Matricellular protein SPARC/Osteonectin expression is regulated by DNA methylation in its core promoter region.

Torres-Núñez, E[1]., Cal, L[1]., Suarez-Bregua, P[1,3]., Gomez-Marin C[2]., Moran, P[3]., Gomez-Skarmeta JL[2]., Rotllant, J[1*].

[1]Aquatic Molecular Pathobiology Lab. Instituto de Investigaciones Marinas (IIM-CSIC). Eduardo Cabello 6, 36208-Vigo, Spain.

[2] Centro Andaluz de Biología del Desarrollo (CABD), CSIC-UPO-Junta de Andalucía. 41013 Sevilla. Spain

[3]Dep. Bioquímica, Xenética e Inmunoloxía. Facultade de Bioloxía. Universidade de Vigo. 36310 Vigo. Spain.

*Corresponding author

Josep Rotllant,

Instituto de Investigaciones Marinas, CSIC. Eduardo Cabello, 36208, Vigo (Pontevedra), Spain

Tel. (+34) 986-231930

Fax (+34) 986 292 762

E-mail: rotllant@iim.csic.es

ABSTRACT

Background: SPARC/Osteonectin is an evolutionarily conserved matricellular protein that modulates cell-matrix interaction and cell function. In all vertebrates, *SPARC* is dynamically expressed during embryogenesis. However, the precise function of SPARC and the regulatory elements required for its expression in particular during early embryogenesis are largely unknown. Results: The present study was undertaken to explore the molecular mechanisms that regulate *sparc* gene expression by *in-vivo* functional characterization of the *sparc* promoter and identification of possible putative regulatory elements that govern basal promoter activity. We report here transient expression analyses of *eGFP* expression from transgenic zebrafish containing a *Sparc-iTol2-eGFP-BAC* and/or 7,25kb-*sparc-Tol2-eGFP* constructs. *eGFP* expression was specifically found in the notochord, otic vesicle, fin fold, intermediate cell mass and olfactory placode of BAC and Tol2 transposon vectors injected embryos. Deletion analysis revealed that promoter activity resides in the unique 5′-UTR intronic region. Computer-based analysis revealed a putative CpG island immediately proximal to the translation start site within the intron sequence. Global inhibition of methylation with 5-Aza-2-deoxycytidine promoted *sparc* expression in association with decreasing CpG methylation. Conclusion: Taken together, these data identify a contributory role for DNA methylation in regulating *sparc* expression in zebrafish embryogenesis.

Key words: Sparc, osteonectin, zebrafish, transcriptional regulation, methylation, CpG island

INTRODUCTION

SPARC (secreted protein acidic and rich in cysteine), also named osteonectin or BM-40, is a non-structural component of the extracellular matrix that it is thought to modulate cell–matrix interactions, particularly during tissue remodelling and at sites of high cellular turnover during development, wound-healing and carcinogenesis. SPARC is spatially and temporally regulated during development and displays a high degree of sequence conservation (Laize et al., 2005; Rotllant et al., 2008), indicating a conserved, essential functional role in vertebrates. We have previously shown that the zebrafish *sparc* gene is also expressed in a temporally and spatially specific manner, with strong expression in the developing inner ear and pharyngeal cartilage (Rotllant et al., 2008). We showed further that *sparc* interacts with genes in known genetic networks, unveiling its novel functions in regulating pharyngeal cartilage and inner ear development. We also demonstrated a critical role of Sparc in embryonic haematopoiesis during early development of zebrafish (Ceinos et al., 2013). Specifically, we showed that *sparc* is a modulator of lineage-specific transcription factor *gata 1* expression levels or activity. Furthermore, we demonstrated a UV exposure-dependent increase in the expression of the matricellular protein osteonectin in zebrafish embryos (Torres-Nuñez et al., 2012). Therefore, because of its spatially and temporally regulated expression during development and its multifunctional role, *sparc* gene is expected to be tightly regulated. Although isolation of *sparc* promoter regions from different species has been reported (Mcvey et al., 1988; Young et al., 1989; Damjonovski et al., 1998), little is known about regulation of *sparc* gene expression at the transcriptional level during embryogenesis. It has been shown that the 5' flanking region of the *sparc* gene contains *cis*-regulatory elements that might be responsible for differential expression during normal development in different vertebrate species. Possible regulatory sequences found on *sparc* promoter include GATA factor binding sites, growth hormone consensus sequences, heat shock factors, metal responsive elements, NF1 and SP1 binding and myogenic elements. Another important class of regulatory regions, the so-called CpG islands, has also been found in *sparc* promoter regions (Yang et al., 2007; Gao et al., 2010). CpG islands are specific regions of 200 base pairs (bp) with over 50% G+C content and a CpG frequency of 0.6 (observed/expected ratio) susceptible to transcriptional gene regulation by DNA methylation. Therefore, aberrant methylation of *sparc* promoter has been associated with disc degeneration (Tajerian et al., 2011) and pancreatic, colorectal and ovarian cancers (Cheetham et al., 2008; Socha et al., 2009; Gao et al., 2010). In summary, although numerous factors appear to be involved in transcriptional

regulation of the *sparc* gene, detailed information on the molecular mechanisms regulating *sparc* gene activity is still lacking, and a consistent promoter study has not yet been performed in non-mammalian vertebrates, particularly teleosts, although it has been shown that they apparently have less SPARC or SPARCL1 functional homologues than mammals. Therefore, observations in non-mammalian vertebrates might reveal key functions of SPARC and its regulatory mechanisms.

The present study was undertaken to explore the molecular mechanisms that regulate *sparc* gene expression by *in-vivo* functional characterization of the *sparc* promoter and identification of possible putative regulatory elements that govern basal promoter activity. Specifically, the investigation was focused on the occurrence and role of CpG dinucleotides methylation in the sparc putative promoter sequence and on a possible epigenetic level of regulation of *sparc* expression in fish.

RESULTS

Analysis of 5'-untranslated region of *sparc* sequence in zebrafish

The 5'UTR of *sparc* was characterized by RACE. The start of transcription was found to be separated from the start of translation by a 7-kb intron (Fig. 1A). The whole 5'UTR sequence comprises 138 nucleotides, of which 13 correspond to the region proximal to the *sparc* ATG in exon 2. Therefore, the non-coding exon I contains 125 bp, while exon II contains 70 bp (13 pb from the untranslated region plus 57 bp from the coding sequence).

Functional analysis of the putative promoter region of *sparc* gene in zebrafish

Functional analysis of the putative promoter region of zebrafish *sparc* was done through expression of *eGFP*. BAC clone (number 97O1) (190Kb) from DanioKey zebrafish BAC library containing the *sparc* gene (Fig.1B) was used to explore the putative *sparc* promoter activity. Additionally, to particularly assess whether the 5′-flanking region of the *sparc* gene could support transcription, the 0.25 kb *sparc* promoter and its 5'-flanking sequence 7 kb upstream of the translated exon 2 were linked with the *eGFP* reporter gene (Fig. 1C). To identify the key regulatory region within the P2 construct sequence responsible for *sparc* expression, a third

construct similar to construct P2 but lacking the UTR intron (+126/+7168) (Fig.1D) was also generated (*Tol2-0,25kb-sparc-eGFP;* P3 construct). The resulting gene constructs, *Sparc-iTol2-eGFP-BAC* (P1)*, 7,25kb-sparc-Tol2-eGFP* (P2) and *0,25kb-sparc-Tol2-eGFP* (P3) were microinjected into zebrafish embryos for transient expression analysis. *eGFP* expression in injected embryos was monitored by direct observation under a confocal fluorescence microscope. The promoter activities of P1 and P2 reporter gene constructs are shown in Fig. 2. *eGFP* expression was initially observed at 19 hpf in the notochord (Fig. 1E). At 30 hpf, strong *eGFP* expression was seen in the otic vesicle, pharyngeal region, caudal fin fold and in the hematopoietic ICM region (Fig.1F,G,H). At 48 hpf, strong *eGFP* expression was seen in the olfactory placode, specifically in the olfactory sensory neurons (Fig. 1I). Observed *eGFP* expression from transient expression of *Sparc-iTol2-eGFP-BAC* (P1 construct) and *Tol2-7,25kb-sparc-eGFP* (P2 construt) constructs were equivalent. The P3 construct containing only the proximal promoter and the 5' untranslated region of exon 1 (-127bp/+125bp) did not express *eGFP*, indicating that removal of the 5'UTR-intron (nt+126/+7168) resulted in complete reduction of promoter activity.

To determine whether the endogenous *sparc* gene was specifically expressed in the same domains, *in situ* hybridization was performed with a *sparc* antisense probe (2) in zebrafish embryos. *Sparc* mRNA was indeed expressed in almost all the same domains as in transient *eGFP* expression studies (Fig. 1J,K,L). Therefore, *eGFP* expression in the *Sparc-iTol2-eGFP-BAC* and *Tol2-7,25kb-sparc-eGFP* constructs injected fish mostly recapitulated the endogenous expression pattern of *sparc* mRNA (2).

These data indicated that the 0.25-kb *sparc* promoter and its 5'-flanking sequence 7 kb upstream of the translated exon contained the regulatory element(s) that temporally and dynamically drive tissue-specific expression in the notochord, pharyngeal region, otic vesicle, intermediate cell mass, fin fold and olfactory sensory neurons.

On the basis of our finding that the 5'UTR-intron +126bp to +7168 bp region is key to the transcriptional regulation of *sparc,* we sought to identify possible putative regulatory elements that could be involved in this regulation. Using Computer base-analysis, we identified that the +6634 to +6905 region within the *sparc* 5′-UTR intron 1 sequence display all the characteristic signatures of a genomic CpG island; therefore we hypothesized a role of CpG dinucleotide methylation in the regulation of *sparc* expression.

In addition, using the MatInspector database, we also identified several putative transcription factor binding sequences in the (+126/+7168) region, including sites for heat shock elements, cAMP responsive element binding proteins, gata factors, sox factors, myoblasts factors, glucocorticoid response elements, retinoic acid receptors and activating protein-1. These putative regulators have been shown to be involved in sparc transcriptional regulation in other species, but most were located upstream the initial transcription site instead of the intronic sequence as we found in zebrafish (Young et al., 1989; Damjanovski et al., 1998). Moreover, zebrafish sparc promoter lacks both the consensus CAAT box and TATA box, elements usually associated with developmentally regulated genes.

Experimentally induced 5-Aza-2-deoxycytidine DNA hypomethylation

As we detected a CpG island immediately proximal to the translation start site in the intron (i1-2) sequence (Fig.1A), we hypothesized a role of CpG dinucleotide methylation in the regulation of sparc expression. Therefore, to investigate the relationship between sparc transcriptional activation and CpG methylation, we treated developing zebrafish larvae with 50 μM 5-Aza-2-deoxycytidine (5-aza-dC) starting at 11 dpf. Larvae were examined at the end of treatment, at 40 dpf. Approximately 40% of 5-aza-dC-treated larvae showed distinctive phenotypic abnormalities, with a shortened tail, torsion of the spinal cord, head malformations and depigmentation (data not shown).

In order to verify the experimentally induced global inhibition of methylation with 5-aza-dC, samples from untreated and 5-aza-dC-treated fish were MSAP genotyped. The number of loci obtained for the primer combination used in this experiment was 423, of which 355 were classified as methylation-susceptible loci. The percentage of polymorphic methylation-susceptible loci was 63%, and the mean Shannon's diversity index was 0,598 ± 0,101. The proportions of the four methylation states (unmethylated, hemi-methylated, internal cytosine methylation and full methylation) are shown in Table 2. The major difference between groups, 25%, was detected for the unmethylated state, the 5-aza-dC-treated fish showing 41.6% demethylation and the control fish 16.6%. The percentage internal cytosine methylation and full methylation states were higher in the control group, and small difference was seen for the hemi-methylated state. The differences between groups in genome-wide methylation were statistically significant (AMOVA; Φ_{st}= 0.5232, $p < 0.0001$). In the principal coordinates analysis, the control

group was clearly separated from the 5-aza-dC-treated group along the first coordinate (44.1% of variance explained) (Fig. 2). The two groups were also differentiated along the second coordinate, with 10.1% of variance explained. These data support the use of 5-Aza-2-deoxycytidine as a demethylating agent in zebrafish larvae.

Global inhibition of methylation with 5-Aza-2-deoxycytidine promoted *sparc* transcriptional activation in association with decreasing CpG methylation.

To determine more specifically the methylation status of the CpG sites in the +6634 to +6905 nt region upstream of the *sparc* translation starting site and included in first intron (i1-2) (Fig. 3A), bisulfite-treated DNA from control and 5-aza-dC-treated fish was sequenced.

Exposure to 5-aza-dC decreased *sparc* CpG island methylation from 57.3±3.48 to 22.4±5.17 % in control and treated fish, respectively ($t = 4.153$, $p = 0.001$) (Fig. 3B). Significant differences in eight of nine positions were detected between treated and untreated fish, positions +6661 and +6705 showing the most significant differences ($t = 2.381$, $p = 0.004$ and $t = 4.394$; $p = 0.003$, respectively). Significant differences in *sparc* gene expression were also seen according to treatment (Fig.1S). *Sparc* levels were increased in fish treated with 5-aza-dC by up to threefold in respect of control fish ($t = -4.86$; $p = 0.001$) (Fig. 1S). Therefore, global inhibition of methylation with 5-aza-dC promoted *sparc* expression in association with decreasing CpG methylation.

Methylation of whole plasmid containing the *sparc 7,25kb* upstream promoter driving the *eGFP* reporter gene suppresses transcription.

As described before, the major activity of the *sparc* promoter seems to be contained within the 5'UTR-intron sequence, lying between +126bp to +7168 bp from the transcription start site. Therefore, the P2 construct was methylated in vitro by Sss I methylase which specifically methylates the cytosine in the dinucleotide sequence 5'-CG-3'. This methylation pattern closely mimics that found in the genomic DNA. In vitro methylation of the *Tol2-7,25kb-sparc-eGFP* (P2 construct) containing the 0.25-kb *Sparc* promoter and its 5'-flanking sequence 7 kb upstream of the translated exon 2 resulted in an approximately 4-fold decrease in *eGFP* fluorescence (Fig. 4A,B,E). To avoid the possible interfering methylation of backbone CpGs, the CpG free *mustn1b* promoter (data non-published) was ligated into the *pToleGFPDest* vector (Lawson Lab # 274), SssI methylated and microinjected. Zebrafish *mustn1b* (Mustang 1b, musculoskeletal temporally activated novel gene) promoter specifically drives the *eGFP* expression in the

skeletal muscle. In vitro methylation of the *Tol2-0.75Kb-mustn1b-eGFP* construct containing the 0.75-kb CpG *free mustn1b* promoter had no effect in the *mustn1b* promoter activity (Fig. 4A,B,E). Suggesting therefore, that complete methylation of the vector backbone, had a negligible effect on transcription from both promoters (Fig. 4).

DISCUSSION

SPARC is a highly conserved extracellular matrix protein, which is actively involved, in many cellular processes including development, wound healing, angiogenesis, tumorogenesis and inflammation. This critical role of SPARC in a variety of different biological processes imposes a tight control of its transcriptional regulation. Although isolation of *sparc* promoter regions from different species has been reported (McVey et al., 1988; Young et al., 1989; Damjanovski et al., 1998) and numerous factors appear to be involved in its transcriptional regulation, detailed information on the molecular mechanisms regulating *sparc* gene activity is still lacking, and a consistent promoter study has not yet been performed in non-mammalian vertebrates.

In this study, we investigated *sparc* gene expression and the regulatory elements required for its expression in particular during early embryogenesis by using transient expression analyses in zebrafish embryos. Comparative molecular analysis of *sparc* promoter and its 5'-flanking sequence between zebrafish and other vertebrate species showed no nucleotide homology at the 5' ends (Damjanovski et al., 1998). However, a number of similarities in their overall organization were found (Young et al., 1989). Thus, the molecular organization of the first and second exons is conserved. Exon I, containing 125 bp in zebrafish, represents the majority of the 5' untranslated region in zebrafish and other vertebrate species (McVey et al., 1988; Young et al., 1989; Damjanovski et al., 1998), while exon II, which comprises 70bp in zebrafish, contains the remainder of the 5' untranslated region and encodes the entire signal peptide like other vertebrates species. Another characteristic found in zebrafish and other vertebrate species is the presence of the 5'CCTG3' motif in the *sparc* promoter and its 5'-flanking sequence. The function of this conserved sequence has been shown to be important in either regulation on the gene or the stability of RNA (McVey et al., 1988). A common characteristic of *sparc* gene organization in all vertebrates species studied is the presence of an intronic sequence between the first non-coding and the second coding exon. However, the size of the first intron seems to be species-specific, being 7 kb in zebrafish, 10 kb in humans and 2 kb in *Xenopus* (Damjanovski et al., 1998). Similar to mammalian vertebrates, the promoter of *sparc* gene in zebrafish lacks the

classical CAAT and TATA box motifs found in many eukaryotic promoters. We found that the 0.2-kb *sparc* promoter and its 5' flanking sequence 7 kb upstream of the translated exon drive *eGFP* expression in the notochord, otic vesicle, pharyngeal region, fin fold and intermediate cell mass, which mimicked the already well described expression pattern of the endogenous *sparc* mRNAs (Laize et al., 2005; Rotllant et al., 2008). Similar results were found in mice, where *sparc* transcripts were detected in developing tissues, such as the otic vesicle (Mothe et al., 2001), notochord and embryonic skeleton (Mason et al., 1986; Holland et al., 1987). In addition, the 0.2-kb *sparc* promoter and its 5'-flanking sequence 7 kb upstream of the translated exon drove the *eGFP* expression in the olfactory placode, specifically in the olfactory sensory neurons. Although specific expression of *sparc* in the olfactory bulb of mice has already been reported (Mendis et al., 1994), this is the first demonstration of the expression of *sparc* in the olfactory sensory neurons of the olfactory epithelium in non-mammalian vertebrate. We were unable to detect *sparc* expression in the olfactory placode by whole-mount in-situ hybridization. One possible explanation may be the limited sensitivity of our assay to detect faint expression of *sparc* in some regions. It should also be noted that, although the conclusion was based on transient expression analysis, it is unlikely that the tissue-specific spatial expression pattern of the *eGFP* was due to a position effect of the integration site, because the pattern of *eGFP* expression mimicked endogenous *sparc* expression in many ways. However, we cannot exclude the possibility that there might be a position effect on the activity of the promoter, which might explain the specific *eGFP* expression in the olfactory sensory neurons in mosaic fish.

Transient expression analyses in zebrafish embryos demonstrated that promoter activity resides in the unique 5′-UTR intronic region (nt+126/+7168). Specific deletion of this region resulted in a complete reduction of promoter activity. Transcriptional regulation of other genes (such as ubiquitine C) has also been shown to be exclusively regulated by the 5'UTR intron sequence (Bianchi et al., 2009). Therefore, the 5′-UTR unique intronic region (nt+126/+7168) provides the regulatory elements required for expression of a reporter gene in a subset of tissues that normally express the endogenous *sparc* gene in zebrafish embryos.

Sequence analyses of the zebrafish *sparc* 5′-UTR intron 1 region revealed a number of transcription factor binding sites. Because of the similarities in the overall organization of the *sparc* promoter and *sparc* expression domains and the highly conserved amino acid sequence in diverse vertebrate species (Laize et al., 2005), we compared conservation of *cis*-acting genetic elements that regulate *sparc* expression in the zebrafish 5′-UTR intron 1 sequence. Several

transcription factors in zebrafish were common to other *sparc* promoter sequences, including heat shock elements, cAMP- responsive element binding proteins, myoblast factors, gata binding factors, activating protein 1, retinoic acid receptor and glucocorticoid elements. All these factors have been shown to regulate *sparc* expression *in vitro* (Brekken and Sage, 2000). Additionally, transcription factor binding sites belonging to the Sox family were identified in the zebrafish 5′-UTR intron 1 region. This finding is in agreement with several other studies showing the role of SOX elements in *sparc* transcriptional regulation (Rotllant et al., 2008; Huang et al., 2008).

A CpG-rich sequence (CpG island) was also identified in the zebrafish 5′-UTR intron 1 region. It has been shown that *sparc* is transcriptionally regulated by DNA methylation, and CpG-rich sequences were also identified in mammalian *SPARC* promoter sequences (Rodriguez-Jimenez et al., 2007; Gao et al., 2010; Tajerian et al., 2011). In order to obtain insights into the transcriptional regulation of *sparc* expression, we investigated the role of DNA methylation in the expression of *sparc* in zebrafish embryos. 5-aza-dC was used to artificially induce DNA hypomethylation. This method has already been used to induce aberrant DNA hypomethylation in zebrafish embryos (Martin et al., 1999; Christman, 2002). Our results show that (i) exposure to 5-aza-dC produces distinctive phenotypic abnormalities in zebrafish larvae, including shortened tail, torsion of spinal cord, head malformations and depigmentation (data not shown); (ii) exposure to 5-aza-dC produces significant global DNA demethylation in zebrafish larvae; (iii) exposure to 5-aza-dC specifically reduced CpG-rich sequence (CpG island) methylation in the zebrafish *sparc* 5′-UTR intron 1 region; and (iv) *sparc* is highly expressed in 5-aza-dC treated zebrafish larvae. Additionally, we also show that in vitro methylation of the *sparc* gene promoter is capable of attenuating transcription in zebrafish embryos. Taken together, these data identify a contributory role for DNA methylation in regulating *sparc* expression in zebrafish embryogenesis.

In summary, our study provides the first evidence that the 5′-UTR intron of zebrafish *sparc* gene contains the functional and regulatory elements required for its expression, in particular during early embryogenesis. We also provide evidence that *sparc* is transcriptionally regulated by DNA methylation. Our findings should provide a basis for further studies to characterize critical regulatory elements and to elucidate the molecular mechanisms underlying transcriptional regulation of the *sparc* gene under both physiological and pathological conditions.

EXPERIMENTAL PROCEDURES

Experimental animals

Zebrafish embryos were cultured as previously described (Westerfield, 2007) and staged by standard criteria Kimmel et al., 1995) or by hours (hpf) or days (dpf) post fertilization. Experiments were performed with the TU (Tuebingen) wild-type strain (Nüsslein-Volhard Laboratory). To inhibit embryo pigmentation, embryo medium was supplemented with 0.003% (w/v) 2-phenylthiourea (Westerfield, 2007). For histology, dechorionated embryos were fixed overnight at 4 °C in 4% paraformaldehyde in 1XPBS, washed in PBS, and either stored at 4 °C in 1XPBS for confocal imaging or dehydrated through a methanol series and stored at -20 °C in 100% methanol for in situ hybridization. Ethical approval for all studies was obtained from the Institutional Animal Care and Use Committee of the IIM-CSIC Institute in accordance with the National Advisory Committee for Laboratory Animal Research Guidelines licensed by the Spanish Authority (1201/2005) and conformed to European animal directive (86/609/EEC) for the protection of experimental animals.

Determination of transcription start site by 5-RACE

RACE was carried out to determine the transcription start site according to the instructions provided by SMART RACE cDNA Amplification Kit (BD Bioscience, Clontech Laboratory). 5'-RACE-Ready cDNA was amplified with the adapter primer (see manual protocol) and a 5' gene specific primer (5'- GCCAGCGAGGCAGAACAGGAAGAAG -3'). The polymerase chain reaction (PCR) product was subcloned into the pGEM-T easy vector for sequencing.

Construction of BAC and Tol2 transposon vectors

BAC Clone (number 97O1) from DanioKey zebrafish BAC library containing the *sparc* gene was purchased from Source BioScience, UK. The constructions of *Sparc-iTol2-eGFP* BAC DNA vector was carried out essentially as previously described (Suster et al., 2009). To generate the *Sparc- iTol2* BAC clone, we used recombineering in *Escherichia coli* to introduce the iTol2-amp cassette from *pCR8GW-iTol2-amp* plasmid (Suster et al., 2009) in the BAC plasmid backbone, which contains the inverted minimal *cis*-sequences required for *Tol2* transposition. The insertion of the reporter gene *eGFP* into the *sparc* locus of *sparc-iTol2* BAC clone was carried out using homologous recombination. Briefly, *eGFP-kan* reporter gene cassette from

linear double-stranded eGFP-*pA-FRT-kan-FRT* plasmid was amplified by PCR, together with 50-bp homologies to the *sparc* translation start site. After transformation of *sparc-iTol2* BAC clone-containing cells with the PCR product, homologous recombination took place between the PCR product and the *sparc-iTol2* BAC clone resulting in integration of the vector insert into the BAC clone, placing the *eGFP* reporter gene under control of the *sparc* promoter within the clone (*sparc-iTol2-eGFP-BAC) (P1)*(Figure.1B).

The zebrafish *sparc* promoter sequences (GenBank accession number: BX640507) for the *Tol2* transposon constructs used in this study were amplified from DanioKey zebrafish genomic BAC library clone number 97O1 (Source BioScience, UK) with KOD Xtreme^{tm} Hot Start DNA Polymerase Platinum (Novagen, 71975). The PCR conditions were: 94 ºC for 2 min and 40 cycles with a 10-s denaturation step at 98 ºC, 55 ºC for 30 s and 68 ºC for 10 min. The primers used were 5'AAGCTTAGCACAATAGGATG -3' and 5'-TTTTGCTTAGGCTGAAACTCAAG-3'. The agarose band was extracted and purified with the QIAquick® Gel Extraction Kit. The PCR product was diluted 1:10, and 1 µl was ligated into 1 µl of P-ENTR^{Tm} /D-TOPO$^{®}$ Cloning Kit (Invitrogen) and transformed according to the protocol. Both promoter sequences were then ligated by LR clonase into the destination vector *pToleGFPDest* (Lawson Lab # 274) containing the *eGFP* sequence downstream, and sequenced. Construct P2 (*7,25kb-sparc-Tol2-eGFP)* contains 127 nt upstream to the transcription start (herein referred to as the proximal promoter, PP), the 125 nt of exon 1, the 7043 nt of intron 1 and the 13 nt of exon 2, excluding the translation starting site (+7182) (Figure. 1C). Construct P3 (*0,25kb-sparc-Tol2-eGFP*) was similar to construct P2, except that it lacked the UTR intron (+126/+7168) (Figure 1C). *Cis*-Acting transcription factor binding sites located in the zebrafish *sparc* promoter sequence were identified with MatInspector software (Cartharius et al., 2005). CpG island predictor analyses were done with MethPrimer software (Li et al., 2002).

Analysis of *eGFP* reporter gene expression

BAC construct DNA and *Tol2* constructs were dissolved in distilled RNAase free H2O to a final concentration of 200 ng/µL and 50 µg/ml respectively. Approximately 2 nl of DNA solution with transposase mRNA (200 ng/µL and 60 µg/ml respectively) were microinjected into the cytoplasm of zebrafish embryos at the one- or two-cell stage. Microinjection was carried out under a dissection microscope (MZ8, Leica) fitted with a MPPI-2 pressure injector (ASI

systems). *eGFP* expression in the injected embryos was analysed at 19, 24, 48 and 72 h post fertilization (hpf). *eGFP* expression was analysed by direct observation of *eGFP* expression under a fluorescence stereoscope and a confocal microscope. The number of embryos showing *eGFP* fluorescence was determined and the BAC and Tol2 vector constructs were compared to score activity and tissue specificity.

In situ hybridization

Whole-mount in situ hybridization was performed with digoxigenin-labelled antisense probes as previously described (Rotllant et al., 2008). Antisense riboprobes were made from linearized full-length *Danio rerio sparc* cDNA (GenBank Accession number: BC071436) (primers: forward 5'-TGCTTAGGCTGAAACTCAAGATGAG-3'; reverse 5'-GCATCAATGGAAGACGTCCTTAGAT-3').

In vitro methylation reactions

Whole reporter gene plasmids were in vitro methylated using Sss1 methylase (New England BioLabs) following the manufacturer's protocol. Sss1 methylation was performed with 10 mM Tris, pH 7.9, 50 mM NaCl, 10 mM MgCl2, 1 mM DTT, and 160 mM S-adenosylmethionine at 37°C for 1 h. Successful vector methylation was checked by analyzing band patterns on gel electrophoresis after digestion of the purified plasmids with the McrBC enzyme, which digests only methylated DNA, according to the manufacturer's instructions (New England BioLabs). The fully methylated and non-methylated plasmids were microinjected into the cytoplasm of zebrafish embryos at one- or two-cell stage following the above protocol. To visualize *eGFP* fluorescence in the developing embryos, 19 and/or 22hpf embryos microinjected with fully methylated and non-methylated plasmids were placed in a glass-bottomed dish and observed. We used a Leica M165 FC stereomicroscope equipped with a DFC310FX camera (Leica). Fluorescence images of live embryos were captured and analyzed using Leica Application Suite (LAS) software. The fluorescence intensity in all embryos was measured at the indicated stages using ImageJ software (National Institute of Heatlh, NIH) and displayed as vertical bar chart (Burgess et al., 2010). To exclude the possible effect of vector backbone CpG-methylation on the eGFP reporter gene expression, a CpG-free promoter (0.7Kb *mustn1b* promoter, unpublished data) was clone into the destination vector *pToleGFPDest* (Lawson Lab # 274) containing the

388 *eGFP* sequence downstream and used as a positive control in the transient expression
389 experiments.

390 **5-Aza-2'-deoxycytidine treatment**

391 5-Aza-2'-deoxycytidine (5-aza-dC) was purchased from Sigma Aldrich, Spain. A final
392 concentration of 50 µM was used (Martin et al., 1999). Embryos were treated for 30 days with 5-
393 aza-dC with the fish water (Westerfield, 1997) being replaced every 24 h with fresh fish water
394 containing 5-aza-dC. The application was initiated at 11 dpf and finished at 40 dpf.
395 5'Azacytidine-2-deoxycytidine it is a cytidine analog, which it is not known to produce any
396 general cytotoxic effects (Martin et al., 1999). 5-Aza-2'-deoxycytidine was previously diluted
397 with acetic acid: water (1:1).

398 **Genomic DNA and RNA extraction**
399 DNA and total RNA were extracted at 40 dpf from control and 5-aza-dC-treated zebrafish larvae.
400 Additionally, individuals for each treatment were collected, anaesthetized with MS-222 (Sigma–
401 Aldrich, Madrid, Spain), and photographs were taken with a Leica DFC310 FX camera and
402 Leica M165FC stereomicroscope.
403 gDNA was extracted with the NucleoSpin® Tissue Kit (BD Biosciences). DNA quality was
404 verified by electrophoresis on 1% agarose gels. Total RNA was extracted from zebrafish larvae
405 with TrizolReagent (Ambion), and first-strand cDNA was synthesized according to the Maxima
406 First Strand cDNA Synthesis Kit (Fermentas) protocol with 1 µg RNA.
407

408 **Methylation-sensitive amplification polymorphism (MSAP) genotyping**
409 A modification of the MSAP method described by (Reyna-Lopez t al., 1997) and (Xu et al.,
410 2000) was used. In brief, genomic DNA was digested with two methylation-sensitive
411 isoschizomers (MspI and HpaII) as frequent cutters, each in combination with the same rare
412 cutter (EcoRI) in parallel batches, ligation of adaptors and selective PCR amplification with
413 primers complementary to the adaptors but with unique 3' overhangs. The two isoschizomers
414 recognize the same sequence (5'-CCGG) but differ in their sensitivity to DNA methylation.
415 Comparison of the two profiles for each individual allowed assessment of the methylation state
416 of the restriction sites. Methylated CpG are restricted by MspI, and hemiMethylated CpCpG sites
417 are restricted by HpaII (REBASE). Sites that are hypermethylated (i.e. at both the internal and

external Cs) and sites that are fully methylated at the external Cs (i.e. on both strands) are not cut by either enzyme, whereas sites that are free from methylation are restricted by both.

Two primer combinations (EcoRI-AAG- HpaII -TC, EcoRI -ACT- HpaII -TC) were used for selective amplifications. Primer sequences and PCR details are available in (Moran and Perez-Figueroa, 2011) HpaII primers were end-labelled with a 6-FAM reporter molecule. PCR products were loaded simultaneously with a GeneScan 500 ROX size standard into an ABI Prism 310 Genetic Analyzer (Applied Biosystems). Fragment analysis and scoring was performed with GeneMapper v.3.7 software (Applied Biosystems). DNA fragments shorter than 100 bp, longer than 500 bp or less than 70 relative fluorescent units were excluded from the analysis.

Methylation analyses of *sparc* CpG island by bisulfite-mediated genomic sequencing

One microgram of genomic DNA was used for bisulfited DNA conversion, according to the manufacturer's protocol (EZ DNA Methylation-Direct™ Kit, Zymo Research). Primers were designed to amplify a 300-bp fragment in the *sparc* intron 1, where a CpG island was detected with MethPrimer software. Primer F: 5'-AATTAAAAGGAAGAGAGATTTTTGG-3' and primer R: 5' -TCAAACCACCAAACCTACTCTA-3' were used. Two microlitres of bisulfited DNA were taken for the PCR reaction. DreamTaq MasterMix (Fermentas) was used to amplify the fragment. Bands were gel-purified and cloned into PGem-Teasy (Promega). Five individuals and 10 clones of each individual per treatment were taken and sequenced with SP6. A total of 100 clones were analyzed. Using the program BDPC DNA methylation analysis platform (available at: http://biochem.jacobs-university.de/BDPC/) (Rohde et al., 2010), different methylation levels of CpG dinucleotides were computed.

Quantification of *sparc* gene expression by real-time RT-PCR

cDNA was made from total RNA with superscript III (Invitrogen) according to the manufacturer's recommendations. The primers designed to detect *sparc* transcripts were: OsteoRT (F): 5'-CCCTCTGCGTGCTCCTCTTA-3' and OsteoRT (R): 5'-GCATCGCACTGCTCAAAGAA-3'. Expression levels were standardized to *18S* by the $2^{-\Delta\Delta T}$ method (Livak and Schmittgen, 2001). Dilutions of 1:10 of cDNAs were made for quantifying the number of *sparc* transcripts. Real-time quantitative PCR (qPCR) reactions were performed in an AB 7300 real-time PCR System (Applied Biosystems) and incorporation of Maxima® SYBR Green/ROX qPCR Master Mix (2X) (Fermentas). The two-step cycling conditions for the two

primer sets were: denaturation at 95 ºC for 10 min, followed by 40 cycles at 95 ºC for 15 s and 60 ºC for 1 min. All samples were done in triplicate. Finally, a melting-curve analysis was carried out at 95 ºC for 15 s, 60 ºC for 30 s and 95 ºC for 15 s for testing the specificity of the primers.

Statistical analysis

MSAP profiles, pooled from both primer combinations, were analysed using the R package MSAP ((Perez-Figueroa, 2013), http://msap.r-forge.r-project.org). We scored the MSAP fragments as follows: fragments present in both *Eco*RI-*Hpa*II and *Eco*RI-*Msp*I products (1/1), denoting a non-methylated state; those fragments present only in either *Eco*RI-*Hpa*II (1/0) or *Eco*RI-*MspI* (0/1) products, corresponding to a methylated state; or absent from both *Eco*RI-*Hpa*II and *Eco*RI-*Msp*I products (0/0), which we considered as an hyper-methylation of the target. Individual fragments (loci) were, therefore, classified into 'methylation-susceptible loci' (MSL) if the observed proportion of methylated scores (1/0, 0/1 and 0/0) exceeded a 5%, and "methylation-susceptible fragments" if the methylated state was the dominant marker (1 for the methylated state and 0 for the non-methylated state). Epigenetic differentiation among treatmets was assessed by means of principal coordinates analysis (PCoA) followed by analyses of molecular variance (AMOVA; (Excoffier et al., 1992)). Student's t test was used to test differences in methylation levels on *sparc* CpG islands and also to test the significant significance of differences in *sparc* levels between treatments. Statistical software of Statistical Package for the Social Sciences (IBM SPSS Statistics 2.1) was used for statistical analyses. Differences were considered statistically significant at $p < 0,05$.

Acknowledgements

We thank Prof. M. Suster and Prof. K. Kawakami for providing the iTol2 constructs. This work was funded by the Spanish Science and Innovation Ministry project AGL2008-00392/ACU and ALG2011-23581 to JR. Partial funding was also obtained from Science and Innovation Ministry projects BFU2010-14839, CSD2007-00008, and P08-CVI-3488 to JLGS. Eva Torres-Nuñez was supported by PhD fellowship FPI BES-2009-016797.

REFERENCES

Bianchi M, Crinelli R, Giacomini E, Carloni E, Magnani M. 2009. A potent enhancer element in the 5'-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene. *Gene*. **448**(1): 88-101.

Brekken RA, Sage EH. 2000. SPARC, a matricellular protein: at the crossroads of cell-matrix. *Matrix Biol*. **19**: 569-580.

Burgess A, Vigneron S, Brioudes E, Labbé J-C, Lorca T & Castro A. 2010. Loss of human Greatwall resultsin G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proc Natl Acad Sci USA*. **107:** 12564–12569

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. **21**: 2933-42.

Ceinos RM, Torres-Núñez E, Chamorro R, Novoa B, Figueras A, Ruane NM, Rotllant J. 2013. Critical role of the matricellular protein SPARC in mediating erythroid progenitor cell development in zebrafish. *Cells Tissues Organs*. **197**(3): 196-208.

Cheetham S, Tang MJ, Mesak F, Kennecke H, Owen D, Tai IT. 2008. SPARC promoter hypermethylation in colorectal cancers can be reversed by 5-Aza-2'-deoxycytidine to increase SPARC expression and improve therapy response. *Br J Cancer*. **98** (11): 1810-1819.

Christman JK. 2002. 5-Azacytidine and 5'-aza-2'deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*. **21**(35): 5483-95.

Damjanovski S, Huynh MH, Motamed K, Sage EH, Ringuette M. 1998. Regulation of SPARC expression during early Xenopus development: evolutionary divergence and conservation of DNA regulatory elements between amphibians and mammals. *Dev Genes Evol.* **207**: 453-461.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. **131**(2): 479-91.

509 Gao J, Song J, Huang H, Li Z, Du Y, Cao J, Li M, Lv S, Lin H, Gong Y. 2010. Methylation
510 of the SPARC gene promoter and its clinical implication in pancreatic cancer. *J Exp Clin*
511 *Cancer Res*. **29**:28.

512 Holland PW, Harper SJ, McVey JH, Hogan BL. 1987. In vivo expression of mRNA for the
513 Ca++-binding protein SPARC (osteonectin) revealed by in situ hybridization. *J Cell Biol*.
514 **105**(1): 473-82.

515 Huang DY, Lin YT, Jan PS, Hwang YC, Liang ST, Peng Y, Huang CYF, Wu HC, Lin CT.
516 2008. Transcription factor SOX-5 enhances nasopharyngeal carcinoma progression by down-
517 regulating SPARC gene expression. *J Pathol*. **214**: 445-455.

518 Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic
519 development of the zebrafish. *Dev Dyn*. **203** (3): 253–310.

520 Laizé V, Pombinho AR, Cancela ML. 2005. Characterization of *Sparus aurata* osteonectin
521 cDNA and in *silico* analysis of protein conserved features: Evidence for more than one
522 osteonectin in Salmonidae. *Biochimie.* **87**: 411-420.

523 Li LC, Dahiya R. 2002. MethPrimer: designing primers for methylation PCRs.
524 *Bioinformatics.* **18**(11): 1427-31.

525 Livak KJ, Schmittgen TD. 2001. Analysis of relative g ene expression data using real-time
526 quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods.* **25**(4): 402-8.

527 Mason IJ, Murphy D, Münke M, Francke U, Elliott RW, Hogan BL. 1986. Developmental
528 and transformation-sensitive expression of the Sparc gene on mouse chromosome 11. *EMBO*
529 *J*. **5**(8): 1831–1837.

530 Martin CC, Laforest L, Akimenko MA, Ekker M. 1999. A role of DNA methylation in
531 gastrulation and somite patterning. *Dev Biol*. **206**: 189-205.

532 Mendis DB, Brown IR. 1994. Expression of the gene encoding the extracellular matrix
533 glycoprotein SPARC in the developing and adult mouse brain. *Brain Res Mol Brain Res*. **24**
534 (1-4): 11-19.

535 Morán, P. and Pérez-Figueroa, A. 2011. Methylation changes associated with early
536 maturation stages in the Atlantic salmon. *BMC Genetics*. **12**: 86.

Mothe AJ, Brown IR. 2001. Expression of mRNA encoding extracellular matrix glycoproteins SPARC and SC1 is temporally and spatially regulated in the developing cochlea of the rat inner ear. *Hear Res*. **155**(1-2): 161-74.

McVey JH, Nomura S, Kelly P, Mason IJ, Hogan BLM. 1988. Characterization of the mouse SPARC/Osteonectin Gene. *J Biol Chem.* **263** (23): 11111-11116.

Pérez- Figueroa, A. 2013. *msap*: a tool for the statistical analysis of methylation- sensitive amplified polymorphism data. *Mol Ecol Resour*. **13**: 522- 527.

Reyna-López E, Simpson J, Ruíz-Herrera J. 1997. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Mol Gen Genet*. **253**(6): 703-710.

Rodríguez-Jiménez FJ, Caldés T, Iniesta P, Vidart JA, García-Asenjo JL, Benito M. 2007. Overexpression of SPARC protein contrasts with its transcriptional silencing by aberrant hypermethylation of SPARC CpG-rich region in endometrial carcinoma. *Oncol Rep*. **17**(6): 1301-1307.

Rohde C, Zhang Y, Reinhardt R, Jeltsch A. 2010. BISMA-Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*. **11**: 230.

Rotllant J, Liu D, Yan YL, Postlethwait JH, Westerfield M, Du SJ. 2008. Sparc (Osteonectin) functions in morphogenesis of the pharyngeal skeleton and inner ear. *Matrix Biol.* **27**(6): 561-72.

Socha MJ, Said N, Dai Y, Kwong J, Ramalingam P, Trieu V, Desai N, Mok SC, Motamed K. 2009. Aberrant promoter methylation of *SPARC* in ovarian cancer. *Neoplasia*. **11**(2): 126–135.

Suster ML, Sumiyama K, Kawakami K. 2009. Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* **10**:477.

Tajerian M, Alvarado S, Millecamps M, Dashwood T, Anderson KM, Haglund L, Ouellet J, Szyf M, Stone LS. 2011. DNA methylation of SPARC and chronic low back pain. *Mol Pain*. **7**: 65.

565 Torres-Núñez E, Sobrino C, Neale PJ, Ceinos RM, Du S. 2012. Rotllant J. Molecular
566 response to ultraviolet radiation exposure in fish embryos: implications for survival and
567 morphological development. *Photochem Photobiol*. **88** (3): 701-7.

568 Westerfield M. 2007. The Zebrafish Book: a guide for the laboratory use of zebrafish (*Danio*
569 *rerio*). (5th edition) *Univ. of Oregon Press*, Eugene.

570 Xu M, Li X. Korban SS. 2000. AFLP-Based Detection of DNA Methylation. *Plant Mol Biol*
571 *Rep*. **18**: 361-368.

572 Yang E, Kang HJ, Koh KH, Rhee H, Kim NK, Kim H. 2007. Frequent inactivation of
573 SPARC by promoter hypermethylation in colon cancers. *Int J Cancer*. **121**(3): 567-75.

574 Young MF, Finday DM, Dominguez P, Burbelo PD, McQuillan C, Kopp JB, Robey PG.
575 Termine JD. 1989. Osteonectin promoter. *J Biol Chem.* **264** (1): 450-456.

576

577

LEGEND TO THE FIGURES

Figure 1. *Sparc* transgene constructs and the *eGFP* expression in zebrafish. A) A schematic diagram of the zebrafish *sparc* locus shows the exons and approximately 7,25 kb of the genomic region upstream of *sparc*. B) *Sparc-iTol2-eGFP-BAC*: *eGFP* modified BAC showing reporter *eGFP* gene inserted before *sparc* translational start codon. C) *7,25kb-sparc-Tol2-eGFP*: Tol2 *eGFP* reporter gene construct contains a promoter region spanning nt −127 to the transcription start site (+1), the first exon, the unique intron of the 5′-UTR of the gene and a small part of the second exon. D) *0,25kb-sparc-Tol2-eGFP* is an *eGFP* reporter gene construct having a deletion of the non-coding intron 1-2 sequence. (E-I) *eGFP* expression in zebrafish carring a modified BAC construct (*Sparc-iTol2-eGFP-BAC)* and/or a Tol2 *eGFP* reporter gene construct (*7,25kb-sparc-Tol2-eGFP)*. Observed fluorescence from both construct was equivalent. Expression of *eGFP*(green) together with 4,6-diamidino-2-phenylindole (DAPI) (blue) was analyzed at the indicated developmental stages. (E) At 19 hpf, periodic expression of *eGFP* is detected in the notochord (arrows). At 30dpf, *eGFP* is expressed in a broad domain that contains most of the pharyngeal region (F), otic vesicle (F) (inset is a magnified image of the otic vesicle), fin fold (G) and intermediate cell mass (H). At 48hpf, *eGFP* is detected in the olfactory sensory neurons within the olfactory placode (I, inset is a magnified image of the olfactory placode, showing the olfactory sensory neurons and the olfactory epithelium). (J,K,L) Whole mount *in situ* hybridization shows *sparc* mRNA expressed in zebrafish notochord (periodic expression, arrows), pharyngeal region, otic vesicle, find fold and hematopoietic ICM region at 19, 25 and 40 hpf stages. Lateral views with anterior to the left and dorsal to the top. Abbreviations: E, exon; i, intron; n, notochord; pr, pharyngeal region; ov, otic vesicle; ff, fin fold; e, eye;op, olfactory placode; osn, olfactory sensory neurons; oe, olfactory epithelium; icm, intermediate cell mass; Scale bars: 100μm. Black boxes indicate untranslated regions; open boxes indicate coding exons; lollipops indicate CpG island detected.

Figure 2. Genoma-wide DNA methylation changes in control and 5-aza-dC-treated juvenile 40 dpf zebrafish. Principal Coordinates Analysis (PCoA) results for epigenetic differentiation between control and 5'-azacytidine-treated juvenile 40 dpf zebrafish. The two coordinates (C1 and C2) show the percentage variance in parentheses. (◆) represents control fish while (•) represents 5-aza-dC-treated larvae. Group's labels show the centroid for the points cloud in each group. Ellipses represent the dispersion of each group around its centre. The long axis of the

ellipse shows the direction of maximum dispersion and the short axis, the direction of minimum dispersion.

Figure 3. Methylation patterns of zebrafish *Sparc* putative promoter at 40dpf. A) A schematic depiction of the zebrafish *Sparc* gene, including the CpG island and the relative location of the PCR amplicon between positions −6634 and −6905. B) The sequence of the PCR amplicon, including the positions of the nine CpG sites. C) Resulting differences in zebrafish Sparc promoter DNA methylation in 40dpf according to 5-aza-dC treatment. On the left, one fish representative of the level of methylation is shown per each treatment. Open and filled circles denote unmethylated and methylated positions, respectively. Ten clones per fish were analyzed. Average methylation was calculated specifically for each CpG position(number below each column).Numbers with a positive sign indicate CpG positions in respect to the transcription starting site. On the right, percentage of methylated CpGs in both groups. Data are expressed as mean ± SEM. Statistically significant differences ($p<0.05$) are indicated by asterisks (*).

Figure 4. Effects of methylation on zebrafish *sparc* promoter activity *in vitro***.** Zebrafish embryos were microinjected with *7,25kb-sparc-Tol2-eGFP* or *0,7kb-Musnt1-Tol2-eGFP* methylated (B,D) and unmethylated (A,C) promoter vectors. Microinjected methylated and unmethylated groups were as follows: 1) zebrafish *7,5Kb Sparc promoter* cloned into *pTolEGFPDest* reporter plasmid (*7,25kb-sparc-Tol2-eGFP*) (A,B); 2) zebrafish *0,75Kb Mustnt1 promoter* cloned into *pTolEGFPDest* reporter plasmid (*0,7kb-Mustn1b-Tol2-eGFP*) (Ctrl, CpG-free Promoter) (C,D). E, intensity of *eGFP* fluorescent measurements were performed with Image J and displayed as vertical bar chart. Student *t* tests were performed to determine statistical relevance. Values are shown as mean ± S.E.M. (n = 10). E Inset: Successful vector methylation verification by analysis of band patterns on electrophoresis gel after digestion of the purified plasmids with the *McrBC* enzyme. Lane M, 1 Kb marker; lane 1 and 5 1 µg *7,25kb-sparc-Tol2-eGF*P or *0,7kb-mustn1b-Tol2-eGFP*; lane 2 and 6 1 µg *7,25kb-sparc-Tol2-eGF or 0,7kb-mustn1b-Tol2-eGFP* treated with *McrBC*; lane 3 and 7 1 µg *Sss*I-methylated *7,25kb-sparc-Tol2-eGFP or 0,7kb-mustn1b-Tol2-eGFP*; lane 4 and 8, 1 µg *Sss*I-methylated *7,25kb-sparc-Tol2-eGFP* or *0,7kb-mustn1b-Tol2-eGFP* treated with *McrBC*. As expected, only the methylated vector was digested. Scale bars: 250µm.

643

644 **Supplementary data**

645

646 **Figure 1S.** Expression of *sparc* in response to 5-aza-dC treatment. The relative expression of
647 *sparc* was determined by real-time PCR and standardized to 18S. The results are expressed as
648 mean ± SEM with respect to the control, which was set at 1. Statistically significant differences
649 ($p<0.05$) are indicated by asterisks (*).

650

651

652 **Table 1.** Percentage of each methylation state in control and 5-aza-dC groups

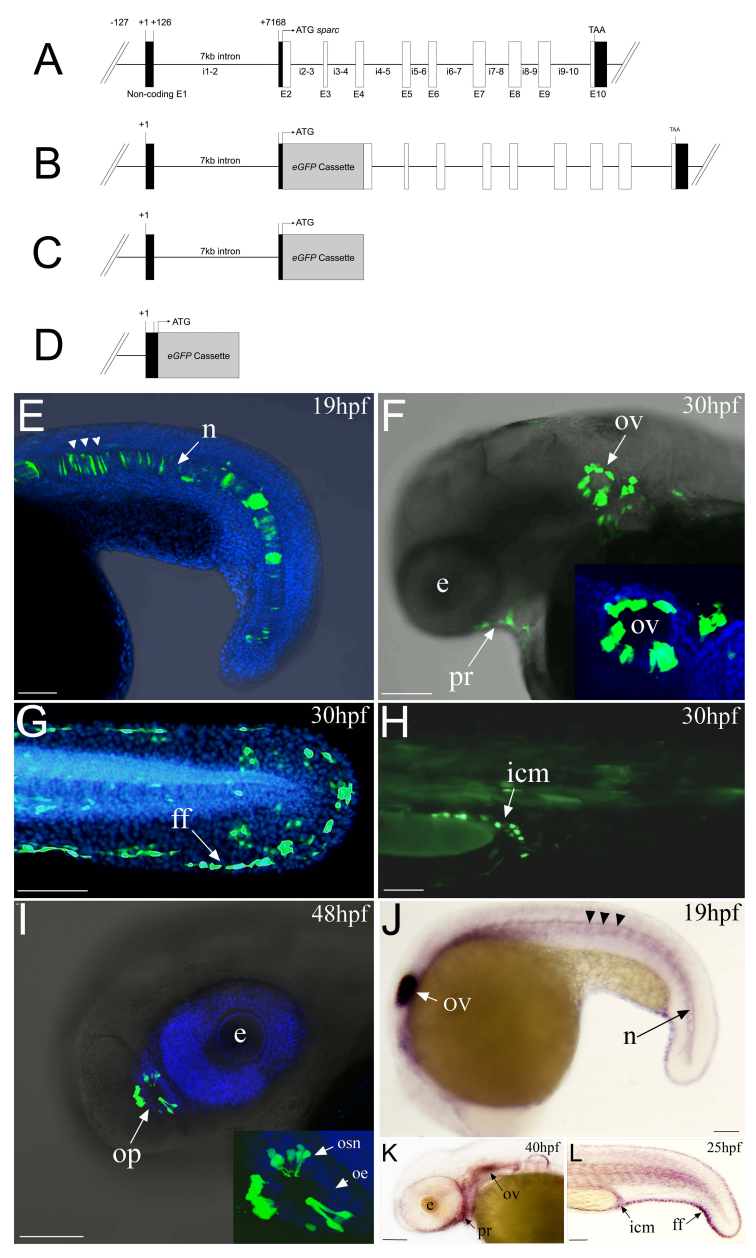
653

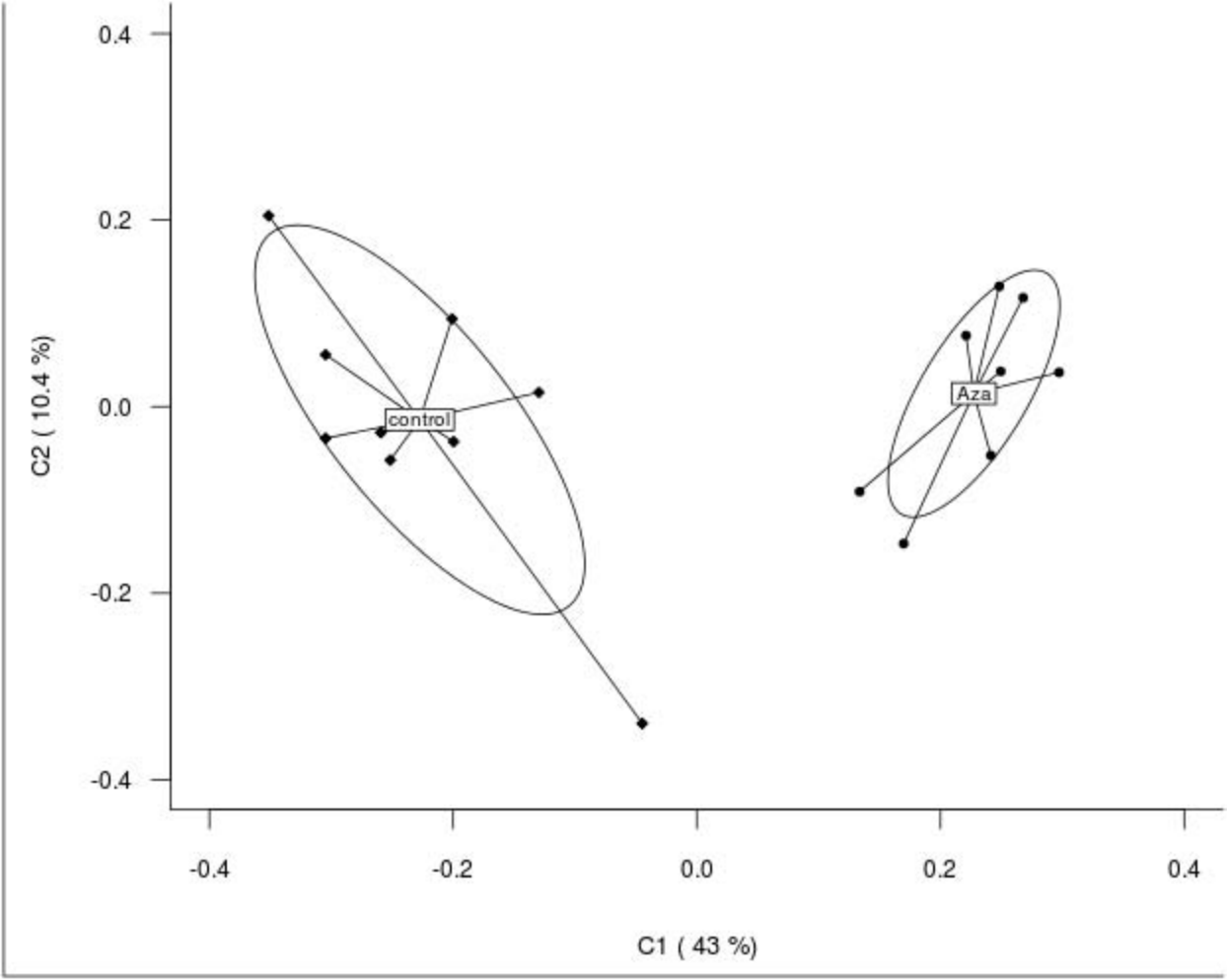
	HPA+/MSP+ (Unmethylated)	HPA+/MSP- (Hemi-methylated)	HPA-/MSP+ (Internal C methylation)	HPA-/MSP- (Full methylation)
Control	16.6	9.4	31	43
5-aza-dC	41.6	10.2	17.1	31.1

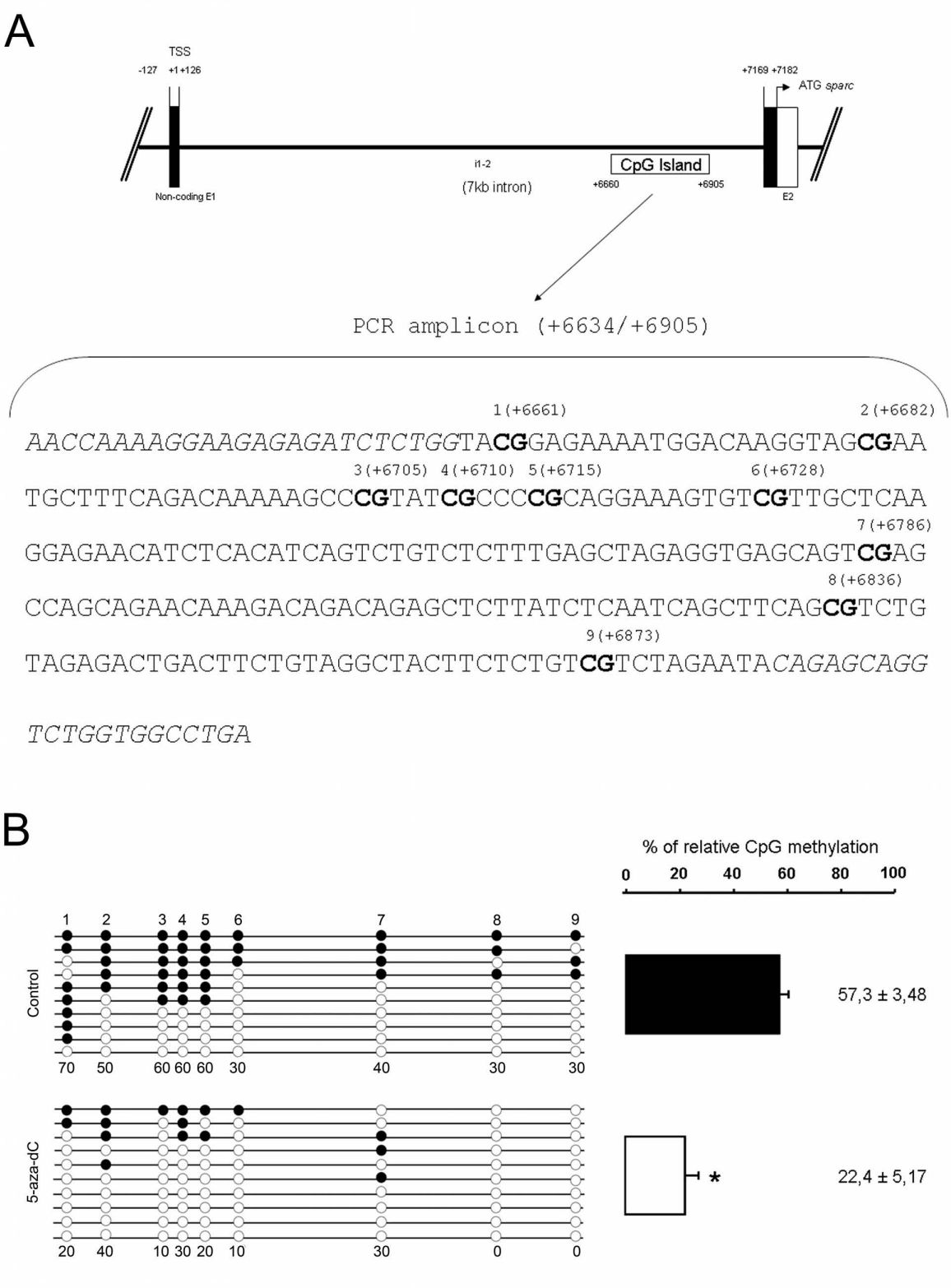
654

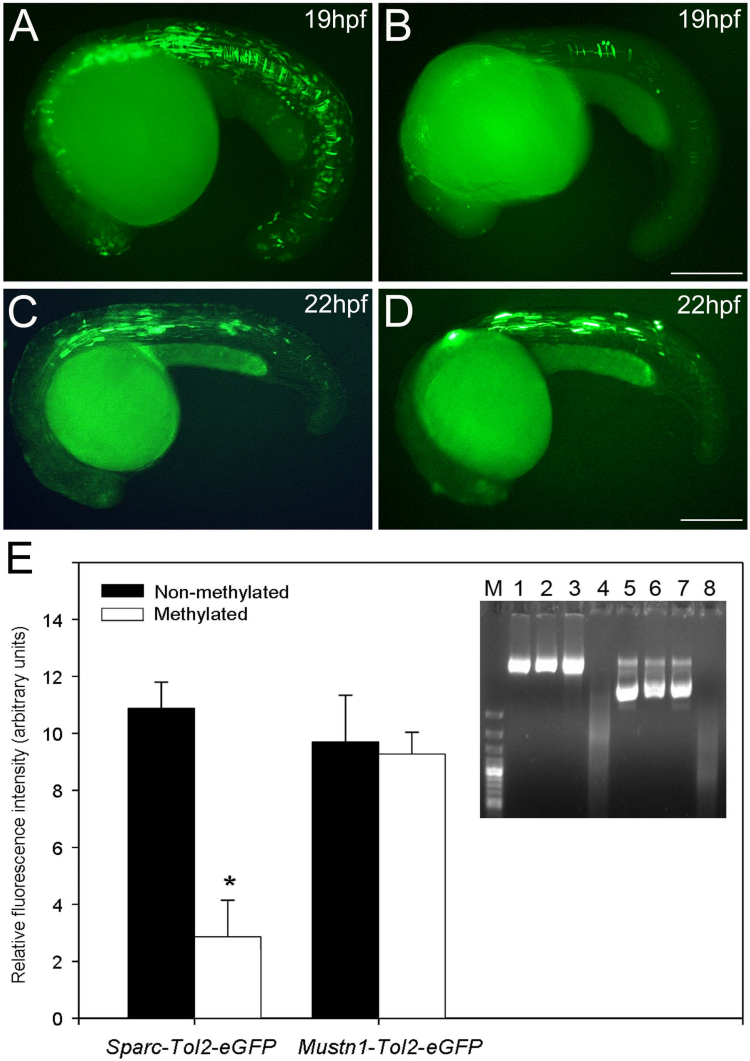
655

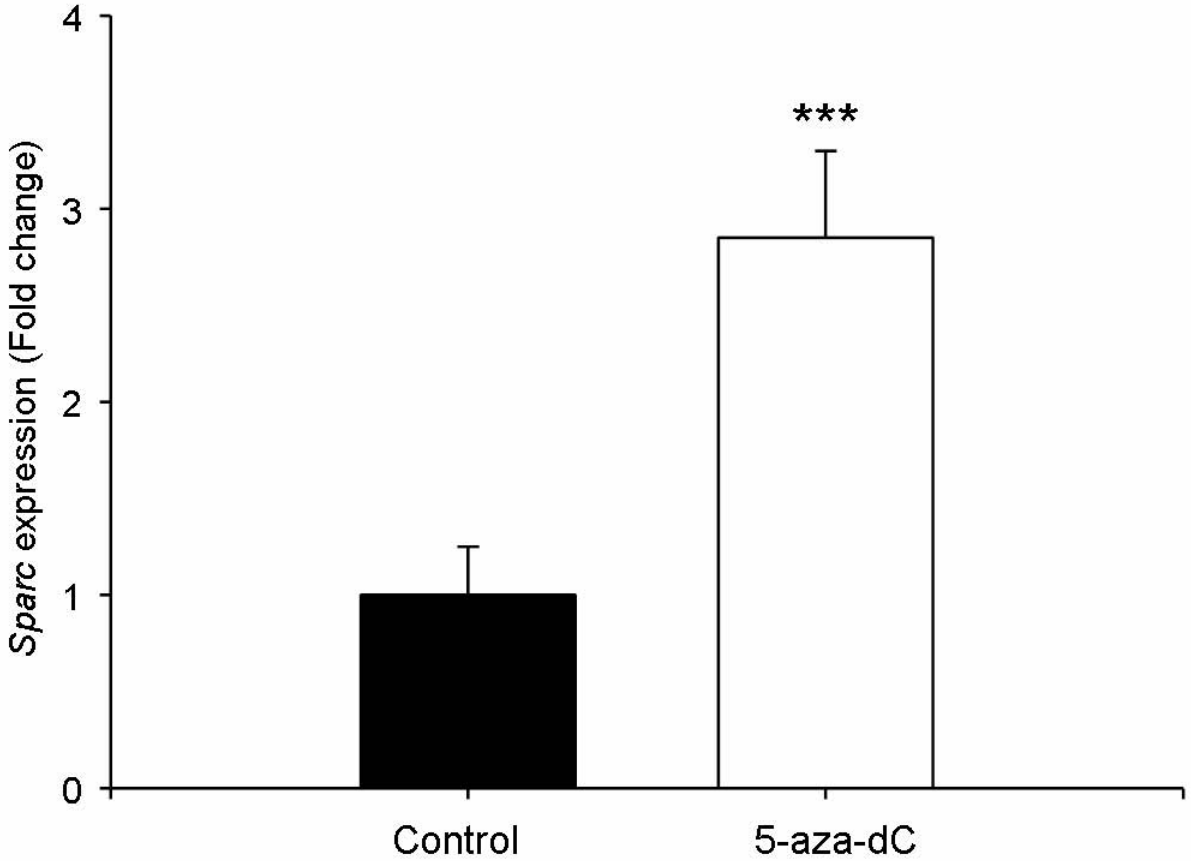
656


A
-127
+1 +126
+7168
ATG sparc
TAA
7kb intron
i1-2
i2-3
i3-4
i4-5
i5-6
i6-7
i7-8
i8-9
i9-10
Non-coding E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
B
+1
ATG
TAA
7kb intron
eGFP Cassette
C
+1
ATG
7kb intron
eGFP Cassette
D
+1
ATG
eGFP Cassette
E
19hpf
n
F
30hpf
ov
e
ov
pr
G
30hpf
ff
H
30hpf
icm
I
48hpf
e
op
osn
oe
J
19hpf
ov
n
K
40hpf
e
ov
pr
L
25hpf
icm
ff


0.4
0.2
0.0
-0.2
-0.4
C2 (10.4 %)
control
Aza
-0.4
-0.2
0.0
0.2
0.4
C1 (43 %)


A
TSS
-127 +1 +126
+7169 +7182
ATG sparc
i1-2
(7kb intron)
CpG Island
+6660
+6905
Non-coding E1
E2
PCR amplicon (+6634/+6905)
1(+6661)
2(+6682)
AACCAAAAGGAAGAGAGATCTCTGGTACGGAGAAAATGGACAAGGTAGCGAA
3(+6705) 4(+6710) 5(+6715)
6(+6728)
TGCTTTCAGACAAAAAGCCCGTATCGCCCGCAGGAAAGTGTCGTTGCTCAA
7(+6786)
GGAGAACATCTCACATCAGTCTGTCTCTTTGAGCTAGAGGTGAGCAGTCGAG
8(+6836)
CCAGCAGAACAAAGACAGACAGAGCTCTTATCTCAATCAGCTTCAGCGTCTG
9(+6873)
TAGAGACTGACTTCTGTAGGCTACTTCTCTGTCGTCTAGAATACAGAGCAGG
TCTGGTGGCCTGA
B
% of relative CpG methylation
0 20 40 60 80 100
1 2 3 4 5 6 7 8 9
Control
70 50 60 60 60 30 40 30 30
57,3 ± 3,48
5-aza-dC
20 40 10 30 20 10 30 0 0
*
22,4 ± 5,17


A
19hpf
B
19hpf
C
22hpf
D
22hpf
E
Non-methylated
Methylated
M 1 2 3 4 5 6 7 8
Relative fluorescence intensity (arbitrary units)
0
2
4
6
8
10
12
14
*
Sparc-Tol2-eGFP
Mustn1-Tol2-eGFP


4
3
2
1
0
Sparc expression (Fold change)
***
Control
5-aza-dC