SEAP activity using Great EscAPe™ SEAP Chemiluminescence Detection
144 Kit (Clontech), and then the luciferase and SEAP activity were assayed using a plate Luminometer,
145 Luminoskan Ascent (Thermo). The assay was performed thrice.
146
147 *2.6. Tissue specific expression of GmEomes in healthy cod*
148 Atlantic cod (~50 g) were supplied from the Aquaculture Research Station (Tromsø, Norway). Six
149 fish were killed by immersion the fish in 100 mg L^{-1} Metacaine, and then, gill, head kidney, liver,
150 heart, gut, pyloric caeca, skin and muscle were sampled and quickly immersed in RNAlater
151 (Invitrogen). Total RNA was isolated by using RNeasy Mini Kit (Qiagen). RNase-Free DNase Set
152 (Qiagen) was additionally used to remove genomic DNA contamination and QuantiTect Reverse
153 Transcription Kit (Qiagen) was applied for cDNA synthesis. The synthesized cDNA was diluted
154 10-fold with MilliQ water and 1 μl of this dilution was used as template in a 20 μl reaction volume.
155 Fast SYBR® Green Master Mix (Applied Biosystems) was used as reagents. qPCR was carried
156 out in ABI PRISM 7500 Fast Real-Time PCR System (Applied Biosystems). Primers for qPCR

157 have been listed in Table 1. An evaluation of gene expression of the two endogenous controls,

158 β-actin and eF1α, No significant differences in gene expression ($P > 0.05$) were detected between

159 the two endogenous controls, and eF1αwas used as endogenous control in this study [33,34]. The

160 results were expressed as fold change in comparison with the lowest expression level. All

161 experiments on fish were approved by the Norwegian Food Safety Authority, and experiments

162 were in accordance the animal welfare act.

163

164 *2.7. Expression of GmEomes in spleen leucocytes after Poly I:C and PMA stimulation*

165 Isolation of spleen primary cells was performed in line with the previously published protocol [27].

166 Briefly, the spleen was removed and minced through a 100 μm nylon Falcon cell strainer (BD

167 Bioscience) in L-15 culture medium (Invitrogen) supplemented with heparin (20 U ml^{-1}). The cell

168 suspension was loaded on a discontinuous 25/50% Percoll (Amersham Pharmacia Biotec) gradient,

169 and centrifuged at 850 *g* for 40 min at 4°C. The cells at the interface were collected and washed

170 twice with L-15 medium. Cells were seeded at a density of 5×10^{6} cells per well in 24-well cell

171 culture plates in L-15 with 1% FCS. The cells were divided into 3 groups with 6 replications for

172 each group and incubated with polyinosinic-polycytidylic acid (poly I:C; Sigma) (100 ng ml^{-1}),

173 phorbol 12-myristate 13-acetate (PMA; Sigma) (100 ng ml^{-1}) or left unstimulated, respectively.

174 After incubation for 0, 6, 12, 24 and 48 h, cells were harvested for qPCR analysis, and the results

175 were expressed as fold change in comparison with the unstimulated group at 0 h. The methods of

176 RNA isolation and cDNA synthesis have been described in 2.6.

177

178 *2.8. Expression of GmEomes in head kidney and spleen after infection by V. anguillarum*

179 *V. anguillarum* (serotype O2b; isolate 4299) [35] were inoculated on Tryptic Soya Agar (Oxoid)
180 supplemented with 5% human blood and 1.5% NaCl and incubated for 3 days and grown in
181 Marine Broth (MB-2216, Difco) at 12°C with gentle shaking until optical density ($OD_{600\ nm}$) was
182 = 0.5-0.6. The culture was washed, centrifuged and diluted with 0.9% NaCl (saline) giving OD_{600}
183 $_{nm}$ = 0.2 corresponding to approximately 10^{8} bacteria ml^{-1}. Infection doses were adjusted by
184 diluting the bacteria suspension to ~10^{7} bacteria ml^{-1}. Colony forming units (CFU) of the
185 infection doses were determined using plates which were incubated at 22° C for five days. 18 fish
186 (~50 g) were intraperitoneally (i.p) injected with 0.1 ml bacterial suspension after being
187 anesthetized, whereas the time-control fish received 0.1 ml of saline. Fish in these two treatment
188 groups were kept apart in two tanks (500L) with continuous supply of sea water (10 ° C).
189 Additional six untreated fish were sampled as time zero control fish at the start of the experiment.
190 The head kidney and spleen from six fish of each group were sampled at 1, 2 and 4 days after
191 injection and immersed in RNAlater (Thermo) for qPCR analysis, and the results were expressed
192 as fold change in comparison with the control group at 0 day. The methods of RNA isolation and
193 cDNA synthesis have been described in 2.6. The experiment with *V. anguillarum* infection was
194 approved by the Norwegian Food Safety Authority.
195
196 *2.9. Overexpression plasmid construction of pGmEomes-RFP and promoter-reporter plasmid*
197 *construction of pLuc2-IFNγ*
198 To construct pGmEomes-RFP which express GmEomes fused to red fluorescent protein (RFP), the
199 coding sequences of GmEomes were amplified with primers GmEomesORFfw / GmEomesORFrv
200 (Table 1), and the PCR products were inserted into pTagRFP-N (Evrogen). The construction of

201 pSsT-bet-RFP plasmid which expresses salmon T-bet fused to RFP has been described previously

202 [36]. Genomic DNA was isolated from Atlantic cod spleen with the DNA isolation kit (Qiagen).

203 About 588 bp of the 5' flanking region sequences of the cod IFNγ (GenBank No: FJ356236.1)

204 were obtained from the genomic DNA by PCR using the primers GmIFNrpromfw/GmIFNrpromrv

205 (Table 1), and the PCR products were inserted into pMetLuc-2 vector (Clontech). All plasmid

206 DNA constructs were isolated using EndoFree Plasmid Kit (Qiagen) to have high quality plasmid

207 for transfection. All plasmid constructs were verified by restriction map analysis and DNA

208 sequencing. The cell culture, transfection and reporter activity assay have been described above in

209 2.5. The reporter assay experiment was done in triplicate fashion.

210 At 48 h after transfection, the medium in the wells was used for Metridia luciferase and SEAP

211 analysis. The cells transfected by pGmEomes-RFP and pTagRFP-N expressing plasmids were

212 fixed with 4% formaldehyde (w/v) (Thermo) for a half hour, then DAPI (Invitrogen) was used for

213 nucleic acid (nucleus) staining in line with the protocol supplied by the manufacturer. Micrographs

214 were obtained by inverted fluorescence microscope (Zeiss).

215 *2.11. Statistical analysis*

216 Statistical analyses were performed using one-way ANOVA followed by LSD multiple group

217 comparisons in the SPSS 18.0 software package (SPSS Inc., Chicago, IL, USA). Data are

218 presented as means ± SD, and statistical significance was defined as $P < 0.05$.

219

220 **3 Results**

221 *3.1. Characterization of the nucleotide and amino acid sequences of Eomes*

222 The GmEomes cDNA (GenBank accession no. JF968408) consisted of 2618 bp in length with an

223 open reading frame of 1989 bp encoding a polypeptide of 662 amino acids, a 278 bp 5'

224 un-translated region (UTR) and a 2531 bp 3' UTR. Within the 3' UTR, one polyadenylation
225 signals (AATAAA) and four mRNA instability motifs (ATTTA) were found (Fig. 1). The putative
226 protein had a molecular weight of 72321.9 Da and an estimated pI of 6.15.
227 According to a BLAST search, the cloned GmEomes gene possessed a high degree of sequence
228 similarity (64.8–89.9%) (Table 2) to various members of the Eomes family in different animals,
229 indicating that it has been conserved during evolution (Fig. 2). Alignment of the predicted protein
230 of GmEomes revealed the presence of conserved TBOX_3 domain (219-414). A phylogenetic tree
231 was developed based on multiple alignments of Eomes from various species, including fish, frog,
232 chicken and mammals (Fig. 3). GmEomes was more closely related to Atlantic salmon Eomes
233 than to the other species.
234
235 *3.2. Structure of 5'-flanking region of GmEomes gene*
236 As the first step towards understanding the transcriptional regulation of the GmEomes gene, a
237 sequence of 694 bp lying 5' to the transcription start site (TSS) was determined (GenBank
238 accession no. JF968407). Sequence analysis using the transcription factor binding site prediction
239 program MatInspector and TRANSFAC® revealed several notable features. The TATA-box,
240 considered being the core promoter sequence, was present at position -309 relative to the putative
241 GmEomes transcription start site (Fig. 4). In addition, we found two potential GATA consensus
242 sequences present within this region at -211, and -269 which are known to be enriched in specific
243 gene regulatory regions of immune cells. Oct-1, HSF2, MyoD, CdxA and SRY, known to be
244 important for transcriptional regulation of genes were also found in this region. Lastly, the NF-E2
245 was found close to the TSS. Other putative transcription factor binding sites are shown in Fig. 4.

3.3. Activity of the GmEomes promoter

To precisely define the 5'-end of TSS of GmEomes that was required for the induced activation, progressive deletion constructs of the GmEomes promoter region were generated and transiently transfected into CHSE-214 cells (Fig. 4). All promoter constructs were active at 12 h, and induced increasing luciferase activities from 24 h to 48 h ($P < 0.001$). The minimal promoter construct induced low luciferase activity at all time points and similar time kinetics as the p(-376/+23)Luc ($P > 0.05$), but showed significant difference compared to p(-694/+23)Luc ($P < 0.001$). The full-length p(-694/+23)Luc promoter construct induced relative luciferase activity 40.16-fold higher than the promoterless controls at 48h, whereas the p(-376/+23)Luc construct and minimal Eomes promoter induced only 2.97-fold and 2.65-fold, compared to promoterless controls, respectively (Fig. 4).

3.4. Tissue distribution of Eomes mRNA

As shown in Fig. 5, the GmEomes gene was widely expressed in all the sampled tissues of healthy fish. The largest quantity of GmEomes mRNA was found in spleen, followed by gill, head kidney, liver, heart, gut, pyloric caeca and skin. The level of GmEomes transcripts in muscle was relatively low.

3.5. Expression profile of GmEomes in head kidney and spleen in vivo after V. anguillarum infection and in spleen leukocytes in vitro after PMA and poly I:C stimulation

No mortality and abnormal behavior were observed during the experimental challenge with *V.*

268 *anguillarum*. In spleen, the expression of GmEomes went 3.85-fold higher than in the control
269 group at day 1 ($P < 0.01$), whereas it was decreased at day 4 post infection (Fig. 6A). No
270 statistically significant changes of GmIFNγ mRNA levels were observed in the saline injected fish
271 in the control group or during challenge with *V. anguillarum*. In head kidney, both GmEomes and
272 GmIFNγ expressions were significantly upregulated at 4 d post-infection, with 5.36 and 3.56-fold
273 increase, respectively ($P < 0.01$) (Fig. 6B). No statistically significant differences in the expression
274 of GmEomes and GmIFNγ between fish in the control group at the other time points was found.
275 The spleen leukocytes stimulated with PMA showed significantly increased expression of
276 GmEomes at 6, 12, 24 and 48 h ($P < 0.01$), compared to non-stimulated cells. Similarly, the
277 expression of GmIFNγ was also significantly upregulated at 12, 24 and 48 h after PMA
278 stimulation. The GmIFNγ expression was significantly higher at 48 h post poly I:C stimulation
279 compared to control cells ($P < 0.01$), whereas no statistically significant changes of GmEomes
280 mRNA levels was observed in control cells and during poly I:C stimulation ($P > 0.05$) (Fig. 6C).

281 3.6 *Localization of over expressed GmEomes in CHSE-214 cells*

282 To obtain information from overexpression of GmEomes, microscopic analysis of cells transfected
283 with GmEomes encoding plasmid fused to red fluorescent protein (RFP) were carried out. The
284 complete coding sequence of GmEomes was inserted into the plasmids encoding red fluorescent
285 protein fused to the 3'-terminal ends. Approximately 15–20% of RFP positive CHSE-214 cells
286 were observed after transfection with the plasmids, as shown in Fig. 7A. GmEomes fused with
287 RFP were identified in or close to the nuclei at 48 h post transfection, whereas the cells transfected
288 with empty control vector (pTagRFP-N) but still containing a RFP gene, showed RFP widespread
289 in the cytoplasms.

290

291 *3.7 The effect from over expression of GmEomes on Atlantic cod IFNγ promoter*

292 Approximately 588 bp 5' flanking region of GmIFNγ were cloned, in which Eomes binding sites
293 were found (Fig. 7B). Overexpression of GmEomes in CHSE-214 cells slightly enhanced the
294 activity of GmIFNγ promoter region, but not with statistically significant difference compared to
295 controls ($P = 0.083$), while an overexpression of salmon T-bet significantly enhanced the activity
296 of the IFNγ promoter ($P < 0.01$) (Fig. 7C).

297

298 **4 Discussion**

299 In this study, we identified an Eomes homologue, GmEomes, from Atlantic cod and analysed its
300 structure, expression, and regulatory property. We found that GmEomes shared 64.8 and 89.9%
301 overall amino acid sequence identities with human and zebrafish Eomes, respectively. This
302 observation indicated that Eomes is considerably conserved among lower and higher vertebrates,
303 which may be consistent with the fundamental role of Eomes. GmEomes contains a T-box domain
304 which was defined as the minimal region within the T-box family proteins that is both necessary
305 and sufficient for sequence-specific DNA binding. The presence of this domain in GmEomes
306 suggested a potential capacity of the translated protein to bind DNA as a transcription factor.
307 Sequence analysis of 5' flanking region of GmEomes revealed that the presence of TATA box in
308 the proximal promoter (694 bp to the TSS) are in line with the previous study on the Xenopus and
309 salmon Eomes promoters [9,24]. By PCR-aided deletion of the putative promoter region we
310 detected a strong positive regulatory element between positions -694 and-377 bp in the GmEomes
311 promoter. Further analysis of this element indicated that it shared a number of consensus
312 transcription factor binding sites, such as binding sites for MyoD, Oct-1, HSF and SRY. Recent

313 reports have shown that MyoD is a transcription factor involved in regulating muscle
314 differentiation, whereas both Oct-1 and Cdx are involved in regulation of the genes essential for
315 growth and embryonic development [37,38]. The sex-determining region Y (SRY) is responsible
316 for the initiation of male sex determination in humans. HSF is needed for proper animal
317 development and the survival of cancer cells [39,40]. The presence of MyoD, Oct-1, Cdx and SRY
318 transcription factor binding sites in the promoter of GmEomes indicated similar mechanisms of
319 development and cell differentiation in cod as in mammals. By PCR-aided deletion of the putative
320 promoter region we detected a weak positive regulatory element between positions -376 and +23
321 bp in the GmEomes promoter. Further analysis of this element indicated that it shared a number of
322 consensus transcription factor binding sites, including GATA and Cdx for positive control
323 transcription factor binding sites, and S8 and HF-E2 negative control transcription factor binding
324 sites. This suggests that GATA and Cdx are not the main transcription factors for the expression of
325 GmEomes.

326 In mammals, Eomes is expressed highly at the certain stages during development [41] and has
327 been well defined as key drivers of neuronic, cardiac and immune cell development associated
328 with effector function [41-43]. In the immune system, Eomes expression has been reported to be
329 present in different lymphoid tissues, and in cells and their subsets, including $CD8^+$ T-cells, γδ
330 T-cells, invariant NKT cells, natural killer cells, B-cells, and dendritic cells [44]. In this study, we
331 found that the expression of GmEomes occurred in multiple tissues and was highly expressed in
332 spleen, head kidney and gill. This is not in full consistence with the reports on Atlantic salmon,
333 trout and zebrafish [24,26]. The expression of Eomes mRNA in gills is rather at low levels when
334 compared to head kidney in trout and ginbuna carp [25,26]. However, the levels of GmEomes

335 transcripts were similar in gill and head kidney of Atlantic cod. In salmonid fishes and ginbuna
336 carp, constitutive high expression of GATA-3 and IL-4 in gill suggests that this tissue may be
337 capable to form a Th2-skewed immune environment [45,46]. A presence of high level of Eomes in
338 the gills of cod may also lead to differentiation of naïve T-cells into $CD8^+$ T cells to compensate
339 the absence of of $CD4^+$T cells and subsets.
340 *V. anguillarum* is an abundant pathogen in aquaculture, and causes vibriosis in a range of different
341 fish species [47-51]. In our study, the expression of GmEomes was enhanced in spleen (day 2) and
342 in the head kidney (day 4) after experimental infection with the *V. anguillarum*. These results
343 suggest a role for GmEomes in host immune defence against microbial pathogens. PMA, through
344 its activation of protein kinase C, is often used experimentally to induce leucocyte activation -
345 especially in T-cells [52,53]. In this study, the expression of GmEomes was indeed induced in
346 spleen leucocytes following PMA stimulation (24 h and 48 h).
347 In mammals, Eomes is highly expressed in $CD8^+$ T cells, but not in $CD4^+$ T cells. Even though
348 Eomes/Tbr2 plays a role in the ability of $CD8^+$ T cells to penetrate tumors, it only plays a small
349 role in production of IFNγ [16,17,54]. In our study, the transcripts levels of GmIFNγ were
350 increased with the increase of GmEomes expression at some time points after PMA stimulation *in*
351 *vitro* or during *V. anguillarum* infection *in vivo*. In addition, we found that some Eomes binding
352 sites were present in the 5'-flanking regions of the cod IFNγ gene predicted by bioinformatics. In
353 CHSE-214 cells transfected with pGmEomes-RFP, the fusion protein was detected mainly in the
354 nucleus, which is consistent with the known function of Eomes in regulation of gene expression.
355 However, GmEomes only slightly increased the activity of this IFNγ promoter suggesting that
356 GmEomes does not significantly contribute to IFNγ production, while the GmT-bet significantly

357 enhanced the activity of GmIFNγ promoter. These results were in line with some reports from
358 mice, where similar numbers of IFNγ producing cells were observed in WT and Eomes KO
359 effector $CD8^{+}$T cells. T-bet KO $CD8^{+}$ T cells and T-bet/Eomes double KO $CD8^{+}$ T cells showed
360 only a slight reduction (approx. 10%) of IFN-γ^{+} cells [54]. As such, Eomes is required, but not
361 crucial, for IFN-γ production by T cells. Mechanistic studies on the interplay between Eomes and
362 T-bet should be performed to find out their relative importance to induce IFNγ production in fish
363 lymphocytes.
364 In summary, we identified and characterized Gmeomes which is a T-box transcription factor
365 molecule and its expression pattern during homeostasis and during contains of PAMPs and
366 pathogen challenge in A. cod. This work presents new knowledge about the promoter region and
367 the promoter activity of the GmEomes. Furthermore, we found that GmEomes mainly was
368 localized in the nucleus, but played a minor role in activation of the INFγ promoter.
369

370 **Acknowledgement**

371 This project was supported by National Natural Science Foundation of China (grants 31872594 &
372 31402326), the Research Council of Norway (grants 237315 & 239140) and the Open Fund of
373 Shandong Key Laboratory of Disease Control in Mariculture (grant KF201801), Tromsø Research
374 Foundation and The Key Laboratory of Mariculture（KLM), Ministry of Education, OUC.
375

376 **References**

377 [1] J. Smith, T-box genes: what they do and how they do it, Trends Genet. 15 (1999) 154-158.
378 [2] V.E. Papaioannou, L.M. Silver, The T-box gene family, BioEssays 20 (1998) 9-19.
379 [3] A. Kispert, B.G. Herrmann, The Brachyury gene encodes a novel DNA binding protein,

380 EMBO. J. 12 (1993) 3211-3220.

381 [4] C. Showell, O. Binder, F.L. Conlon, T-box genes in early embryogenesis, Dev. Dyn. 229
382 (2004) 201–218.

383 [5] B.G. Ciruna, J. Rossant, Expression of the T-box gene eomesodermin during early mouse
384 development, Mech. Dev. 81 (1999) 199-203.

385 [6] A.P. Russ, S. Wattler, W.H. Colledge, S.A. Aparicio, M.B. Carlton, J.J. Pearce, et al.,
386 Eomesodermin is required for mouse trophoblast development and mesoderm formation,
387 Nature 404(2000) 95-99.

388 [7] K. Tagawa, T. Humphreys, N. Satoh, T-Brain expression in the apical organ of hemichordate
389 tornaria larvae suggests its evolutionary link to the vertebrate forebrain, J. Exp. Zool. 2000,
390 288:23-31.

391 [8] S.J. Szabo, S.T. Kim, G.L. Costa, X. Zhang, C.G. Fathman, L.H. Glimcher, A novel
392 transcription factor, T-bet, directs Th1 lineage commitment, Cell 100 (2000) 655-669.

393 [9] K. Ryan, N. Garrett, P. Bourillot, F. Stennard, J.B. Gurdon, The Xenopus eomesodermin
394 promoter and its concentration-dependent response to activing, Mech. Dev. 94 (2000) 133–
395 146.

396 [10] A. Bulfone, S. Martinez, V. Marigo, M. Campanella, A. Basile, N. Quaderi, et al., Expression
397 pattern of the Tbr2 (Eomesodermin) gene during mouse and chick brain development, Mech.
398 Dev. 84 (1999) 133–138.

399 [11] K. Sone, T.C. Takahashi, Y. Takabatake, K. Takeshima, T. Takabatake, Expression of five
400 novel T-box genes and brachyury during embryogenesis, and in developing and regenerating
401 limbs and tails of newts, Develop. Growth Differ. 41 (1999) 321–333.

402 [12] M. Mione, S. Shanmugalingama, D. Kimelmanb, K. Griffin, Overlapping expression of
403 zebrafish T-brain-1 and eomesodermin during forebrain development, Mech. Dev. 100 (2001)
404 93–97.
405 [13] A.E. Bruce, C. Howley, Y. Zhou, S.L. Vickers, L.M. Silver, M.L. King, et al., The maternally
406 expressed zebrafish T-box gene eomesodermin regulates organizer formation, Development
407 130 (2003) 5503–5517.
408 [14] A.E. Bruce, C. Howley, D. Dixon Fox, R.K. Ho, T-box gene eomesodermin and the
409 homeobox-containing Mix/Bix gene mtx2 regulate epiboly movements in the zebrafish, Dev.
410 Dyn. 233 (2005) 105–114.
411 [15] C.R.R. Bjornson, K.J.P. Griffin, G.H. Farr III, A. Terashima, C. Himeda, et al.,
412 Eomesodermin is a localized maternal determinant required for endoderm induction in
413 zebrafish, Dev. Cell 9 (2005) 523–533.
414 [16] E.L. Pearce, A.C. Mullen, G.A. Martins, C.M. Krawczyk, A.S. Hutchins, V.P. Zediak, et al.,
415 Control of effector CD8+ T cell function by the transcription factor Eomesodermin. Science
416 302 (2003) 1041–1043.
417 [17] C.N.R. Lino, J. Barros-Martins, L. Oberdörfer, T. Walzer, I. Prinz, Eomes expression reports
418 the progressive differentiation of IFN-γ-producing Th1-like γδ T cells, Eur. J. Immunol. 47
419 (2017) 970–981.
420 [18] E. Lupar, M. Brack, L. Garnier, S. Laffont, K.S. Rauch, K. Schachtrup, et al., Eomesodermin
421 Expression in CD4+ T Cells Restricts Peripheral Foxp3 Induction, J. Immunol. 195 (2015)
422 4742-4752.
423 [19] G. Rosenlund, O. Halldorsson, Cod juvenile production: research and commercial

424 developments, Aquaculture 268 (2007) 188-194.

425 [20] B. Star, A.J. Nederbragt, S. Jentoft, U. Grimholt, M. Malmstrøm, T.F. Gregers, et al., The
426 genome sequence of Atlantic cod reveals a unique immune system, Nature 477 (2011)
427 207-210.

428 [21] O.K. Tørresen, B. Star, S. Jentoft, W.B. Reinar, H. Grove, J.R. Miller, et al., An improved
429 genome assembly uncovers prolific tandem repeats in Atlantic cod, BMC Genomics, 18
430 (2017), 95.

431 [22] S. Amigorena, A. Savina, Intracellular mechanisms of antigen cross presentation in dendritic
432 cells, Curr. Opin. Immunol. 22 (2010) 109–117.

433 [23] F. Buonocorea. M. Gerdol. Alternative adaptive immunity strategies: coelacanth, cod and
434 shark. Immunity. Mol. Immunol. 69 (2016) 157–169.

435 [24] J. Kumari, J. Bøgwald, R.A. Dalmo, Eomesodermin of Atlantic salmon: an important
436 regulator of cytolytic gene and interferon gamma expression in spleen lymphocytes, PLoS.
437 One 8 (2013) e55893.

438 [25] F. Takizawa, K. Araki, K. Ito, T. Moritomo, T. Nakanishi, Expression analysis of two
439 eomesodermin homologues in zebrafish lymphoid tissues and cells, Mol. Immunol. 44 (2007)
440 2324–2331.

441 [26] F. Takizawa, K. Araki, M. Ohtani, H. Toda, Y. Saito, V.S.Lampe, J.M. Dijkstra,
442 Transcription analysis of two Eomesodermin genes in lymphocyte subsets of two teleost
443 species. Fish Shellfish Immunol. 36 (2014):215–222.

444 [27] H. Chi, Z. Zhang, J. Bøgwald, W. Zhan, R.A. Dalmo, Cloning, expression analysis and
445 promoter structure of TBK1 (TANK-binding kinase 1) in Atlantic cod (*Gadus morhua* L.),

446 Fish Shellfish Immunol. 30(2011) 1055-1063.

447 [28] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, J.S. Sammut, H.R. Hotz, et al., The Pfam protein

448 families database, Nucleic Acids Res. 36 (2008) D281-D288.

449 [29] J.J. Campanella, L. Bitincka, J. Smalley, MatGAT: an application that generates

450 similarity/identity matrices using protein or DNA sequences, BMC. Bioinform. 4 (2003) 29.

451 [30] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: molecular evolutionary genetics analysis

452 (MEGA) software version 4.0, Mol. Biol. Evol. 24 (2007) 1596-1599.

453 [31] T. Heinemeyer, E. Wingender, I. Reuter, H. Hermjakob, A.E. Kel, O.V. Kel, et al., Data

454 bases on transcriptional regulation: TRANSFAC, TRRD and COMPEL, Nucleic Acids Res.

455 26 (1998) 362-367.

456 [32] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, et al.,

457 MatInspector and beyond: promoter analysis based on transcription factor binding sites,

458 Bioinformatics 21 (2005) 2933-2942.

459 [33] C. Furnes, M. Seppola, B. Robertsen, Molecular characterisation and expression analysis of

460 interferon gamma in Atlantic cod (*Gadus morhua*). Fish Shellfish Immunol. 26 (2009)

461 285-292.

462 [34] I.A. Aursnes, A.L. Rishovd, H.E. Karlsen, T. Gjøen, Validation of reference genes for

463 quantitative RT-qPCR studies of gene expression in Atlantic cod (*Gadus morhua* L.) during

464 temperature stress. BMC Res Notes. 5 (2011) 104.

465 [35] H. Mikkelsen, V. Lund, L. Martinsen, K. Gravningen, M. B. Schrøderb, Variability among

466 Vibrio anguillarum O2 isolates from Atlantic cod (*Gadus morhua* L.): Characterisation and

467 vaccination studies, Aquaculture 266 (2007) 16-25.

468 [36] J. Kumari, Z. Zhang, T. Swain, H. Chi, C. Niu, J. Bøgwald, et al., Transcription factor T-Bet
469 in Atlantic salmon: characterization and gene expression in mucosal tissues during
470 *Aeromonas Salmonicida* infection, Front. Immunol. 6 (2015) 345.
471 [37] X. He, M.N. Treacy, D.M. Simmons, M.A. Ingraham, L.W. Swanson, M.G. Rosenfeld,
472 Expression of a large family of POU-domain regulatory genes in mammalian brain
473 development, Nature 340 (1989) 35-42.
474 [38] S. Kouyasu, R.E. Hussey, L.K. Clayton, A. Lerner, R. Pedersen, P. Delany-Heiken, et al.,
475 Targeted disruption within the CD3 zeta/eta/phi/Oct-1 locus in mouse, EMBO. J. 13 (1994)
476 784-797.
477 [39] H.H. Salamanca, N. Fuda, H. Shi, J.T. Lis, An RNA aptamer perturbs heat shock
478 transcription factor activity in Drosophila melanogaster, Nucleic Acids Res. 39 (2011) 6729–
479 6740
480 [40] H.H. Salamanca, M.A. Antonyak, R.A. Cerione, H. Shi, J.T. Lis, Inhibiting heat shock factor
481 1 in human cancer cells with a potent RNA aptamer, PLoS.ONE 9 (2014) e96330.
482 [41] C. Englund, A. Fink, C. Lau, D. Pham, R.A. Daza, A. Bulfone, et al., Pax6, Tbr2, and Tbr1
483 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic
484 neurons in developing neocortex, J. Neurosci. 25 (2005) 247–251.
485 [42] A. Sessa, C.A. Mao, A.K. Hadjantonakis, W.H. Klein, V. Broccoli, Tbr2 directs conversion
486 of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis
487 in the developing neocortex glish, Neuron. 60 (2008) 56–69.
488 [43] I. Costello, I.M. Pimeisl, S. Dräger, E.K. Bikoff, E.J. Robertson, S.J. Arnold, The T-box
489 transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm

490 during mouse gastrulation, Nat. Cell Biol. 13 (2011) 1084–1091.

491 [44] J.J. Knox, G.L. Cosma, M.R. Betts, L.M. McLane, Characterization of T-bet and eomes in

492 peripheral human immune cells, Front. Immunol. 5 (2014) 217.

493 [45] F. Takizawa, Y. Mizunaga, K. Araki, T. Moritomo, M. Ototake, T. Nakanishi, GATA3

494 mRNA in ginbuna crucian carp (*Carassius auratus langsdorfii*): cDNA cloning, splice

495 variants and expression analysis, Dev Comp Immunol. 32 (2008) 898–907.

496 [46] F. Takizawa, E.O. Koppang, M. Ohtani, T. Nakanishi, K. Hashimoto, U. Fischer, et al.,

497 Constitutive high expression of interleukin-4/13A and GATA-3 in gill and skin of salmonid

498 fishes suggests that these tissues form Th2-skewed immune environments, Mol Immunol. 48

499 (2011):1360-1368.

500 [47] I. Frans, C.W. Michiels, P. Bossier, K.A. Willems, B. Lievens, H. Rediers, *Vibrio*

501 *anguillarum* as a fish pathogen: virulence factors, diagnosis and prevention, J. Fish Dis. 34

502 (2011) 643–661.

503 [48] X. Zhang, S. Wang, S. Chen, Y. Chen, Y. Liu, C. Shao, et al., Transcriptome analysis

504 revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during

505 *Vibrio anguillarum* infection, Fish Shellfish Immunol. 43 (2015) 209–218.

506 [49] I. Frans, K. Dierckens, S. Crauwels, A. Van Assche, J.J. Leisner, J. Leisner, et al., Does

507 virulence assessment of *Vibrio anguillarum* using sea bass (*Dicentrarchus labrax*) larvae

508 correspond with genotypic and phenotypic characterization, PLOS. One 8 (2013): e70477.

509 [50] C. Baker-Austin, L. Stockley, R. Rangdale, J. Martinez-Urtaza, Environmental occurrence

510 and clinical impact of *Vibrio vulnificus* and *Vibrio parahaemolyticus*: a European perspective,

511 Environ. Microbiol. Rep. 2 (2010) 7–18.

512 [51] C. Yang, J. Zhang, F. Li, H. Ma, Q. Zhang, T.A. Jose Priya, et al., A Toll receptor from
513 Chinese shrimp *Fenneropenaeus chinensis* is responsive to *Vibrio anguillarum* infection,
514 Fish Shellfish Immunol. 24 (May 2008) 564–574.

515 [52] A. Weiss, R.L. Wiskocil, J.D. Stobo, The role of T3 surface molecules in the activation of
516 human T cells: a two-stimulus requirement for IL 2 production reflects events occurring at a
517 pre-translational level, J. Immunol. 133 (1984) 123-128.

518 [53] B. Manger, K.J. Hardy, A. Weiss, J.D. Stobo, Differential effect of cyclosporin A on
519 activation signaling in human T cell lines, J. Clin. Invest. 77 (1986) 1501-1506.

520 [54] Y. Zhu, S. Ju, E. Chen, S. Dai, C. Li, P. Morel, et al., T-bet and eomesodermin are required
521 for T cell-mediated antitumor immune responses, J. Immunol. 185 (2010) 3174–3183.

522

523 **Table 1.** List of primers and their designated applications

524 **Table 2.** Identities (%) of Eomes in 10 vertebrate species at nucleotide and amino acid levels. The
525 percentage identity values of nucleotides are marked with numbers in bold and the percentage
526 identity of amino acids are marked as numbers in italic font. Matrix Global Alignment Tool
527 (MatGAT) was applied in the calculation.

528 **Fig.1.** Nucleotide and deduced amino acid sequences of GmEomes cDNA. Start and stop codons
529 are in bold. The TBOX_3 domain is bordered. The RNA instability motifs (attta) are marked with
530 underlines. The putative polyadenylation (aataaa) signal is in bold and underlined.

531 **Fig.2.** Multiple alignment of the deduced amino acid sequences of Eomes in cod and other
532 vertebrates by the ClustalW program. Residues shaded in the same colour are completely
533 conserved across all species aligned, and residues shaded in different colour to 50–90% identity.
534 Dashes indicate gaps. The GenBank accession numbers of the Eomes sequences are as follows:
535 human: BAH13105.1; monkey: XP_002803087.1; chimpanzee: XP_526157.2; dog: XP_542755.2;
536 Norway rat: XP_001061749.2; mouse: NP_034266.2; newt: BAA84718.1; frog: AAI25987.1;
537 gray bichir: BAH58788.1; Zebrafish: NP_571754.3.

538 **Fig.3.** Phylogenetic tree showing the relationship between GmEomes and other vertebrate amino
539 acid sequences. The phylogram was constructed with the MEGA 4.0 software using the
540 neighbor-joining method based on an amino acid alignment (ClustalW) of the full-length protein.
541 Numbers beside the internal branches indicate bootstrap values based on 1 000 replications. The
542 0.1 scale indicates the genetic distance. The position of Atlantic cod (*Gadus morhua* L.) Eomes
543 was marked with "●". GenBank accession numbers for some Eomes sequences are listed in the
544 legend of Fig.2, while the others are listed as follows: Eomes: *Salmo salar*: NP_001191029.1;

545 *Gallus gallus*: XP_426003.2; TBR1: *Danio rerio*: AAG_48249.1; *Xenopus tropicalis*: AAI36087.1;

546 *Ailuropoda melanoleuca*: XP_002924534.1; *Canis lupus*: XP_545492.2; *Callithrix jacchus*:

547 XP_002798955.1; *Pan troglodytes*: XP_001150172.1; *Homo sapiens*: AAI13419.1; *Rattus*

548 *norvegicus*: NP_001177999.1; *Bos taurus*: NP_ 001178978.1; *Equus caballus*: XP_001493157.2;

549 TBX21: *Salmo salar*: ADP36855.1; *Carassius auratus langsdorfii*: BAF73805.1; *Danio rerio*:

550 NP_001164070.1; *Mus musculus*: NP_062380.2; *Rattus norvegicus*: NP_001100513.1; *Canis*

551 *lupus familiaris*: XP_548164.1; *Homo sapiens*: BAJ21009.1; *Pan troglodytes*: XP_001173500.1.

552 **Fig.4.** The sequence and activity of Eomes promoter. A**.** The nucleotide sequence of promoter

553 regions (694 bp) was determined. The transcription start site is designated as +1 and boxed.

554 Transcription factor binding sites were predicted by MatInspector and TRANSFAC®. Consensus

555 elements of transcription factor binding sites are underlined, while (−) sign indicates the binding

556 sites identified on the negative strand. B. CHSE-214 cells were transiently transfected with the

557 promoter constructs plus pSEAP2 internal control vector in 24-well plates. Luciferase activity is

558 expressed relative to SEAP (mean ± SD from six wells). Double asterisks (**) above the bars

559 show significant differences ($P < 0.01$) compared to the different promoter constructs with the

560 same treatment. The data are from six cell wells per treatment in one experiment and are

561 representative of three independent experiments.

562 **Fig.5.** Tissue distribution of GmEomes transcripts. The expression of GmEomes mRNA was

563 determined by qRT-PCR in different organs. The results were calculated by relative expression

564 with β-actin as the house keeping gene and muscle as a calibrator. The value above the bars shows

565 average real-time CT values of six fish. Data are represented as mean ± SD (N=6).

566 **Fig.6.** Specific expression of GmEomes in Atlantic cod spleen (A) and head kidney (B) during *V.*

567 *anguillarum* infection and spleen leukocytes (C) at different time-points upon PolyI:C or PMA

568 stimulation. (A) Spleen. Data are presented as means ± SD (N = 6). Untreated fish spleen as

569 calibrator for GmEomes. (B) Head kidney. Data are presented as means ± SD (N = 6). Untreated

570 fish head kidney as calibrator for GmEomes. (C) Head kidney leucocytes. Data are presented as

571 means ± SD (N = 6). Untreated fish spleen leucocytes as calibrator for GmEomes. ** $P < 0.01$ and

572 * $P < 0.05$ show the significantly differences compared the control group.

573 **Fig.7.** The regulation of GmEomes on Atlantic cod IFNγ (GmIFNγ) promoter**.** A. The micrograph

574 of overexpression of RFP and GmEomes-RFP (red) in CHSE-214 cells. The colour of blue

575 showed nuclei stained by DAPI (blue). B. Gene sequence and schematic representation of the cod

576 IFNγ promoter. The transcription start site is designated as +1 and boxed. Eomes binding sites

577 were predicted by MatInspector and TRANSFAC®. C. GmIFNγ promoter activity in CHSE-214

578 cells. Luciferase and SEAP activities were measured at 48 h after the transfection. Luciferase

579 activity is expressed relative to SEAP (mean ± SD, n=6). Double asterisks (**) above the bars

580 show significant differences ($P < 0.01$) compared to the control group. The data are from six cell

581 wells per treatment in one experiment and are representative of at least three independent

582 experiments.

583

584

585

586 Table 1

Oligo Name	Sequence 5' to 3'	Use
GmEomefw1	ccccggctgcacatcgtggaggttaccga	3’-RACE
GmEomefw2	cgtggaggttaccgaggagggcgtggagg	3’-RACE
GmEomefw3	agcaacgaggcgcgagacacagaccttca	3’-RACE
GmRomerv1	tttcgcgaaggggttgtggtctatcttcag	5’-RACE
GmRomerv2	tggtaggcggtcacggcgatgaactggtt	5’-RACE
GmRomerv3	aggtctgtgtctcgcgcctcgttgctcat	5’-RACE
AP1	gtaatacgactcactatagggc	Genome walking
AP2	actatagggcacgcgtggt	Genome walking
GmEomes gwrv1	atactcgatctgtgaggagccggggctgt	Genome walking
GmEomes gwrv2	ttgcgctgtcggacgacgagaggttataaa	Genome walking
GmEomesEcoRIfw1	gcttcgaattcaaattgtgaactaaatgccc	Promoter cloning
GmEomesEcoRIfw2	gcttcgaattctcaatcttccgagaatcctattta	Promoter cloning
GmEomesEcoRIfw3	gcttcgaattcctatgtcgataaggcaagtgcattt	Promoter cloning
GmEomesSacIIrv	taagcccgcggatgtgcaactccgattatgatct	Promoter cloning
GmEomesRTfw1	cgacatggccaacccgtgct	Real-time PCR
GmEomesRTrv1	ccgatgctggatcccgtcgc	Real-time PCR
GmIFNγ F	tggtctgcatgtcagtttgtctg	Real-time PCR
GmIFNγ R	ttctgtggatgttgttggctaaga	Real-time PCR
GmEF1α F	atgtgagcggtgtggcaatc	Real-time PCR
GmEF1α R	tcatcatcctgaaccaccctg	Real-time PCR
GmEomesORFfw	cagatctcgagatgcagttggagaacatccttcct	Plasmid construction
GmEomesORFrv	ttgagctcgagggggctcgtgtagaacgcatagta	Plasmid construction
GmIFNγpromfw	cagatctcgagtctgccttcctagtaagtgaggctgc	Plasmid construction
GmIFNγpromrv	ttgagctcgagccctgcagctgtacacgctgaaagtcgc	Plasmid construction

587

588

589

Table 2

	1	2	3	4	5	6	7	8	9	10
1. Human		97.9	72.7	66.3	69.6	59.8	59.2	59.8	55.7	56.5
2. Monkey	*98.3*		73.3	66.1	70.3	59.8	59.2	60.9	56.4	55.9
3. Dog	*72.8*	*73.6*		80.9	89.8	63.4	63.5	66.9	59.1	59.9
4. Norway rat	*69.8*	*70.1*	*85*		88.4	60.9	60.4	62.5	55.9	57.1
5. Mouse	*70.9*	*71.4*	*92.8*	*90.4*		63.3	62.8	65	58.5	59.5
6. Gray bichir	*67.8*	*67.6*	*72.4*	*73.3*	*72.4*		64.7	67.8	73.6	74.4
7. Newt	*65.8*	*65.9*	*73.7*	*72.8*	*73.1*	*75.9*		70.9	62.7	62.5
8. Frog	*66.4*	*67.9*	*76*	*74.6*	*74.7*	*79.4*	*79.8*		61.7	64.4
9. Zebrafish	*64.1*	*65.1*	*71.7*	*72.1*	*71.3*	*85.8*	*73.8*	*76.1*		78.1
10. Cod	*64.8*	*64.4*	*72.3*	*73*	*71.9*	*84.7*	*74.7*	*78.5*	*89.9*	

590

591

592

593 Figure 1

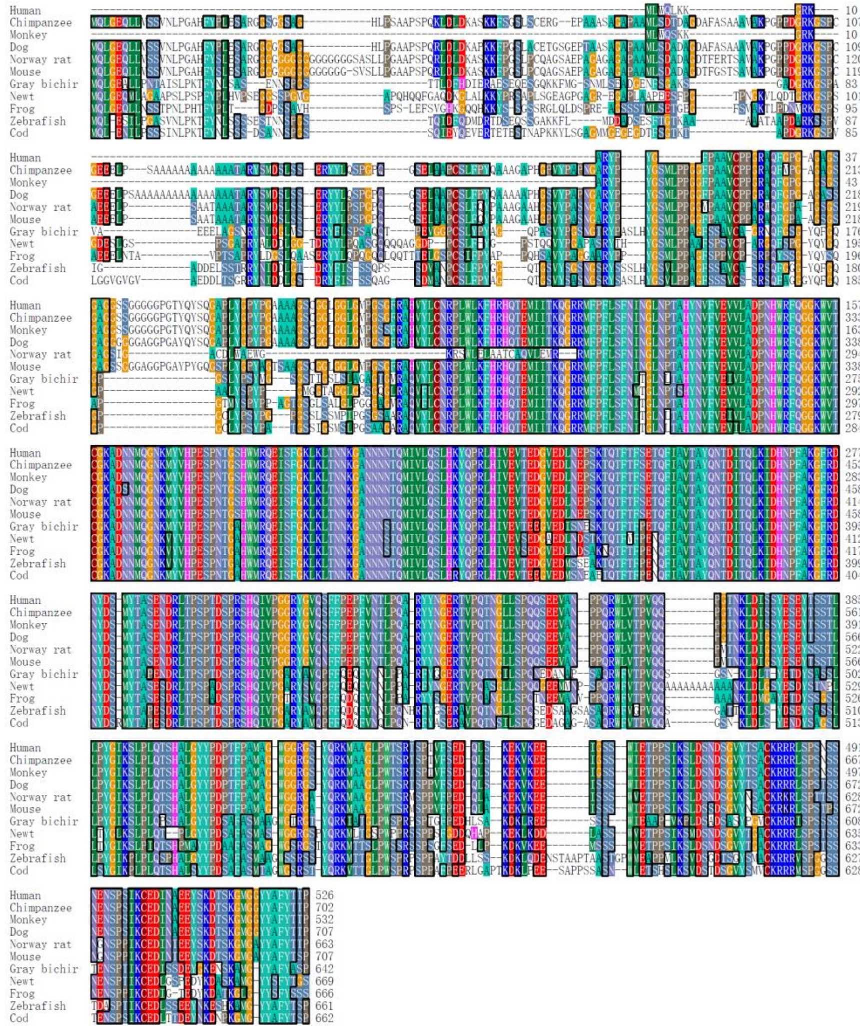
```
agatcataatcggagttgcacatctctgggtgaggaagagaaggccaagtcctgaagtgaggagcaggaacgggaggacgcgcaccgcag 90
atctaccaaccgaggcgacactcgtacagtgtcgctcctgggcttatactcacttttggggtcccgagtcttttttggttactccaaag 180
gttaaacgatgcagttggagaacatccttcctagctcgagcatcaacttacccaagacgttttataacctctcgtcgtccgacagcgcaa 270
         M  Q  L  E  N  I  L  P  S  S  S  I  N  L  P  K  T  F  Y  N  L  S  S  S  D  S  A  27
ataacagccccggctcctcacagatcgagtatcaagaagtggagcgaaccgagacggaatccacgaacgccccgaagaagtacctgagcg 360
N  N  S  P  G  S  S  Q  I  E  Y  Q  E  V  E  R  T  E  T  E  S  T  N  A  P  K  K  Y  L  S  57
gagcggggatgatgggagaggggaaggggacactttctctgggactaagaccgcccccgatgggaggaaaggctctccggttctcggtg 450
G  A  G  M  M  G  E  G  E  G  D  T  F  S  G  T  K  T  A  P  D  G  R  K  G  S  P  V  L  G  87
gtgttggtgttggtgttgcagaggacgacctgacaagtggccggcgatacaacatagacgatctgggctccgacagatactttatctcgt 540
G  V  G  V  G  V  A  E  D  D  L  T  S  G  R  R  Y  N  I  D  D  L  G  S  D  R  Y  F  I  S  117
cctctcaggcgggttccgacatggccaacccgtgctccttgttcccctacgcgggacagaccggctcggtgtacagcgcctccaacggct 630
S  S  Q  A  G  S  D  M  A  N  P  C  S  L  F  P  Y  A  G  Q  T  G  S  V  Y  S  A  S  N  G  147
ccaggtattcggcgtctctgcactacggatccgtgctgccccccgcgggcttctcgtcctctgtgtgctccagtcgcagccagtttgccg 720
S  R  Y  S  A  S  L  H  Y  G  S  V  L  P  P  A  G  F  S  S  S  V  C  S  S  R  S  Q  F  A  177
gcggaggataccagttcggccagggtccgggctgtctgtacccttcttacccagcgacgggatccagcatcggctccatgtcgctgcccg 810
G  G  G  Y  Q  F  G  Q  G  P  G  C  L  Y  P  S  Y  P  A  T  G  S  S  I  G  S  M  S  L  P  207
ggtccgcggccggagccagggcgcaggtttatctgtgcaaccggcccctctggctcaagtttcaccggcaccagacggagatgatcatca 900
G  S  A  A  G  A  R  A  Q  V  Y  L  C  N  R  P  L  W  L  K  F  H  R  H  Q  T  E  M  I  I  237
ccaaacaggggcggcgaatgttcccattcctcagtttcaacatcaccgggctcaacctgacagctcattacaacgtgtttgtagaagtca 990
T  K  Q  G  R  R  M  F  P  F  L  S  F  N  I  T  G  L  N  L  T  A  H  Y  N  V  F  V  E  V  267
tcctggcagaccccaatcactggcgctttcagggaggcaaatgggtcacctgtgggaaagcggacaataatatgcaggggaacaagatgt 1080
I  L  A  D  P  N  H  W  R  F  Q  G  G  K  W  V  T  C  G  K  A  D  N  N  M  Q  G  N  K  M  297
acgtccaccccgaatcccccaataccggggcgcactggatgagacaagaaatctcatttggcaaactgaagctgaccaacaacaagggtg 1170
Y  V  H  P  E  S  P  N  T  G  A  H  W  M  R  Q  E  I  S  F  G  K  L  K  L  T  N  N  K  G  327
ccaacaataataacacacagatgatcgtgctacagtccctccacaggtaccagccccggctgcacatcgtggaggttaccgaggagggcg 1260
A  N  N  N  N  T  Q  M  I  V  L  Q  S  L  H  R  Y  Q  P  R  L  H  I  V  E  V  T  E  E  G  357
tggaggacatgagcaacgaggccgagacacagaccttcacctttccggagaaccagttcatcgccgtgaccgcctaccaaaacactgaca 1350
V  E  D  M  S  N  E  A  E  T  Q  T  F  T  F  P  E  N  Q  F  I  A  V  T  A  Y  Q  N  T  D  387
tcacacaactgaagatagaccacaaccccttcgcgaaaggtttccgggacaattatgactccaggatgtacacggccccggagagcgaca 1440
I  T  Q  L  K  I  D  H  N  P  F  A  K  G  F  R  D  N  Y  D  S  R  M  Y  T  A  P  E  S  D  417
ggttgaccccgtcccccaccgactcgccgcgctcccaccagatcgtcccgggcgcccgctacgccatgcagcccttcttccaggaccagt 1530
R  L  T  P  S  P  T  D  S  P  R  S  H  Q  I  V  P  G  A  R  Y  A  M  Q  P  F  F  Q  D  Q  447
tcgtcaaccagctgccgcagaaccgcttctacgccagcgagcgcgccgtgccacagaccaacagcatcctgtccccgcagggtgaggacg 1620
F  V  N  Q  L  P  Q  N  R  F  Y  A  S  E  R  A  V  P  Q  T  N  S  I  L  S  P  Q  G  E  D  477
cgggcgccggcgcctccgcgcagcgctggttcgtcacgcccgtgcaacaggcgggctccaacaaactggacttgtcctatgagaacgact 1710
A  G  A  G  A  S  A  Q  R  W  F  V  T  P  V  Q  Q  A  G  S  N  K  L  D  L  S  Y  E  N  D  507
actccgccggcagcctgctgtcgtacggcatcaagccgctgcccctgcagacgtcccacgccctcagctactaccccgactcggccttcg 1800
Y  S  A  G  S  L  L  S  Y  G  I  K  P  L  P  L  Q  T  S  H  A  L  S  Y  Y  P  D  S  A  F  537
cctccatgacggcgggctggggcagccgcagcacttaccagcgcaaggtgaccacgggcctgccctggtccccgcggcccagcccccccg 1890
A  S  M  T  A  G  W  G  S  R  S  T  Y  Q  R  K  V  T  T  G  L  P  W  S  P  R  P  S  P  P  567
ccttccccgaggagcggctgggtgcccccacgaaggacaagctgccggaggagagcgcgccgccgtcgtcggcctcgaactggctggaga 1980
A  F  P  E  E  R  L  G  A  P  T  K  D  K  L  P  E  E  S  A  P  P  S  S  A  S  N  W  L  E  597
cgtcgcactcgctgaagtcggtggactcgacggactctggggtgtactccatggtctgcaagaggcgccggatgtcccccggggggctcca 2070
T  S  H  S  L  K  S  V  D  S  T  D  S  G  V  Y  S  M  V  C  K  R  R  R  M  S  P  G  G  S  627
gcacggagaactccccgagcattaagtgcgaggacttgaccacggacgagtacaacaaggacaacccgaaaggcatgggctactatgcgt 2160
S  T  E  N  S  P  S  I  K  C  E  D  L  T  T  D  E  Y  N  K  D  N  P  K  G  M  G  Y  Y  A  657
tctacacgagcccctaaataagactgttgtgtaccgaaccagaattataggctaaagcatttctttttttaaattctttttttttaattt 2250
F  Y  T  S  P  -                                                                          662
ttgtatccctgataccgttacttcccttgttttccctcaaagcaaaaagcacatggttattttcccaggccagatcccccatacctgagc 2340
ccaatttagttttattctgttttgaggcatgcgactccccctccctttctcccgcgcacgcacacacacctcttttatttaatttttttac 2430
aactgtaattttttgttattttttaatttaaattaaggctatttaaaggtccttttttgatgtacagtatttgtgcactttagaatgtgatg 2520
tatcttgtacaaagtgtcttcatggaggtggtataaaatgggtaataaacagcttttcataaagcattaaaaaaaaaaaaaaaaaaaaa 2610
aaaaaaaa                                                                                   2618
```

594

595

596

597 Figure 2

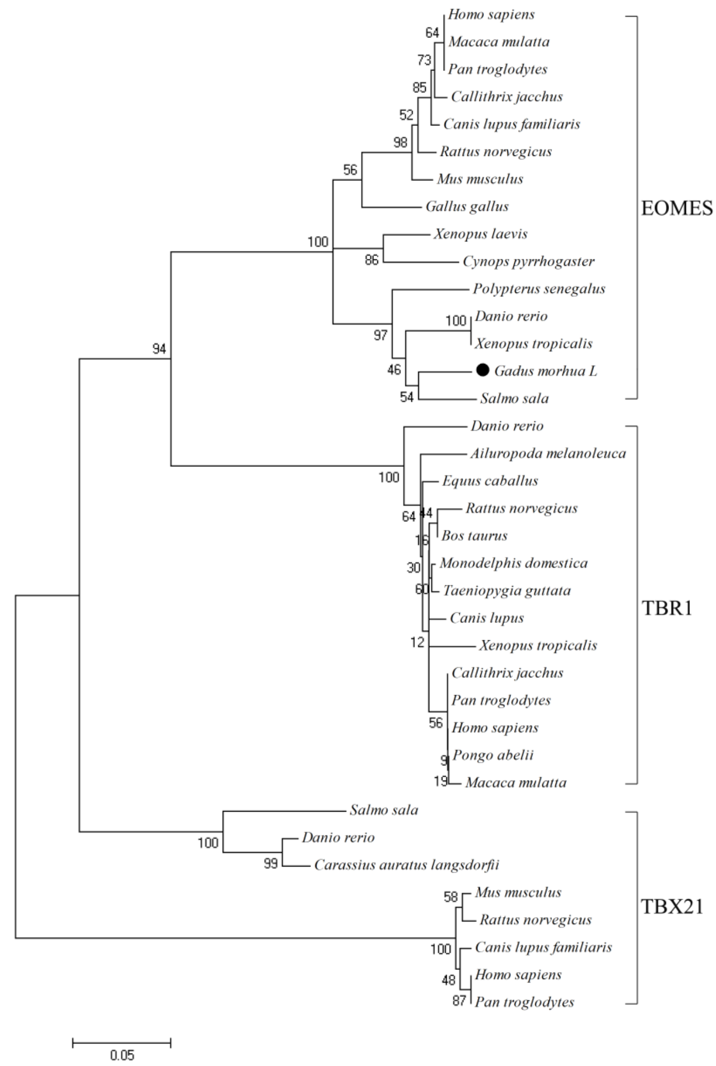


598

599

600

601 Figure 3



602

603

604

605 Figure 4

606 A

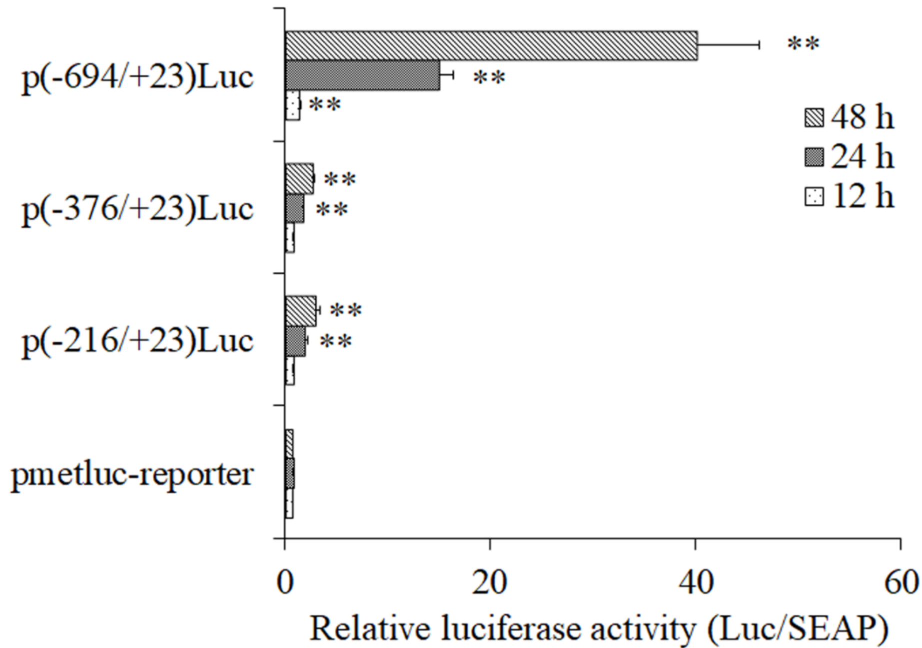
```
-694  AAATTGTGAACTAAATGCCCCACAACCTGTCCTAATAATAAAATATCATAAGGTTTCAGG
                                       Oct-1
-634  TTGTTCATTCTAAAGTGGGTATTGGCACACCGGTCACCTGCGGTGCGGAGGTGGGCGCGT
                                      E47(-)
-574  TCAGCGACATGAGGAGGCCAATTAGAAAGTTCCACATTTCTTGTCGGAAGTCTTGCACAT
                                HSF2
-514  GTGTTGTCTAATTGCGTCGCTGCATGAACATCTGATCATTGTATGTGACGGCACCGAAAA
       MyoD
-454  TGCTGAACTATGCTACTGGTTTAGTTTCTTTTTTAAGTTTTACGTTTTAATTTATCTACT
                              SRY
-394  AATTATGTAGGATTGATTCAATCTTCCGAGAATCCTATTTATCTGAAACATTTTAATGGA
        S8(-)                                GATA-1(-)
-334  CTTGCTTTGTGGATTTCAATCTGTAAATAAATGTAGACTAAACACTTGCTCTCGAAATGC
                                 CdxA
-274  CTGTCCAGGATCGGGGTAAGTCTGACATCCTTCTTATAAAACATGTGTAATTTAGTGCTA
          GATA-1/2                          TATA
-214  TGTCGATAAGGCAAGTGCATTTAGTTCGACTAATTGATTGGGAGAAATATGACCATCGGA
          GATA-x                      S8(-)
-154  CGTGCAGAGGACTGGAACCAATTCCAAGGATTCAGAAAGCTCCTCCCATGTCGCTTGCAG

 -94  CGAGTCGTGGGTGTGTGTGACTGAGCGTGTGTGTGTGAGAGTGTCTGTGTGTTGAGAACA
                     NF-E2(-)
 -34  GCGTCTTATAAGAGCCATTGAACCACCCCTGACCAGATCATAATCGGAGTTGCACATCTC
                                         +1
 +27  TGGGTGAGGAAGAGAAGGCCAAGTCCTGAAGTGAGGAGCAGGAACGGGAGGACGCGCACC
 +87  GCAGATCTACC
```

607

608 B.

609



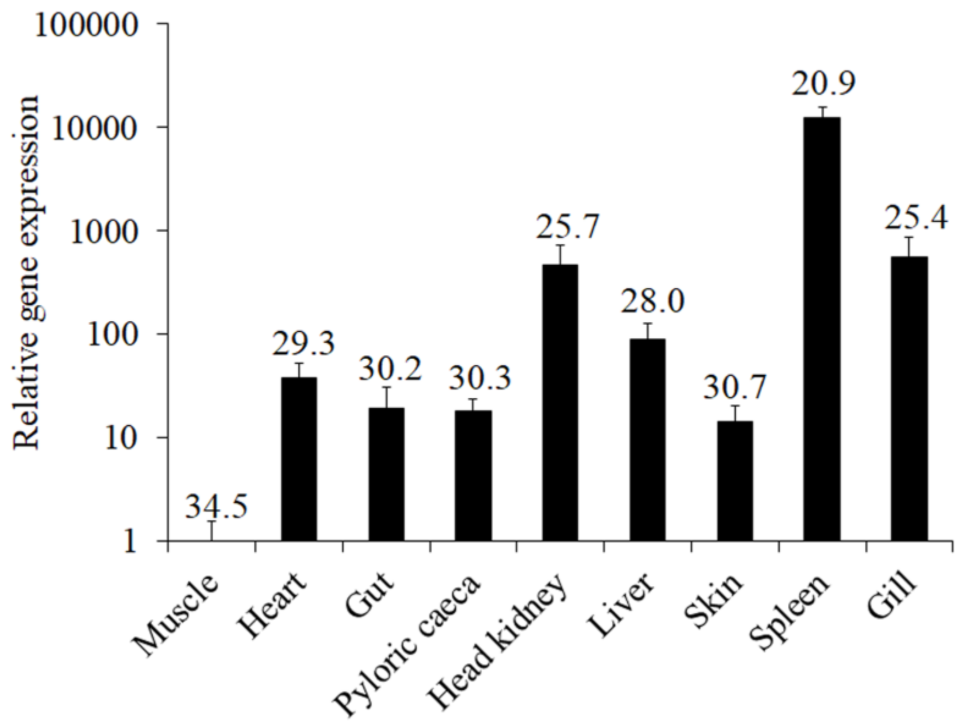
610

611

612

613

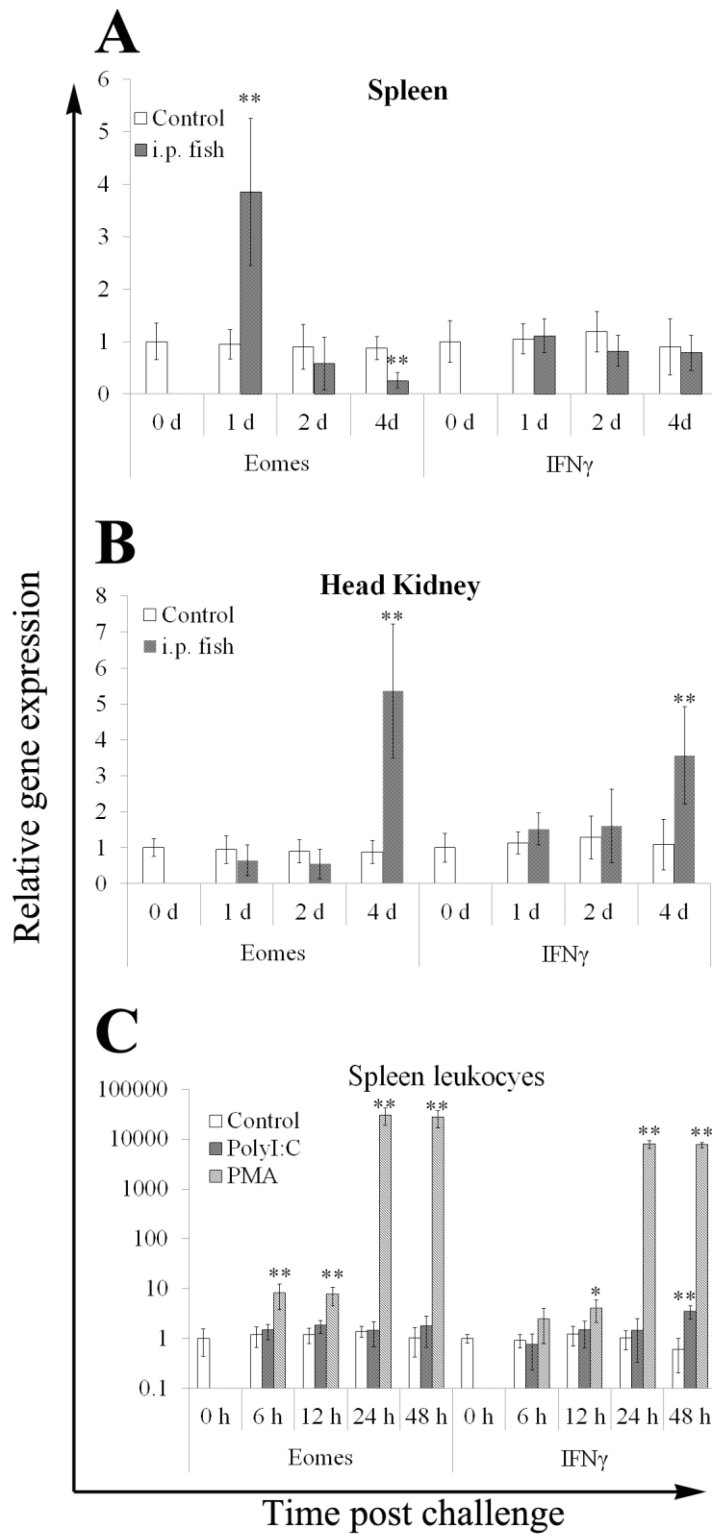
614 Figure 5



615

616

617 **Figure 6**

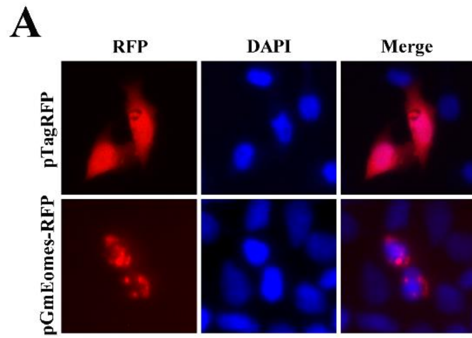


618

619

620

621 **Figure 7**



B

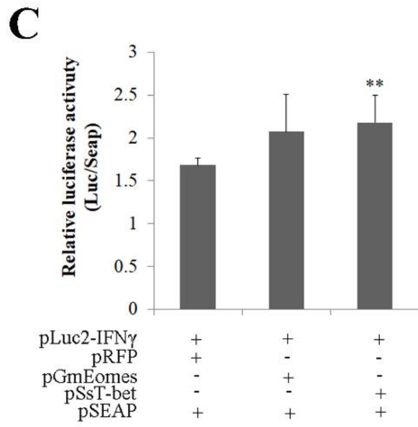
```
GCATATAGTCACGTATACTCTTATATAGTAGTACATAGTCAATTATCAACATACAGTATA -528

CTCTTATATAGTCGTATATGGTGTTATATAGTCATATTGAGTGTTATATTATAATATGTA -468
                 Eomes              Eomes
GTCTTCTACGGGGTAAGTCGTTGCTGTCGTCTATCTCTAAGCTGTGGTTGTATGGGCTGT -408
                                       Eomes
GGTCGTCATAGAATGCTCTTCACAGCAACAGTTGCACAAGCTAGGATAACCGAGGCTAGC -348

TAACACCAGCCAGACTAACACACAAGCCTAACATTGTGGCTAACTAGCATAGCCAAGCTA -288

CCTTAAAGGGTAGTAAATCTCTTACTTCCTCTTACTTTTGGATAGCTTACGTGTGTGTGT -228
                                                  Eomes
GTGTGTGTGTGTGTGTGTGTGTGTGCGTGTGTGTGTGCGTGTGTGCTGTGAAGTTTGACG -168
                                                  Eomes
ATTTAAAGGAAACTTTAGACCACATTTTCACACCTCGATCGTCCCACCTCTCTCTCTCTC -108

TCTCATATATAAAGTGCTCATCTTAACTGATAACACTCACTTCATTCACCTGAACTCTGC -48
                                             +1
ACTGGACGGATATATGGGCGGAAAGACAGAGAGACAGACAGACTGACGGGCAGAAAGACA +13
GACAGAGAGACAATAGAAAAGTAGACAGACAAGCAGAAGGACAGACAGAAAGGCTGTATC +73
GTCCTGCCGGGGATTCGAACAGGCGACTTTCAGCGTGTACAGCTGCAGGG           +123
```



622

623

624