

RESEARCH ARTICLE

Open Access

Regulation and expression of sexual differentiation factors in embryonic and extragonadal tissues of Atlantic salmon

Kristian R von Schalburg¹, Motoshige Yasuike¹, Ryosuke Yazawa², Johan G de Boer¹, Linda Reid¹, Stacy So¹, Adrienne Robb¹, Eric B Rondeau¹, Ruth B Phillips³, William S Davidson⁴, Ben F Koop^{1*}

Abstract

Background: The products of *cyp19*, *dax*, *foxl2*, *mis*, *sf1* and *sox9* have each been associated with sex-determining processes among vertebrates. We provide evidence for expression of these regulators very early in salmonid development and in tissues outside of the hypothalamic-pituitary-adrenal/gonadal (HPAG) axis. Although the function of these factors in sexual differentiation have been defined, their roles in early development before sexual fate decisions and in tissues beyond the brain or gonad are essentially unknown.

Results: Bacterial artificial chromosomes containing salmon *dax1* and *dax2*, *foxl2b* and *mis* were isolated and the regulatory regions that control their expression were characterized. Transposon integrations are implicated in the shaping of the *dax* and *foxl2* loci. Splice variants for *cyp19b1* and *mis* in both embryonic and adult tissues were detected and characterized. We found that *cyp19b1* transcripts are generated that contain 5'-untranslated regions of different lengths due to cryptic splicing of the 3'-end of intron 1. We also demonstrate that salmon *mis* transcripts can encode prodomain products that present different C-termini and terminate before translation of the MIS hormone. Regulatory differences in the expression of two distinct aromatases *cyp19a* and *cyp19b1* are exerted, despite transcription of their transactivators (ie; *dax1*, *foxl2*, *sf1*) occurring much earlier during embryonic development.

Conclusions: We report the embryonic and extragonadal expression of *dax*, *foxl2*, *mis* and other differentiation factors that indicate that they have functions that are more general and not restricted to steroidogenesis and gonadogenesis. Spliced *cyp19b1* and *mis* transcripts are generated that may provide regulatory controls for tissue- or development-specific activities. Selection of *cyp19b1* transcripts may be regulated by DAX-1, FOXL2 and SF-1 complexes that bind motifs in intron 1, or by signals within exon 2 that recruit splicing factors, or both. The potential translation of proteins bearing only the N-terminal MIS prodomain may modulate the functions of other TGF β family members in different tissues. The expression patterns of *dax1* early in salmon embryogenesis implicate its role as a lineage determination factor. Other roles for these factors during embryogenesis and outside the HPAG axis are discussed.

Background

Sex determination and early gonadal development depends upon specific embryonic gene programs that activate differentiation of the bipotential primodium. These processes hinge upon the expression of both sex-linked and autosomal genes that promote or antagonize

male or female cell fate pathways. The products of some genes are essential for female (WNT-4 (Wingless-related MMTV integration site-4) and follistatin) [1] and male (SRY (Sex-determining region of the Y chromosome) and FGF-9 (fibroblast growth factor-9)) [2] sex-determination in mammals. In mammals, as well as teleosts, DAX-1 (Dosage-sensitive sex reversal, adrenal hypoplasia congenital, critical region on the X-chromosome, gene-1), FOXL2 (forkhead box L2), cytochrome P450 aromatase, MIS (Mullerian inhibiting substance), SF-1 (steroidogenic

* Correspondence: bkoop@uvic.ca

¹Department of Biology, Centre for Biomedical Research, University of Victoria, Victoria, British Columbia, V8W 3N5, Canada
Full list of author information is available at the end of the article

factor-1), SOX-9 (SRY-related, high-mobility group (HMG) box (SRY box)-9), WT-1 (Wilm's tumour-1) [3-6] and many other gene products contribute in directing sex-determined fates, as well as to subsequent gonad development and function [7-9 and refs. therein].

SF-1 is central to the activity of many of these genes as well as a large component of steroidogenic genes. Both SF-1 and DAX-1 are orphan receptors for which no ligand has yet been unequivocally identified, but which do possess ligand-binding domains [10]. SF-1 and DAX-1 interact cooperatively or antagonistically to influence various target gene activities [10]. DAX-1 modulates transcription of genes such as aromatase [11] and MIS [12] through direct interactions with SF-1 and the recruitment of other factors. Functional interactions of SF-1 with WT-1 or DAX-1 promotes or inhibits the expression of MIS [13]. The products of *dax1*, *foxl2* and *sf1* each play critical roles in the transcriptional regulation of aromatase (*cyp19*) genes [6,11,14-17].

Elucidation of the factors that promote or repress the transcription of these differentiation factors will also help to characterize the regulatory programs they initiate. Conservation of transcription factor binding motifs in the promoters of mammalian and teleost genes point to shared regulatory regimes. The expression patterns of the genes described here may be deciphered in part by examination of their promoter regions. Antagonism between DAX-1 and FOXL2 or SF-1 and SOX factors may be common themes for activation or repression among many of these differentiation factors [11,18-23].

Although our understanding of the action of various sex determination and differentiation factors in early gonad development is well-established, our knowledge of other roles that these regulators may play before formation of sexual tissues is limited. Sex-specific regulation of various steroidogenic and differentiation regulators in trout have been demonstrated by molecular studies to occur about one month before observable morphological differences [5]. Histologically discernable features of gonad differentiation in salmonids become manifest approximately five weeks after hatching [5,14].

In the Atlantic salmon (*Salmo salar*), we have detected the expression of *dax1/2*, *foxl2a/b* and *mis* (and *sf1* and *sox9* factors) more than three weeks before even the hatching stage. In this paper we show that the genes we examined were expressed several months before morphologically discernable differentiation of the gonad. This is the first time, other than for *dax1*, that this has been demonstrated for the genes we selected. We also demonstrate broad expression of these regulators in various extragonadal adult tissues. Therefore genes previously considered to be involved in gonadogenesis and

steroidogenesis are apparently more general in their transcriptional programming, at least in salmon.

Furthermore, we have isolated, sequenced and assembled BACs containing these genes and analyzed their proximal promoters. Our expression studies also reveal unique splicing activities that occur among some of these regulators. Taken together, we suggest that the roles of *dax1/2*, *foxl2a/b*, *mis* and *sf1* are not confined to differentiation and development of gonadal (and brain) cell lineages and examine the potential influence of the novel splicing and transcriptional regulatory mechanisms of these factors within and outside the HPAG axis.

Results

Stages that present challenges during development

To characterize the stages of development of the Atlantic salmon and identify potentially problematic stages, we assayed mortality rates during embryonic and larval development. The day of fertilization was Day (D) 0, and hatching and yolk sac absorption occurred between D38 to D40 and D68 to D70, respectively. Images of selected stages taken during the course of early development are presented in Figure 1. We counted the mortalities each collection day during the development study. We found that there were at least four stages which can present potential developmental challenges (see Additional file 1). One stage comes at D13.5, the point at which we noted eye formation was initiated. Next, from D38 to D50 is a period when hatching had begun. Most larvae hatched between D38 and D40, but there also were a large number for which hatching was delayed. The large peak at D49 (see Additional file 1) marks 184 mortalities due to late, incomplete or unsuccessful hatching. Many larvae in this group failed to complete hatching or had survival problems once hatched. The two small peaks that follow this period represent morphological abnormalities that arose during alevin development (ie; curved backs).

Early expression of sex differentiation factors

It became increasingly clear from preliminary studies that the expression of known sex differentiation regulators were expressed much earlier in salmonid development than we had expected. To determine when expression of these sex differentiation factors was initiated, we prepared RNA from embryos two days post-fertilization (D2) and succeeding time points. Our RT-PCR results show that the mRNA for many of these genes is present very early in development (Figure 2), at least one month ahead of previously characterized expression of regulators linked to differentiation of the salmonid gonad [5,14]. It was necessary to reamplify the early amplification products of

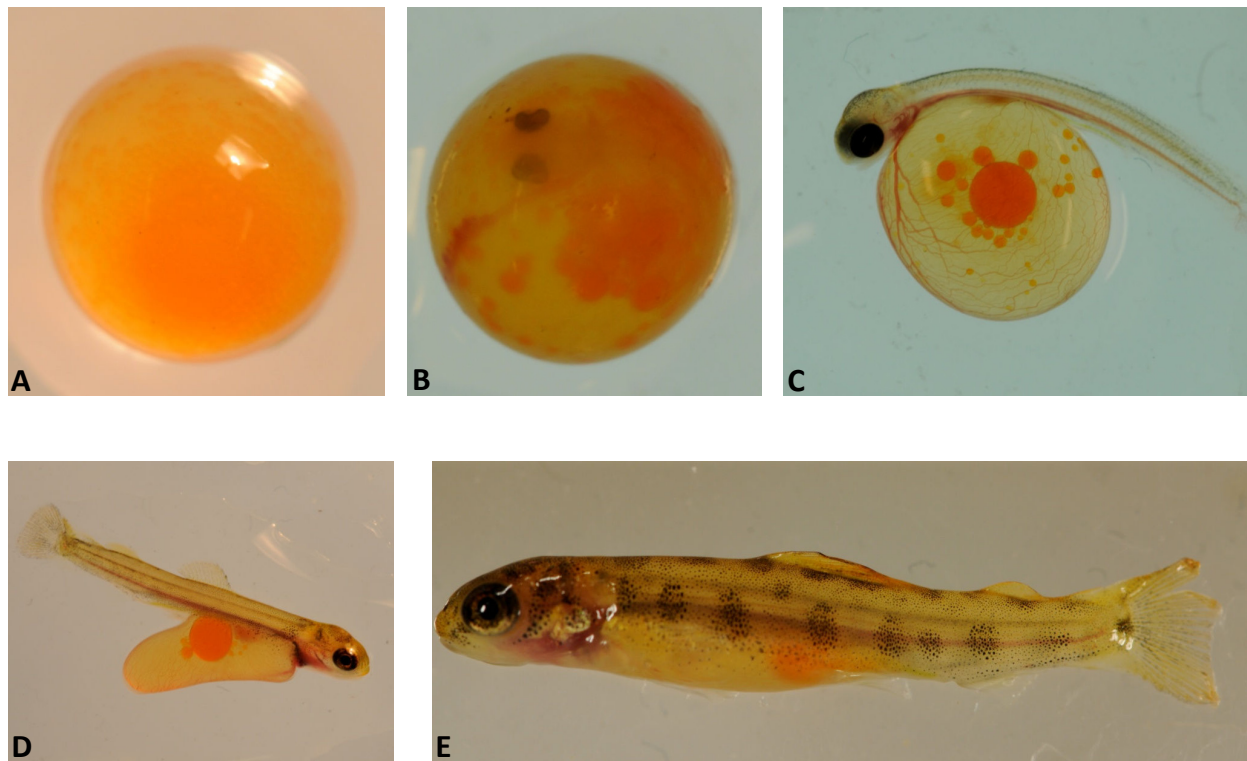


Figure 1 Images of selected development stages. Ages based on day ((D) post-fertilization. A. D7. B. D23. C. D38. D. D53. E. D71.

dax2, *foxl2a*, *foxl2b*, *mis* and *sf1* for presentation in Figure 2. We do not know if these transcripts are maternally contributed or if they are the result of zygotic transcription. To the best of our knowledge, for the genes we detected at these early stages, there presently is only evidence for the expression of *dax1* in embryonic cells in mammals [24].

Also of interest are the gene expression results for larvae and alevin between D38 to 77. It is known that *dax1*, *foxl2* and *sf1* expression correlates with transcription of *cyp19a* in many species [6,11,14,16,17,25,26]. We therefore were interested in examining the expression of these regulators in relationship to *cyp19a* transcription. Most if not all of the genes known to be involved in aromatase transactivation were expressed during this period (Figure 3). *cyp19b1* was expressed weakly as early as D21 (Figure 2), but weak expression of *cyp19a* was not detected until at least D47 (Figure 3). The earliest “strong” detection of the expression of *cyp19a* was D59, nearly two months after the earliest expression of its regulators (Figure 3). Thus, although the genes (and presumably their protein products) important to *cyp19a* regulation were present, it was not until three weeks later (D59) that *cyp19a* transcription was strongly detected (Figure 3). This is similar to findings that were

made for ovary-specific *cyp19a* expression in trout [5]. This time point may be considered a mark of the earliest molecular expression of steroidogenic and differentiation factors before morphologically distinguishable features of sex differentiation [5].

Expression of sex differentiation factors in various adult tissues

Most of the information available on the genes we examined is from investigations done on the brain or gonadal tissue (eg; endocrine organs). We therefore explored the expression of each gene in twelve tissues extracted and prepared from three different male and female adult fish. Analysis of the RT-PCR results indicates that most of these genes were expressed quite broadly (see Additional file 2). We were most surprised to find the expression of *sf1*, *mis* and both *sox9* genes across all tissues examined, particularly in the male. It was also interesting to compare the expression patterns between paralogs where it was possible to distinguish potential regulatory differences between *foxl2a* and *foxl2b*, and *cyp19a* and *cyp19b1* transcripts (see Additional file 2). We also detected the expression of multiple products for *cyp19b1* and *mis* and these amplicons were isolated, cloned and sequenced.

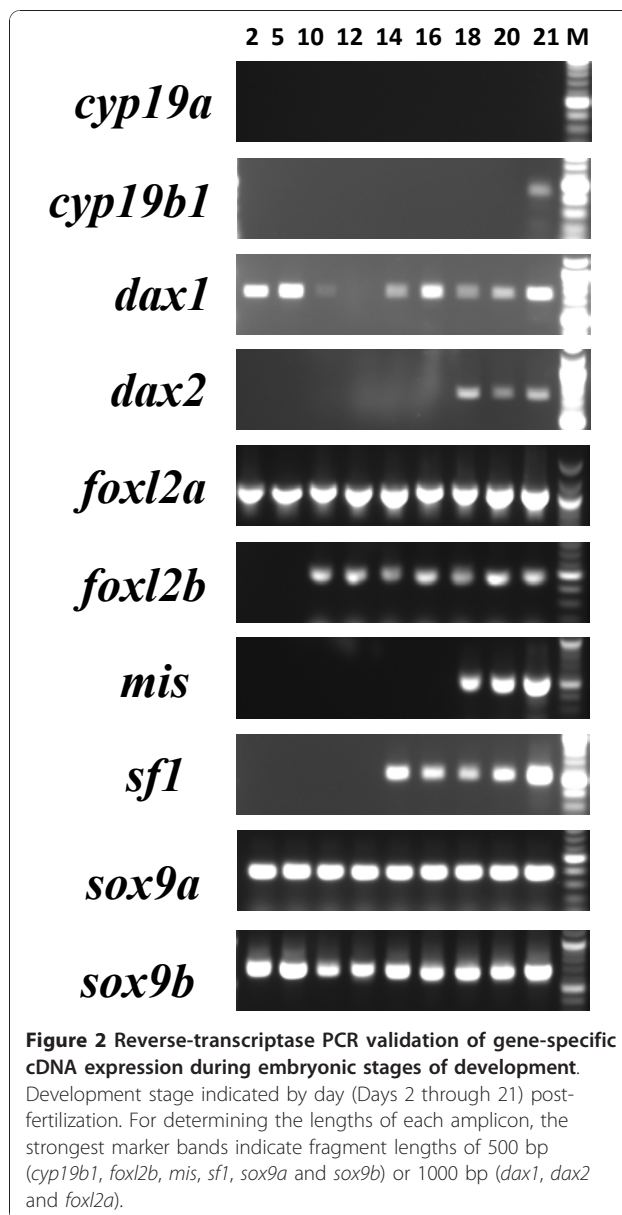


Figure 2 Reverse-transcriptase PCR validation of gene-specific cDNA expression during embryonic stages of development. Development stage indicated by day (Days 2 through 21) post-fertilization. For determining the lengths of each amplicon, the strongest marker bands indicate fragment lengths of 500 bp (*cyp19b1*, *foxl2b*, *mis*, *sf1*, *sox9a* and *sox9b*) or 1000 bp (*dax1*, *dax2* and *foxl2a*).

Alternative splicing of *mis* transcripts

MIS is a member of the TGF superfamily of proteins that share features that are conserved across many species from fish to mammals [27], each containing an N-terminal prodomain and a C-terminal domain that presents MIS. Proteolysis of the prodomain from the C-terminal domain at specific protease recognition motifs is required to release the hormone activity of MIS [27]. We isolated three transcripts of different sizes for *mis*. The primer set used was designed against sequence that spans a region of DNA that encodes potential enzymatic cleavage recognition residues and that is upstream from the bioactive hormone region of the molecule (nts 27979 to 28608 [GenBank:HM159473]) (Figure 4A).

We found expression of a complete sequence (629 bp), and two alternatively-spliced transcripts of 436 and 396 bp in length (see Additional file 3). Similar alternatively-spliced variants of *mis* have also been characterized in the European sea bass [27]. The *mis* (629 bp) and *mis* (436 bp) transcripts terminate 22 and 47 codons downstream from their putative protease RLRR recognition motifs, respectively (Figure 4B). Translation of these transcripts thereby does not generate the C-terminal mature hormone portion of salmon MIS. Only *mis* (396 bp) is in-frame to encode the bioactive hormone portion of the gene (Figure 4C).

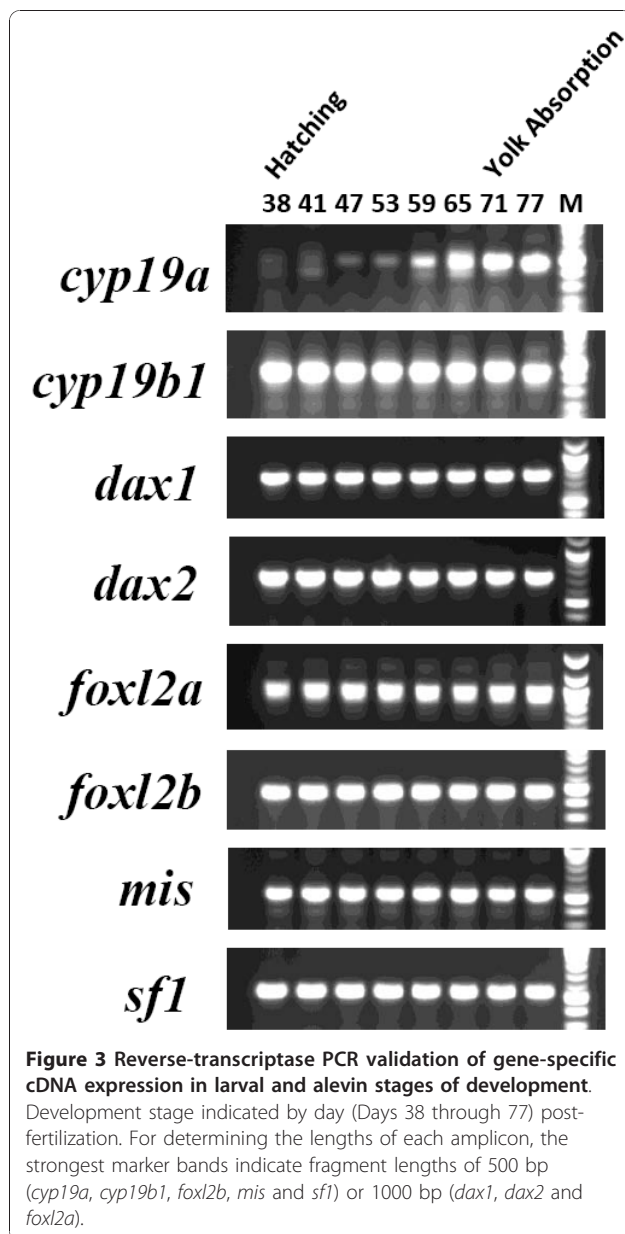
The features of the different *mis* transcripts may be important for different regulatory capacities, depending on the tissue in which they are expressed. For example, *mis* (629 bp) is expressed at relatively high levels in the muscle, skin, gill, spleen, brain and heart; *mis* (436 bp) is most "strongly" expressed in the gonad (as well as the kidney and eye), and *mis* (396 bp) is expressed at a relatively low level in most of the male tissues we examined (see Additional file 2). Other potential alternatively-spliced amplicons of *mis* are observed in several of the female tissues.

In the early development stages, expression of *mis* was first observed on D18 and we did detect the expression of multiple *mis* amplicons by D27 (data not shown). We attempted to clone and sequence each of these PCR products, but were successful in obtaining information for only the larger 629 bp transcript.

Alternative splicing of *cyp19b1* transcripts

During the process of preparing clones for sequencing to confirm their identities, we isolated and sequenced *cyp19b1* transcripts of three different sizes from D38 larvae. Previous work demonstrated that at least five different *cyp19b1* transcripts that bear differences in the lengths of their 5'-UTRs are generated in the brain and gonads of rainbow trout [23,28]. We now show similar results for the transcription of *cyp19b1* transcripts with 5'-UTRs of a variety of lengths in Atlantic salmon.

The salmon transcript 1 utilizes the more conventional intron 1/exon 2 splice sites and is devoid of intron 1 (Figure 5). Transcript 2 has been previously identified in the trout brain and gonads [23,28] and includes a portion of the 3'-end of intron 1 of the salmon *cyp19b1* gene (Figure 5 and 6). Transcript 3 contains an extra 69 nts of intron 1, in addition to the most 3'-end of the intron that is included in the 5'-end of transcript 2. This is the first time that a *cyp19b1* transcript using this more upstream splice site has been reported for a salmonid. Each of these distinct mRNAs use splice sites that follow the canonical pre-mRNA GT-AG rule for intron splicing [29]. It is interesting that the cryptic 3'-splice site utilized in transcript 2 is



bypassed by the spliceosome in favour of the more upstream acceptor AG site in transcript 3. Of eight *cyp19b1* clones isolated from amplifications of D38 cDNA, six were transcript 2 representatives. We presume that other examples of these partial intron 1-containing *cyp19b1* transcripts exist. However, several of the larger (>521 bp), weaker *cyp19b1* PCR products detected in the adult tissues were isolated, cloned and sequenced and determined to be non-specific amplifications (see Additional file 2).

Generally, strong 3'-splice sites in vertebrate intronic regions contain a pyrimidine-rich tract and a terminal AG [29,30]. These sequences are important for spliceosome recognition for precise intron removal [30]. Examining the

30 nts preceding each of the three 3'-splice sites in intron 1 of *cyp19b1* revealed that the conventional intron 1/exon 2 junction, the cryptic 3'-splice site and the most upstream splice site had 33% (1), 53% (2) and 63% (3) CT content, respectively (Figure 5). Interestingly, the splice site with the least pyrimidine content is that of the more conventional intron 1/exon 2 junction. It is therefore possible that the formation of this junction is the least favoured of the three splice sites presented here.

A comparison of the rainbow trout [GenBank: AJ716203] and Atlantic salmon *cyp19b1* intron 1 shows strong conservation in the 5'- and 3'-ends (data not shown). The central region of the two introns contain indels and are less conserved. Interestingly, only one intronic repeat sequence (IRS) segment, previously reported as four contiguous blocks of 46 nts in the trout [23], was present in the salmon *cyp19b1* intron 1 (Figure 6). The loss of the other three IRSs in the Atlantic salmon comprises almost half of the difference in the sizes between the introns (trout: 1264 bp; salmon: 964 bp).

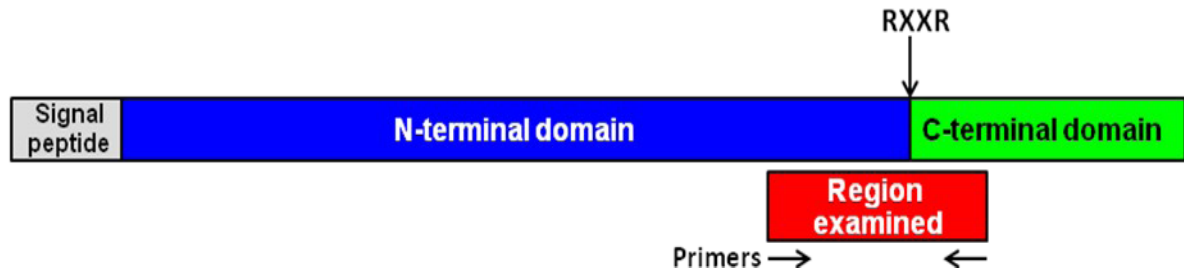
The novelty of the cryptic 3'-splicing in the *cyp19b1* intron 1 [23, 28 and refs therein] suggests this region is a target for regulators that bind signal elements within it to dictate distinct splicing events. Interestingly, a cluster of perfect and imperfect core recognition motifs for FOXL2 (5'-(G/A)(T/C)(C/A)AA(C/T)A-3' [16,25,26]) are located in the salmon *cyp19b1* intron 1 (Figure 6). The IRS contains at least one potential half-site (AGGACCT) that could potentially bind estrogen receptor or retinoic acid receptor-retinoid X receptor heteromers. The IRS is also flanked by two everted SF-1 motifs that present only one mismatch to the consensus sequence YCAAGGYCR, where Y = T/C and R = G/A [31 and refs therein].

Recent viral and mammalian work has demonstrated that exonic splicing signals exist that can enhance (ESEs) or suppress (ESSs) spliceosome recognition for intron splice sites [30]. Although this class of RNA processing signal is ill-defined in fish, we previously identified several putative exonic splicing enhancers and suppressors in exon 2 of trout *cyp19b1* [23]. These ESEs and ESSs are almost identical in the Atlantic salmon *cyp19b1* exon 2 (data not shown), indicating that this exon contains significantly positioned splicing motifs that could also potentially direct the lengths of the 5'-end of *cyp19b1* mRNAs.

Conservation of nucleocytoplasmic transport signals in SOX9

The N-terminal high mobility group (HMG), DNA-binding domain of the characterized salmonid SOX-9 proteins are highly conserved, even in comparison to the mouse and human SOX-9 proteins (Figure 7). Signals important for the import and export of SOX-9 between the cytoplasm and the nucleus have been identified [32,33]. In particular, the bipartite and basic cluster

A.



B.

```

MIS (629 bp) ASVPLYSLDSLPPLSLGVSSSETLLARLLNSSAPTLFSFPTQGSVLQGHGELS LQPALL
MIS (436 bp) ASVPLYSLDSLPPLSLGVSSSETLLARLLNSSAPTLFSFPTQGSVLQGHGELS LQPALL
MIS (396 bp) ASVPLYSLDSLPPLSLGVSSSETLLARLLNSSAPTLFSFPTQGSVLQGHGELS LQPALL
*****

MIS (629 bp) EVLRQRLEEVLVQMRAEEVVGKAGMDRLRRLQELSVLPKES E EAPAGEQASM*
MIS (436 bp) EVLRQRLEEVLVQMRAEEVVGKAGMDRLRRLQELSVLPKES E EAPAGV GSPSETQYPAPLS
MIS (396 bp) EVLRQRLEEVLVQMKALQTVVGAWDVERGQRATRA G QK-DPGNQHL CRLHSLTVSLEKYL
***** * * * *

MIS (436 bp) AAQSHRVPGEIPAVSS*
MIS (396 bp) LSPPGATIYNCQ
    
```

C.

```

MIS (396 bp) KALQTVVGAWDVERGQRATRA G QKDPGNQHL CRLHSLTVSLEKYL LSPPGATIYNCQ
HM159473    KALQTVVGAWDVERGQRATRA G QKDPGNQHL CRLHSLTVSLEKYL LSPPEATIYNCQ
*****
    
```

Figure 4 Amino acid residue alignment of translated *mis* transcripts. A. Schematic presentation of the organization of MIS and the region examined. The organization of MIS is based on alignments presented in Halm et al. [27]. B. Partial amino acid residue alignment of translated MIS transcripts. Open reading frames based on sequence upstream from the C-terminal hormone portion of salmon MIS. Stop codons are denoted by red asterisks. Putative protease recognition motifs are boxed. C. Amino acid residue alignment of translated MIS (396 bp) and HM159473. The region bearing the protease recognition sites is aligned with the corresponding portion of HM159473 to highlight potential of *mis* (396 bp) to encode the C-terminal bioactive portion of salmon MIS.

nuclear localization signals (NLS-1 and -2) [32], as well as the leucine-rich nuclear export signal (NES) [33] are 100% identical across the species examined (Figure 7). The largest divergence between the mammalian and salmonid SOX-9 proteins occurs in the C-terminal non-HMG, protein-interface and transactivation domain (data not shown).

BAC organization and regulatory regions

In order to understand more fully the regulatory regions and chromosomal context of specific factors thought to be important in sexual differentiation processes in salmon, we isolated several BACs containing *dax1* and

dax2, *foxl2b* and *mis*. Once the BACs had been isolated and assembled, we examined approximately 3.5 kb of sequence in the proximal promoter regions and identified a number of different transcription factor response elements for each gene of interest (GOI). As well, the chromosomal environment in which each GOI resides is presented in Figure 8.

The mammalian *dax1* promoter contains binding sites for OCT3/4, SOX-2, SF-1 and other regulatory factors [20]. We show both of the salmonid *dax* promoters contain binding elements for SF-1 and SOX-2 factors, but that only the *dax1* proximal promoter contains OCT3/4 motifs (see Additional files 4 and 5). The possibility exists

```

transcript 1      ACAGCAGTGGCTACACAGGAGACGGACACAAACTGTAGCCAGGCTAGAGG
transcript 2      ACAGCAGTGGCTACACAGGAGACGGACACAAACTGTAGCCAGGCTAGAGG
transcript 3      ACAGCAGTGGCTACACAGGAGACGGACACAAACTGTAGCCAGGCTAGAGG
                    *****

transcript 1      CAGACTACTCTGGGGCAGTCCAGGATGAGAGCG-----
transcript 2      CAGACTACTCTGGGGCAGTCCAGGATGAGAGCG-----
transcript 3      CAGACTACTCTGGGGCAGTCCAGGATGAGAGCGATAAGCGACACCTACTT
                    *****

transcript 1      -----
transcript 2      -----
transcript 3      CGGTGAGGTGCAAACCAGTGGGGCATGATTTACTTTTAAATATTTAACC

transcript 1      -----
transcript 2      --GTAGTCTCCTTGAAGTAAGTCAGAAGACCTTTTCCCTGGGGAGATAAT
transcript 3      AGGTAGTCTCCTTGAAGTAAGTCAGAAGACCTTTTCCCTGGGGAGATAAT

transcript 1      -----CTGTCATCATAAC
transcript 2      ACTTGGAAAGGAGAATAATTAATAGCAGAGTGCTCCAGCTGTCATCATAAC
transcript 3      ACTTGGAAAGGAGAATAATTAATAGCAGAGTGCTCCAGCTGTCATCATAAC
                    *****

                    M A T S I V D D
transcript 1      ACAGCAGCCTCCAGACCTCCTCACTGTAATGGCTACCAGTATAGTGGACG
transcript 2      ACAGCAGCCTCCAGACCTCCTCACTGTAATGGCTACCAGTATAGTGGACG
transcript 3      ACAGCAGCCTCCAGACCTCCTCACTGTAATGGCTACCAGTATAGTGGACG
                    *****

                    T S M S E A L L L L L L L S L L
transcript 1      ACACTAGCATGTCTGAAGCCCTGCTGCTGCTGCTGCTGCTCTCTCTCCTG
transcript 2      ACACTAGCATGTCTGAAGCCCTGCTGCTGCTGCTGCTGCTCTCTCTCCTG
transcript 3      ACACTAGCATGTCTGAAGCCCTGCTGCTGCTGCTGCTGCTCTCTCTCCTG
                    *****
    
```

Figure 5 Sequences of alternative 5'-ends found in three different salmon *cyp19b1* transcripts. Transcripts 2 and 3 contain spliced sections of the 3'-end of intron 1 in comparison to transcript 1. The position of the initiator ATG start codon in exon 2 is highlighted.

that these promoters could also be regulated by estrogen receptor since both have potential palindromic EREs. Other clear differences are discernable in the regulation of these two genes, most notably in the large regions of CT- and GA-rich regions in *dax2* that are not present in *dax1* (see Additional files 4 and 5). The GA-rich region is contiguous with a block of DNA that contains at least four perfect repeats, each 40 nts in length.

We could not identify any canonical SOX binding elements in the proximal *mis* promoter. However, we did detect two potential half-AREs and the presence of two SOX footprints [34] (see Additional file 6). Several OCT3/4, PPARE and WT-1 response elements were also identified. As well, typical GC boxes are located in the *mis* promoter (5'-GGGGCGGGG-3') that could bind SP-1 and Kruppel-like factors [35].

We found a tetrapartite GCAT sequence in the *foxl2b* promoter that might bind SMAD factors. SMADs can act through GCNT, GCCG and GCAT sequence motifs [36]. Interestingly, the *foxl2b* promoter does not contain a strong consensus SF-1 binding motif. However, it is possible that a weaker site at -802 to -795 (5'-TCAGGCCA-3') could interact with SF-1 (see Additional file 7). We also provide evidence that FOXL2 may regulate its own expression through several binding sites present in its promoter. Numerous potential binding sites for ERE, RXR/RAR and WT-1 are also located in the *foxl2* promoter.

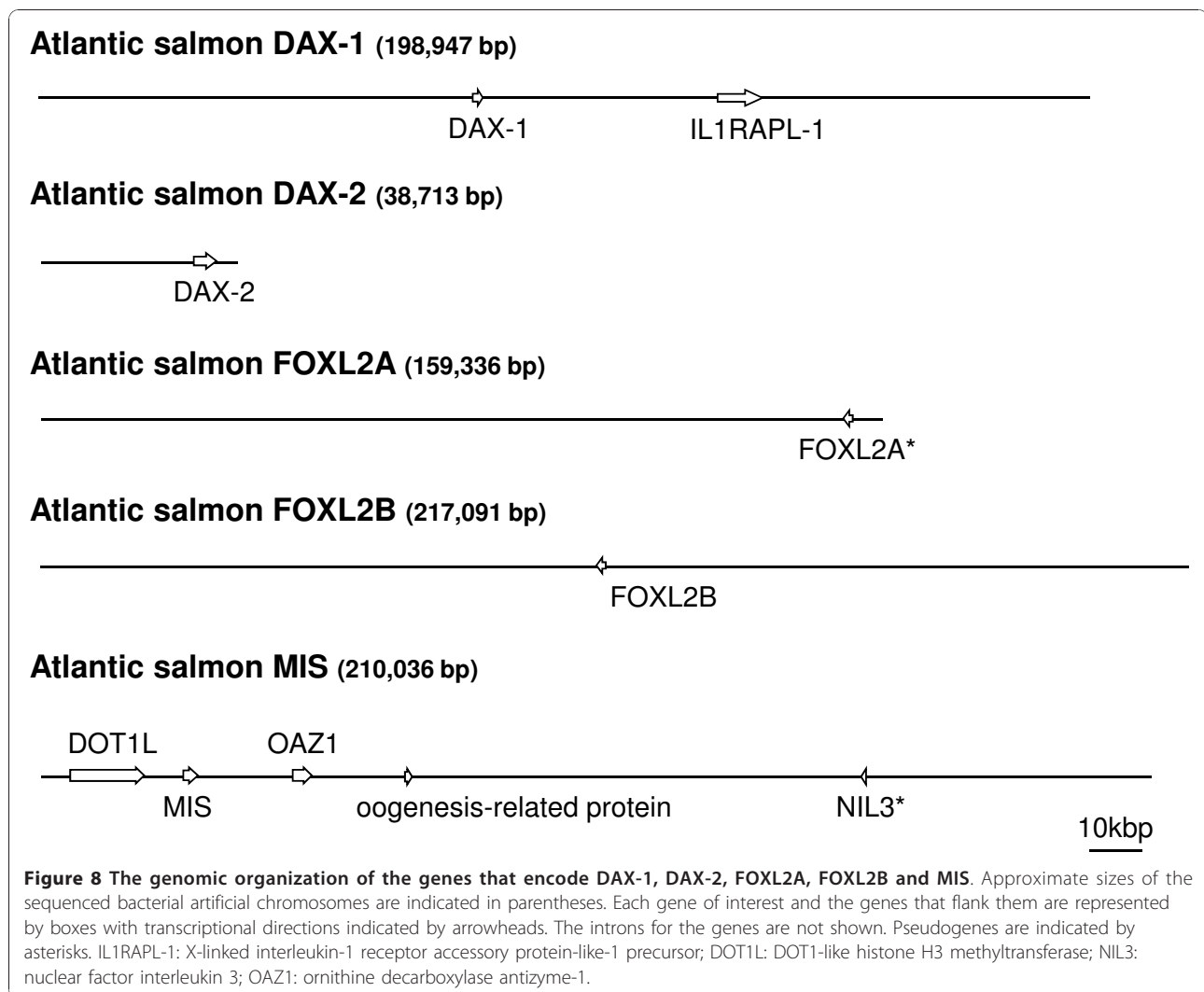
We also isolated a BAC containing the *foxl2a* pseudogene. The pseudogenic *foxl2a* mRNA, if transcribed, would share similarity with the characterized trout orthologs beginning in regions approximately 240 and 260 nts downstream from the start codon of the

ACAGCAGTGGCTACACAGGAGACGGACACAACTGTAGCCAGGCTAGAGGCAGACTACTCTGGGGCAGTC
CAGGATGAGAGCGgtGAGTATTATCCTTTTCTACATAATGACAATTCTATAAATAGGGGAACTTTTTG
 TCTGTATGTCTAATTAAGTGGAGGGAATTGTTTTCTCTCTGTGGAAGAAGAAGTGTCAAGAGGGATATTC
 TTGAAAAGAGGCCTATAACAGAATAAGAAATATAATTGAAATTTGTACAATTATAAACATAAAAAATAAT
 TAAATCTGTATTATAACAATGTTACAGCAAAAAAACTTCTAGAGGGGATGATGACGAGAACAAGCTGA
 CCAAGTTTCTCAATAATAAAGAAACAACAGCATTAGTACATTCTGATAAAATGACAAACCTCAGCC
 TTCCCTCTGGTTTCAGAAGTTCTAATACAATATCGTTGATGTCTAATTTAATAACTCTAGCCTTGGTAA
 GACAt**CTGGA**ACTGTTTCATGCAATAGTACAAC**GTCA**ACTTGAAAAGCTTTCCGTGGGTGTGTTATGGT
 GACCTTGTTACCGTGTTCACATGTTGTACCAACCCCATAGGGCAAGGACAGGAGAAATATTAAGGGG
 ATGAAAAGAGAGCAAGGGTTGTCACGGAGTTAATGCTTCCCTATTTTGTGTTTGACAATTTGAATCCCC
 TGTCTACGTC**CAAAT**g**GACCC**TATTCCCTATTTAGTGC**ACTACTTTTAA**CCAGGACCTATAGGGCTCT
 GGTATTAAGAAAGTGTACACTACATAGGGAATAGGGTGTAAATTTGGT**tCTGGA**AATTCAGGGTGA**TC**
 AATACATATACAGTAGTCTCTTAGAAGCCTTTTTCTTCGGCTGTCCAATag**A**TAAAGCGACACCTACTT
 CGGTGAGGT**GCAAAC**cAGTGGGACATGATTTACTTTTAAATATTTAACCag**G**TAGTCTCCTTGAAGTAA
 GTCAGAAGACCTTTTCCCTGGGGAGATAATACTTGAAGGAGAATA**ATtAATA**GCAGAGTGCTCCag**CTG**
TCATCATAACACAGCAGCCTCCAGACCTCCTCACTGTAATGGCTACCAGTATAGTGGACGACACTAGCAT
GTCTGAAGCCCTGCTGCTGCTGCTGCTCTCTCTCCTGCTCATCACGACCTGGTGCCTCACTAATACC
TCACACATACCAGgtAAGACCAATGATGTATGAATGATGTGTGAATGTGTCTACGGGGAAAAACACAGCCA
 CTCTCTCTCCGTCTAACAACATCACTACCGTGAATATGAAATGTCATAATGTCTTTTTGTTGATCAT
 TTGAAGATTAAACCTTTTGAGGTTTAATACATTTTGAATTGTATCTCTAATATCTCATAAAAAGTTGATA
 GCCAATAGTTAACTAGAAAGACATTGGCTATTCATCTAATATAACGGTTATTACTGTAAGAAGGTGGCAA

Figure 6 Sequence of Atlantic salmon *cyp19b1* exon 1, intron 1 and exon 2. Boxed sequences represent putative binding elements of FOXL2. Arrows indicate the direction of potential regulatory motifs for SF-1 shown in bold. Numbers 1, 2 and 3 show the positions of each of the three AG acceptor 3'-splice sites within intron 1 that can be selected by the spliceosome to generate different 5'-UTRs in *cyp19b1*. An intronic repeat sequence (IRS) segment is shown in italics and enclosed by asterisks. Exons 1 and 2 are highlighted in bold and the ATG translation start site is shown in exon 2.

ortholog [GenBank:AY507927] and the diverged paralog [Genbank:AY507926] (see Additional file 8). We could find no inframe ATG initiation codons for the pseudogene in sequence upstream from these regions. Furthermore, multiple stop codons in the pseudogenic reading frame would disable any possible generation of

a full-length protein. A comparison of the salmon pseudogene with the trout FOXL2 open reading frame shows a region of only short identity with the forkhead domain with the N-and C-terminal ends diverging completely (Figure 9). However, we did isolate an amplicon of 1010 nts that represents the salmon *foxl2a* based on

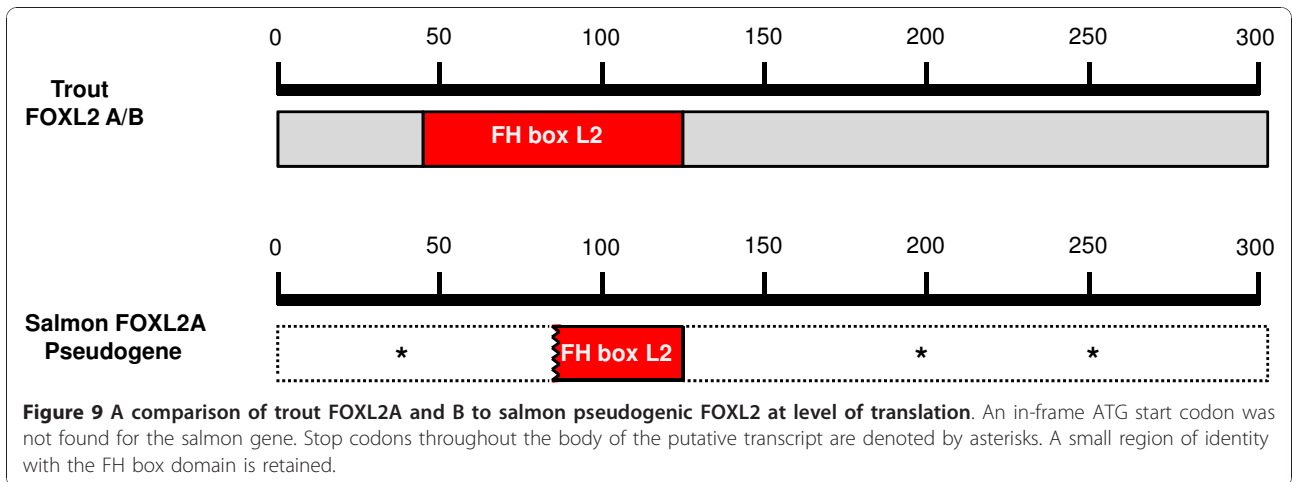


not shown). The activity of *mis* was absent only in precocious or prespawn trout testis. In this study, we show expression of *mis* during early development and in various adult salmon tissues (Figure 2 and 3; see Additional file 2). The expression of *mis* by D18 indicates the potential for a role early in salmonid development. Since MIS is a glycoprotein member of the TGF β family, it may serve in establishing and shaping different cellular phenotypes well before patterning of the reproductive tissues.

Since we observe expression of *mis* and the absence of *cyp19a* in most adult tissues (see Additional file 2), it could be suggested that MIS downregulates *cyp19a*. In contrast, the expression of *cyp19b1* is evident in most adult tissues. Transactivation of *cyp19b1* may not be reliant on gonadotropic inputs like *cyp19a* (ie; cAMP stimulation [17,23,40]), and therefore could be refractory to inhibition by MIS. Nevertheless, unlike in mammalian

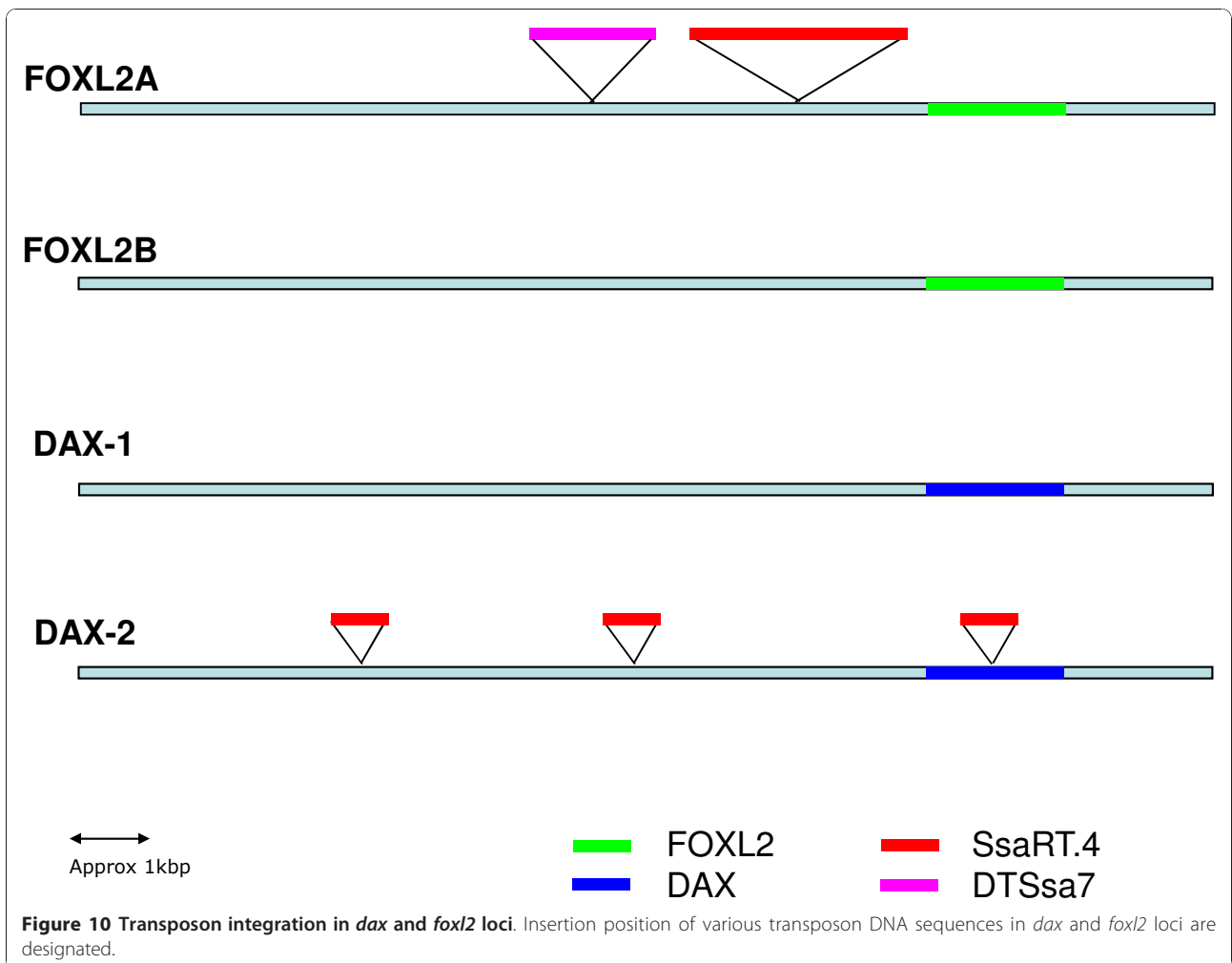
models where *mis* expression is absent during the onset of sex determination, *mis* expression is detected in the embryonic gonads of both trout [5,14] and tilapia [6]. Although the levels of expression in the female are much lower than in the male during the crucial sex determination window [5,6,14], it may still play a role in female development. For example, disruptions to MIS signaling have been shown to impact germ cell proliferation and cell-cycle timing in both sexes of medaka [42].

Furthermore, in mammalian models, *mis* is the downstream regulatory target of SOX-9 activity in the differentiating testis. However, expression of *sox9a* and *sox9b* occurs in both the trout male and female gonad during sex determination [5]. Part of the decoupling of MIS expression from SOX regulation, at least in salmonids, could be due to the lack of canonical SOX binding elements in the proximal *mis* promoter (see Additional file 6). Another control mechanism that may differ in fish



that exists in mammals is that nuclear localization of SOX-9 occurs in the male (to promote subsequent MIS transcription), but is absent in the primordial female gonad [43]. The subcellular localization signals that have been shown to direct SOX-9 nucleocytoplasmic

shuttling during mammalian male sexual differentiation [32,33] are highly conserved in salmonid SOX-9 proteins. Whether these signals are targets for sex-specific transport factors; and whether the *sox9* paralogs are localized to different cell-types, and are involved in



differentiation of Sertoli precursor cells in salmonids, still remains to be determined.

In mammals, the MIS precursor protein consists of two sections - the N-terminal prodomain and the C-terminal, mature hormone region [44]. The mammalian studies indicate that the MIS precursor protein is proteolytically cleaved at consensus RXXR sites [44]. It is not known if fish MIS proprotein is similarly proteolytically activated, but the organization and conservation of these motifs across many species does suggest similar mechanisms of hormone bioactivation [27]. Although the mammalian studies indicate that the N-terminal prodomain section of MIS serves only to stabilize and to target the processed hormone [44], our work suggests that some *mis* transcripts generate only truncated prodomain representatives. The function of these transcripts, in various tissues and during early development, remains to be determined. It is possible that the expression of transcripts that encode the MIS prodomain alone may add another level of control for tissue-specific regulation of MIS activity, or that they influence the bioactivities of other members of the TGF family, such as BMPs, activin or inhibin. Heterodimerization between various members of this large family of growth factors has been shown to direct different cell-patterning decisions [45 and refs therein, 46].

DAX

DAX-1 has been described to be either a possible ovarian- or testicular-determining factor [12,14,44,47]. We previously found *dax1* to be consistently expressed across all reproductive development stages examined in trout ovary and testis [23]. Although both the salmon *dax1* and *dax2* described here were expressed in all adult tissues (see Additional file 2), the appearance of *dax1* was at least two weeks before *dax2* in salmon embryos (Figure 2).

Although the interactions that promote the undifferentiated state in embryonic stem (ES) cells are well-established, the events that antagonize them to initiate cellular differentiation are poorly understood. The expression of *dax1* has been associated with early mammalian ES cells in maintaining the undifferentiated state and to safeguard pluripotency [20,48]. OCT3/4 and SOX-2 are also considered "core" transcription factors in the maintenance of pluripotency [49]. It has been shown that these factors interact cooperatively to regulate *dax1* expression in ES cells [20]. We found putative response elements for OCT3/4 and SOX factors in the *dax1* and *dax2* promoters (see Additional files 4 and 5), that implicate their role in the regulation of the salmonid genes as well.

It also has been proposed that SF-1 expression is not coupled to *dax1* regulation in ES cells [20]. In this

regard, it is also important to note the location of the SOX consensus binding elements in close proximity to the SF-1 motifs in both salmonid *dax* promoters (see Additional files 4 and 5). SOX binding of its response element may interfere with SF-1 binding during early development stages. We also show the expression of *sf1* no earlier than D14, a period that succeeds a potentially important differentiation transition point (see below) (Figure 2).

Regulation of alternative splicing by DAX-1 and SOX factors may control stem cell differentiation, and initiation and development of the gonad [47]. It has been demonstrated that *dax1* inhibits pre-mRNA splicing of target genes and that these activities can be restored by the addition of HMG-domain containing factors [47]. Moreover, it has been proposed that the control of splicing is exerted through competition between DAX-1 and SOX proteins for splicing factor complexes [47]. Antagonism between DAX-1 (and OCT3/4 [20]) and SOX-2 (and other HMG-domain proteins?) [47] may be the mechanism whereby "stemness" versus differentiation-associated genes are regulated in ES cells. The decline (D10), absence (D12) and reappearance (D14) of *dax1* we detected (Figure 2) may reflect a transition that results in differentiation of embryonic cell layers in specific developing organs; similar to the different cell lineages that have been shown to arise following *dax1* downregulation in mouse [21,24].

FOXL2

Both *foxl2* transcripts were found to be expressed very early in development (Figure 2). In the male, we found *foxl2a* expression to be quite restricted in comparison to the *foxl2b* diverged paralog, which was present in most adult extragonadal tissues examined (see Additional file 2). In contrast, expression of *foxl2b* is quite restricted in the adult female tissues, essentially confined to the skin, gill and ovary. The expression of *foxl2b* in the testis was also very strong, correlating well with previous findings that showed *foxl2b* was present, with *foxl2a* absent, in later stages of testis development in trout [23]. Taken together, this may indicate that *foxl2b* is an important regulator of male-specific genes (eg; *cyp19b1*), at least later in development.

The expression of *foxl2a* is associated with upregulation of ovarian-specific *cyp19a* and with differentiation of the ovary in many species [6,14,16,17,25,26]. Upregulation of *foxl2* genes have been correlated with *cyp19a* expression during ovarian differentiation in trout [14], Japanese flounder [17] and Nile tilapia [6,16]. Depending on the study, transactivation of *cyp19a* occurs directly through the FOXL2 binding elements or in concert with SF-1. SF-1-mediated regulation of aromatase activity is modulated by DAX-1, which binds SF-1 directly [11,15].

We previously implicated FOXL2 with the generation of *cyp19b1* transcripts bearing 5'-ends of different lengths, potentially in complexes with SF-1 and DAX-1 [23]. At least two SF-1 binding sites and a number of potential FOXL2 response elements reside in the trout *cyp19b1* intron 1. Because of well-characterized DAX-1 spliceosome activity (see above), and the association of both DAX-1 and SF-1 with FOXL2 at aromatase promoters in fish and mammals [6,11,15-17,25,26], we hypothesize that complexes of these gene products could also control splicing activity of *cyp19b1*.

This mechanism of regulatory control in the synthesis of these transcripts is further reinforced by our present characterization of the Atlantic salmon *cyp19b1* intron 1 (Figure 6) which maintains many of the features described for the trout *cyp19b1* intron 1 [23]. In accord with this model, we demonstrate the generation of differentially spliced, partial intron 1-containing *cyp19b1* transcripts in salmon. We argue that if DAX-1, FOXL2 and SF-1 are known to interact at aromatase promoters, then nucleation at the *cyp19b1* intron 1 is also plausible. Strong evidence exists that indicates that various transcription factors and nuclear hormone receptors can bind to RNA and regulate RNA splicing, migration and localization [50,51]. For example, both DAX-1 and SF-1 have been shown to bind steroid receptor activator [52], a RNA that acts as a bridge and platform to nucleate and stabilize transcription complexes. It has been proposed that nuclear receptors, such as SF-1, which can bind DNA as monomers, may recognize the YCAAG-GYCR motif in RNA [52 and refs therein]. Among the various factors that bind RNA, recognition for YCAAG-GYCR, as well as for other binding elements, may be through different mechanisms of interaction with the nucleic acid (e.g. zinc fingers, anti-parallel α -helices) [31,50,52].

Aromatase

In this study, we confirmed the expression of alternatively-spliced *cyp19b1* transcripts in only the larvae (D38) and gonad of salmon. However, expression of *cyp19b1* transcripts of different sizes and abundances are evident in a broad range of adult tissues (see Additional file 2), such as muscle (521 bp), skin (452 bp) and spleen (367 bp) of the male. Whether differences in processing results in partitioning of specific *cyp19b1* transcripts to distinct tissues remains to be explored.

We also show during early salmon development that at least a three-week delay in the transcription of *cyp19a* occurs in comparison to *cyp19b1* (Figure 2 and 3). Part of these regulatory differences could be dictated by estrogenic or gonadotropic controls [23 and refs therein]. For example, unlike *cyp19b* promoters, *cyp19a* promoters lack consensus EREs for the fish species

characterized to date [17,18,25,53-55]. Another important difference between these genes is that SOX binding elements, often many of them, are in close proximity to SF-1 binding motifs in teleost *cyp19a* promoters [18,25,53,54]. Competition between SOX factors and FOXL2 for access to their binding sites on *cyp19a* promoters could be key to ovarian-specific aromatase synthesis. This is compatible with recent work that shows that the mammalian ovarian phenotype is mediated by cooperative FOXL2 and ER repression of *sox9* transcription [22]; interactions that are opposed by the binding of SRY/SF-1 and SOX-9 in the male pathway [19]. Further work is needed to define the precise roles that FOXL2A and FOXL2B play in activation of the two salmonid *cyp19* genes early in development, as well as their function in the transcription of other promoters.

Genome rearrangements

There is a growing body of evidence that demonstrates that past transposon activity has led to salmonid genome remodeling [38,56]. We show the integration of SsaRT.4 very close to the transcription start site of the pseudogenic *foxl2a* (Figure 10). Although speculative, integration of SsaRT.4 into the *foxl2a* regulatory region may have led to the reorganization of the 5'-end and subsequent accumulation of stop codons into the coding region of this gene. Integration of SsaRT.4 into the *dax2* intron also provides an example of how transposon activity can rearrange genes; potentially leading to the creation of introns and indicating how genes can acquire new functions over time.

Conclusion

We have found expression of many factors classically associated with sex differentiation in embryonic stages well before sex fate determination. Their expression early in development points to the potential for function during embryogenesis. Expression of these regulators in diverse adult tissues also indicates their potential role in transcription programs independent of steroidogenesis and gonadogenesis. Further investigation is required to determine if these genes are translated in the embryonic and adult stages of development and to localize their expression to specific cell-types. Whether these genes are associated with stem cells, or with other differentiation factors in networks that influence cell-patterning throughout development, remain to be unequivocally determined.

Methods

Animals and Sampling

Treatment of the fish used in this study was in compliance with the regulations of the University of Victoria Animal Care Committee. Eggs from Atlantic salmon

(McConnell (Mowi)) were obtained in November, 2009 from Marine Harvest United Hatchery (Fanny Bay, B.C., Canada). The eggs were fertilized by gently mixing the eggs and milt by hand and then washed with partial exchanges of water. Approximately 2000 fertilized eggs were then transferred and placed in Heath trays (Marisource) at the University of Victoria. The embryos and larvae were raised in fresh water at a temperature of 12°C and a flow rate of 200 liters/h. Fry were transferred from the Heath trays to 30-liter holding tanks following yolk sac absorption.

The day of fertilization was Day (D) 0, and hatching and yolk sac absorption occurred between D38 to D40 and D68 to D70, respectively. Whole embryos ($n = 20$) were collected every other day for three weeks. Larvae were collected every third day and then alevin and fry every sixth day for the remainder of the study. Samples were directly placed into dry ice and stored at -80°C until RNA extraction.

The number of mortalities were counted on each collection day. We also took several pictures using a Nikon D300 s with a 105 mm 2.8 G_ED lens to record the progress of the development of the fish.

Twelve different Atlantic salmon (Mowi stock) tissues from three individual adult male and female fish were provided by Fisheries and Oceans Canada (West Vancouver Laboratories, West Vancouver, British Columbia). The six fish were euthanized, followed by rapid dissection of tissues (kidney, muscle, skin, gut, gill, spleen, brain, heart, gonad, liver, eye and pyloric caeca). The tissues were flash frozen in liquid nitrogen or dry ice and stored at -80°C until RNA extraction.

Reverse transcription and cDNA amplification

Total RNAs were extracted in TRIzol reagent (Invitrogen) by mixer-mill homogenization (Retsch) and spin-column purified using RNeasy Mini kits (Qiagen). The RNA from three whole embryos or two larvae or alevins was pooled for D2 to D20 and for D21 to D77 collections, respectively. RNA was extracted from portions of each adult tissue described above for each different fish. Each RNA sample was then quantified and quality-checked by spectrophotometer (NanoDrop Technologies) and agarose gel, respectively.

The cDNAs were synthesized in 25- μL reactions from 200 ng to 1.0 μg total RNA using oligo(dT)₁₅ (Promega) and Superscript II RNase H⁻ reverse transcriptase according to the manufacturer's instructions (Invitrogen). The reactions were incubated at 37°C for 90 min and the transcriptase heat-inactivated at 70°C for 30 min. Approximately 100 to 200 ng of cDNA was used in each 25- μL PCR reaction containing 1.25 U Taq polymerase, 1 \times Taq buffer, 1.25 mM MgCl₂, 10 mM dNTPs (Invitrogen) and 15 pmol of each gene-specific

5'- and 3'-primer. Each PCR was carried out under the following cycling parameters: 94°C for 2 min, then 35 cycles of 94°C for 30 sec, 55°C for 30 sec (or 58°C for *sox9b*), and 72°C for 1 min using a Perkin Elmer GeneAmp 9600. The PCR products were separated by electrophoresis on 1.0 to 1.5% agarose gels and photographs were stored using an UVP GelDoc-It documentation system (Ultraviolet Products).

Genomic DNA to enable isolation of the *cyp19b1* intron 1 was extracted from Atlantic salmon spleen using a DNeasy Blood and Tissue kit following instructions of the manufacturer (Qiagen).

PCR analysis of sex differentiation factor expression in adult tissue

PCRs were performed with an initial denaturation step of 2 min at 95°C and then 35 cycles as follows: 30 sec of denaturation at 95°C, 30 sec of annealing at 55°C and 1 min of extension at 72°C using a Peltier Thermal Cycler PTC-225 (MJ Research). Amplicons for each gene of interest (GOI) were visualized on 1.0% agarose gels stained with ethidium bromide and images were stored using an UVP GelDoc-It documentation system (Ultraviolet Products). The intensity of the amplification products were semiquantitatively measured using ImageJ software [57] and divided by the intensity of the respective ubiquitin signals. The data generated was graphed in Microsoft Excel 2007 and expressed as the mean \pm SE of three individuals studied.

Gene identification and primer design

The primers used to amplify trout ovarian aromatase (*cyp19a*), brain aromatase (*cyp19b1*), dosage-sensitive sex reversal, adrenal hypoplasia congenital, critical region on the X-chromosome, gene-1 and -2 (*dax1* and 2), forkhead box L2 ortholog (*foxl2a*), forkhead box L2 diverged paralog (*foxl2b*), Mullerian inhibiting substance (*mis*), steroidogenic factor-1 (*sf1*), sex-determining region of the Y chromosome (SRY)-related high mobility group (HMG) box (SRY box) Sox9a (*sox9a*), Sox9b (*sox9b*), and ubiquitin (*ubiq*) were designed specifically against the sequences provided for each gene obtained from <http://www.ncbi.nlm.nih.gov>. GenBank accession numbers, primer sets and product sizes for each GOI are shown in Additional file 12. The primer sets for each target gene were designed using Primer3 software [58] and purchased from Integrated DNA Technologies. The specificity of each primer set and the identity of each amplicon were confirmed by subsequent cloning into pCR2.1-TOPO vector (Invitrogen) and sequencing.

BAC screening and sequencing

Atlantic salmon (AS) CHORI-214 [59] bacterial artificial chromosome (BAC) libraries were obtained from

BACPAC Resources, Children's Hospital Oakland Research Institute (CHORI) [60]. AS BAC library filters were hybridized with 75-mer oligonucleotide probes for each GOI (Integrated DNA Technologies) that were 5'-end-labeled with ^{32}P -ATP using T4 polynucleotide kinase (Invitrogen). Filter hybridizations were conducted as described by CHORI [60].

Confirmation of each GOI-containing BAC was performed by comparisons of *Hind*III restriction digests of the isolated clones to *in silico* digests for each BAC. BAC clones were chosen based on the physical BAC fingerprint map for Atlantic salmon that is publicly available on the internet Contig Explorer (iCE) version 3.5 [61]. The BAC end sequence information, that is available in ASalBase [37], was also used for selection of the BAC clones.

BAC preparation and library construction

BAC DNA was isolated by an alkaline lysis procedure using Large Construct Isolation Kit (Qiagen) following the manufacturer's protocol. The isolated BAC DNA was nebulized and the DNA ends were made blunt by filling with T4 polymerase. The blunt-ended, repaired DNA was size fractionated by electrophoresis and the gel region corresponding to 1.5 to 3.0 kb was excised and gel purified (Qiagen). The fragments were blunt-end ligated into *Sma*I-cut M13mp19 vector and transformed into electrocompetent DH5 α *E. coli* cells using a Bio-Rad Gene Pulser system. Library quality was evaluated and high redundancy plating was followed by large-scale colony picking (Genetix). Extracted recombinant plasmid templates were sequenced on an ABI 3730 DNA sequencer.

BAC alignment and annotation

Bases were called using PHRED [62,63]. High quality sequence reads were assembled using PHRAP [64] and then viewed and edited using Consed [65]. Some gaps in BAC assembly were filled by designing primers to the contiguous sequence ends, followed by amplification of the BAC region by PCR and subsequent cloning and sequencing of the fragments.

Each BAC has been deposited in GenBank as follows: BAC 33H18 containing *dax1*

[GenBank:HM159469]; BAC 105M10 containing *dax2* [GenBank:HM159470]; BAC 217E24 containing *foxl2a* ortholog (pseudogene) [GenBank:HM159471]; BAC 261D01 containing *foxl2b* diverged paralog [GenBank:HM159472], and BAC 19C14 containing *mis* [GenBank:HM159473].

BAC clone sequence data were annotated using BLAST homolog searches and EMBOSS programs package [66]. Alignment of cDNA sequences against genomic sequences was done by Sequin [67], which was

obtained from NCBI. Identification of transposon insertions was performed using Dotter plots [68], comparing each GOI-containing BAC with salmon transposon sequences [38].

Fluorescent in situ hybridizations

BAC DNA was labeled with Spectrum Orange using a nick translation kit (Abbott Molecular). Human placental DNA (2 μg) and Cot-1 DNA (1 μg , prepared from Atlantic salmon) or Cot-1 DNA (1 μg , prepared from rainbow trout) were added to the probe mixture for blocking. Hybridizations were carried out at 37°C overnight and post-hybridization washes were as recommended by the manufacturer (Abbott Molecular) with minor modifications [69]. Antibodies to Spectrum Orange (Molecular Probes) were used to amplify the signal. Slides were counter-stained with 4',6-diamidino-2-phenylindole (DAPI) at a concentration of 125 ng DAPI in 1.0 mL antifade solution. Images were captured with a Jai camera and analyzed with Cytovision Genus (Applied Imaging, Inc.) software. Chromosomes were arranged according to size within the metacentric/sub-metacentric and acrocentric groups [70].

The chromosome preparations were obtained from blood of the Norwegian strain of Atlantic salmon (2N = 58) by methods described previously for salmonid fishes [70]. Briefly, the buffy coat was isolated from whole blood and placed in MEM media with pen-strep, L-glutamine, 10% fetal calf serum and 200 $\mu\text{g}/\text{ml}$ LPS and cultured for 6 days at 20°C. Cells were collected by centrifugation and re-suspended in 0.075 M KCl for 30 min, then fixed in 3:1 methanol acetic acid. Cell suspensions were dropped onto clean slides and allowed to dry on a slide warmer with humidity at 40°C.

Promoter analysis

To identify the location of potential transcription factor binding elements in the promoters of our GOI, we preliminarily used MatInspector (Genomatix Software GmbH). Our designation of OCT3/4 consensus binding motifs 5'-ACACGCAT-3' or 5'-ATTTGCAT-3' are based on work done in ES cells [20,71]. The HMG domain that binds to DNA is highly conserved among all SOX factors (5'-(A/T)(A/T)CAA(A/T)G-3') [72]. The SOX footprints that we also include in our analysis are based on work done on mammalian *mis* and *cyp19* promoters [34]. We use the term "immediate early response" element (IER) to identify response elements (5'-GCG(G/T)GGGCG-3') that potentially could bind early response products of genes such as EGR-1 and WT-1 [73]. Sequences presumed or demonstrated to bind CRE, ERE, FOXL2, PPARE, RXR/RAR and SF-1 in other fish species are also presented for each promoter examined [16-18,25,53-55,74].

Additional material

Additional file 1: Graphical presentation of number of mortalities found throughout early development. Day represents post-fertilization age.

Additional file 2: cDNA expression profiles for ten genes of interest in twelve different adult tissues. cDNA expression profiles for ten genes of interest (GOI) in twelve different tissues: 1: kidney, 2: muscle, 3: skin, 4: gut, 5: gill, 6: spleen, 7: brain, 8: heart, 9: testis, 10: liver, 11: eye and 12: pyloric caecum. Semi-quantitative levels of expression were calculated for each GOI as signal intensity relative to ubiquitin levels for three individual fish. Mean values (bars) of three individuals plus standard error are shown.

Additional file 3: Alignment of partial *mis* transcripts. Sequence alignment of *mis* cDNAs and amino acid residues based on sequence upstream from the C-terminal hormone portion of salmon MIS. A. The stop codon for *mis* (629 bp) is highlighted in red. The open reading frame of *mis* (396 bp) cDNA is in-frame for encoding the bioactive hormone portion of MIS and does not contain a stop codon. Two potential protease recognition motifs (RGQR AND RATR) for translated MIS (396 bp) are boxed. B. The stop codons for *mis* (629 bp) and *mis* (436 bp) are highlighted in red. Putative cleavage RLRR recognition sites are boxed.

Additional file 4: Proximal promoter sequences of the Atlantic salmon *dax1* gene. The potential binding elements of various transcription factors, TATA boxes, exon 1 and initiator methionine codon are labeled.

Additional file 5: Proximal promoter sequences of the Atlantic salmon *dax2* gene. The potential binding elements of various transcription factors, TATA boxes, exon 1 and initiator methionine codon are labeled.

Additional file 6: Proximal promoter sequences of the Atlantic salmon *mis* gene. The potential binding elements of various transcription factors, TATA boxes, exon 1 and initiator methionine codon are labeled.

Additional file 7: Proximal promoter sequences of the Atlantic salmon *foxl2b* gene. The potential binding elements of various transcription factors, TATA boxes, exon 1 and initiator methionine codon are labeled.

Additional file 8: Proximal promoter sequences of the Atlantic salmon pseudogenic *foxl2a* gene. Potential TATA boxes and the 5'-end portion of a region of sequence similarity shared with rainbow trout *foxl2* genes are labeled.

Additional file 9: Fluorescent *in situ* hybridization of *dax1*-containing BAC.

Additional file 10: Fluorescent *in situ* hybridization of *dax2*-containing BAC.

Additional file 11: Fluorescent *in situ* hybridization of *mis*-containing BAC.

Additional file 12: Primer oligonucleotide sequences used in RT-PCRs.

Abbreviations

ARE: androgen response element; **BMP:** bone morphogenetic protein; **CRE:** cyclic AMP response element; **EGR-1:** early growth response-1; **ERE:** estrogen response element; **OCT3/4:** octamer binding factor 3/4; **PPARE:** peroxisome proliferator-activated receptor element; **RXR/RAR:** retinoid X receptor/retinoic acid receptor; **SMAD:** mothers against decapentaplegic; **SP-1:** stimulatory protein-1; **TGFβ:** transforming growth factor β; **WNT:** wingless-related MMTV integration site.

Acknowledgements

This research was supported by Genome Canada, Genome BC and the Province of BC, and additionally by the Natural Sciences and Engineering Research Council of Canada (BFK, WSD). We are grateful to Marine Harvest

for the generous donation of the Atlantic salmon eggs and to Robert Devlin (F.O.C., WestVan Lab., West Vancouver, B.C.) for providing adult fish. We are particularly indebted to Carlo Biagi, Wendy Vandersteen (RHD), Yvonne Lai and Krzysztof Lubieniecki (WSD) for their adept extractions and handling of tissues. We appreciate very much the help of Brendon Campbell, Amy Hoare and Brian Ringwood for egg transport and care of the young fish (Animal Care Services, University of Victoria). We thank Roderick Haesevoets, Sherri Brady and Stephanie Puckett for clone sequencing (CBR DNA Sequencing Facility, University of Victoria).

Author details

¹Department of Biology, Centre for Biomedical Research, University of Victoria, Victoria, British Columbia, V8W 3N5, Canada. ²Tokyo University of Marine Science and Technology, Tateyama, Chiba, 294-0308, Japan.

³Biological Sciences, Washington State University, Vancouver, Washington, 98686-9600, USA. ⁴Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada.

Authors' contributions

KRVS designed and coordinated the study, performed GOI-specific PCRs of early development stage cDNAs, promoter analysis and drafted the manuscript. MY performed GOI-specific PCRs of adult tissue cDNAs and semi-quantitative analysis. RY performed BAC sequence data analysis and annotations. JdB performed transposable element analysis. LR performed hybridizations to isolate and identify GOI-containing BACs. SS, AR and EBR performed BAC clone preparation and sequencing. RBP performed FISH on BAC DNA. WSD and BFK obtained funding and contributed to experimental design and analysis. All authors read and approved the final manuscript.

Received: 22 July 2010 Accepted: 13 January 2011

Published: 13 January 2011

References

1. Yao HH, Matzuk MM, Jorgez CJ, Menke DB, Page DC, Swain A, Capel B: Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. *Dev Dyn* 2004, **230**(2):210-215.
2. Kim Y, Capel B: Balancing the bipotential gonad between alternative organ fates: a new perspective on an old problem. *Dev Dyn* 2006, **235**(9):2292-2300.
3. Park SY, Jameson JL: Minireview: transcriptional regulation of gonadal development and differentiation. *Endocrinology* 2005, **146**(3):1035-1042.
4. Yao HH: The pathway to femaleness: current knowledge on embryonic development of the ovary. *Mol Cell Endocrinol* 2005, **230**(1-2):87-93.
5. Vizziano D, Randuineau G, Baron D, Cauty C, Guiguen Y: Characterization of early molecular sex differentiation in rainbow trout, *Oncorhynchus mykiss*. *Dev Dyn* 2007, **236**(8):2198-2206.
6. Ijiri S, Kaneko H, Kobayashi T, Wang DS, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y: Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. *Biol Reprod* 2008, **78**(2):333-341.
7. von Schalburg KR, Rise ML, Brown GD, Davidson WS, Koop BF: A comprehensive survey of the genes involved in maturation and development of the rainbow trout ovary. *Biol Reprod* 2005, **72**(3):687-99.
8. von Schalburg KR, McCarthy SP, Rise ML, Hutson JC, Davidson WS, Koop BF: Expression of morphogenic genes in mature ovarian and testicular tissues: potential stem-cell niche markers and patterning factors. *Mol Reprod Dev* 2006, **73**(2):142-152.
9. von Schalburg KR, Cooper GA, Yazawa R, Davidson WS, Koop BF: Microarray analysis reveals differences in expression of cell surface and extracellular matrix components during development of the trout ovary and testis. *Comp Biochem Physiol* 2008, **3**:78-90.
10. Ikeda Y, Swain A, Weber TJ, Hentges KE, Zanaria E, Lalli E, Tamai KT, Sassone-Corsi P, Lovell-Badge R, Camerino G, Parker KL: Steroidogenic factor 1 and Dax-1 colocalize in multiple cell lineages: potential links in endocrine development. *Mol Endocrinol* 1996, **10**(10):1261-1272.
11. Wang ZJ, Jeffs B, Ito M, Achermann JC, Yu RN, Hales DB, Jameson JL: Aromatase (Cyp19) expression is up-regulated by targeted disruption of Dax1. *Proc Natl Acad Sci USA* 2001, **98**(14):7988-7993.
12. Tremblay JJ, Viger RS: Nuclear receptor Dax-1 represses the transcriptional cooperation between GATA-4 and SF-1 in Sertoli cells. *Biol Reprod* 2001, **64**(4):1191-1199.

13. Nachtigal MW, Hirokawa Y, Eyearnt-VanHouten DL, Flanagan JN, Hammer GD, Ingraham HA: Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell* 1998, **93**(3):445-454.
14. Baron D, Houlgatte R, Fostier A, Guiguen Y: Large-scale temporal gene expression profiling during gonadal differentiation and early gametogenesis in rainbow trout. *Biol Reprod* 2005, **73**(5):959-966.
15. Nakamoto M, Wang DS, Suzuki A, Matsuda M, Nagahama Y, Shibata N: Dax1 suppresses P450arom expression in medaka ovarian follicles. *Mol Reprod Dev* 2007, **74**(10):1239-1246.
16. Wang DS, Kobayashi T, Zhou LY, Paul-Prasanth B, Ijiri S, Sakai F, Okubo K, Morohashi K, Nagahama Y: FoxL2 up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. *Mol Endocrinol* 2007, **21**(3):712-725.
17. Yamaguchi T, Yamaguchi S, Hirai T, Kitano T: Follicle-stimulating hormone signaling and FoxL2 are involved in transcriptional regulation of aromatase gene during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. *Biochem Biophys Res Commun* 2007, **359**(4):935-940.
18. Chang X, Kobayashi T, Senthilkumar B, Kobayashi-Kajura H, Sudhakumari CC, Nagahama Y: Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). *Gen Comp Endocrinol* 2005, **141**(2):101-115.
19. Sekido R, Lovell-Badge R: Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 2008, **453**(7197):930-934.
20. Sun C, Nakatake Y, Ura H, Akagi T, Niwa H, Koide H, Yokota T: Stem cell-specific expression of Dax1 is conferred by STAT3 and Oct3/4 in embryonic stem cells. *Biochem Biophys Res Commun* 2008, **372**(1):91-96.
21. Khalfallah O, Rouleau M, Barbry P, Bardoni B, Lalli E: Dax-1 knockdown in mouse embryonic stem cells induces loss of pluripotency and multilineage differentiation. *Stem Cells* 2009, **27**(7):1529-1537.
22. Uhlenhaut NH, Jakob S, Anlag K, Eisenberger T, Sekido R, Kress J, Treier AC, Klugmann C, Klasen C, Holter NI, Riethmacher D, Schütz G, Cooney AJ, Lovell-Badge R, Treier M: Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* 2009, **139**(6):1130-1142.
23. von Schalburg KR, Yasuike M, Davidson WS, Koop BF: Regulation, expression and characterization of aromatase (*cyp19b1*) transcripts in ovary and testis of rainbow trout (*Oncorhynchus mykiss*). *Comp Biochem Physiol B* 2010, **155**(2):118-125.
24. Clipsham R, Niakan K, McCabe ER: Nr0b1 and its network partners are expressed early in murine embryos prior to steroidogenic axis organogenesis. *Gene Expr Patterns* 2004, **4**(1):3-14.
25. Galay-Burgos M, Gealy C, Navarro-Martín L, Piferrer F, Zanuy S, Sweeney GE: Cloning of the promoter from the gonadal aromatase gene of the European sea bass and identification of single nucleotide polymorphisms. *Comp Biochem Physiol A* 2006, **145**(1):47-53.
26. Pannetier M, Fabre S, Batista F, Kocer A, Renault L, Jolivet G, Mandon-Pépin B, Cotinot C, Veitia R, Pailhoux E: FOXL2 activates P450 aromatase gene transcription: towards a better characterization of the early steps of mammalian ovarian development. *J Mol Endocrinol* 2006, **36**(3):399-413.
27. Halm S, Rocha A, Miura T, Prat F, Zanuy S: Anti-Müllerian hormone (AMH/AMH) in the European sea bass: its gene structure, regulatory elements, and the expression of alternatively-spliced isoforms. *Gene* 2007, **388**(1-2):148-158.
28. Toffolo V, Belvedere P, Colombo L, Valle LD: Tissue-specific transcriptional initiation of the CYP19 genes in rainbow trout, with analysis of splicing patterns and promoter sequences. *Gen Comp Endocrinol* 2007, **153**(1-3):311-319.
29. Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987, **15**(17):7155-7174.
30. Zheng Z-M: Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J Biomed Sci* 2004, **11**:278-294.
31. Little TH, Zhang Y, Matulis CK, Weck J, Zhang Z, Ramachandran A, Mayo KE, Radhakrishnan I: Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol Endocrinol* 2006, **20**(4):831-843.
32. Südbeck P, Scherer G: Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. *J Biol Chem* 1997, **272**(44):27848-27852.
33. Gasca S, Canizares J, De Santa Barbara P, Mejean C, Poulat F, Berta P, Boizet-Bonhoure B: A nuclear export signal within the high mobility group domain regulates the nucleocytoplasmic translocation of SOX9 during sexual determination. *Proc Natl Acad Sci USA* 2002, **99**(17):11199-11204.
34. Haqq CM, King CY, Donahoe PK, Weiss MA: SRY recognizes conserved DNA sites in sex-specific promoters. *Proc Natl Acad Sci USA* 1993, **90**(3):1097-1101.
35. Hagen G, Müller S, Beato M, Suske G: Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic Acids Res* 1992, **20**(21):5519-5525.
36. Park GT, Morasso MI: Bone morphogenetic protein-2 (BMP-2) transactivates Dlx3 through Smad1 and Smad4: alternative mode for Dlx3 induction in mouse keratinocytes. *Nucleic Acids Res* 2002, **30**(2):515-522.
37. ASalBase. [http://www.asalbase.org].
38. de Boer JG, Yazawa R, Davidson WS, Koop BF: Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 2007, **8**:422-431.
39. Teixeira J, Maheswaran S, Donahoe PK: Müllerian inhibiting substance: an instructive developmental hormone with diagnostic and possible therapeutic applications. *Endocr Rev* 2001, **22**(5):657-674.
40. Vigier B, Forest MG, Eychenne B, Bézard J, Garrigou O, Robel P, Josso N: Anti-Müllerian hormone produces endocrine sex reversal of fetal ovaries. *Proc Natl Acad Sci USA* 1989, **86**(10):3684-3688.
41. Rodríguez-Marí A, Yan YL, Bremiller RA, Wilson C, Cañestro C, Postlethwait JH: Characterization and expression pattern of zebrafish Anti-Müllerian hormone (Amh) relative to sox9a, sox9b, and cyp19a1a, during gonad development. *Gene Expr Patterns* 2005, **5**(5):655-667.
42. Morinaga C, Saito D, Nakamura S, Sasaki T, Asakawa S, Shimizu N, Mitani H, Furutani-Seiki M, Tanaka M, Kondoh H: The hotei mutation of medaka in the anti-Müllerian hormone receptor causes the dysregulation of germ cell and sexual development. *Proc Natl Acad Sci USA* 2007, **104**(23):9691-9696.
43. Schmahl J, Kim Y, Colvin JS, Ornitz DM, Capel B: Fgf9 induces proliferation and nuclear localization of FGFR2 in Sertoli precursors during male sex determination. *Development* 2004, **131**(15):3627-3636.
44. Nachtigal MW, Ingraham HA: Bioactivation of Müllerian inhibiting substance during gonadal development by a kex2/subtilisin-like endoprotease. *Proc Natl Acad Sci USA* 1996, **93**(15):7711-7716.
45. Eimon PM, Harland RM: Effects of heterodimerization and proteolytic processing on Derrière and Nodal activity: implications for mesoderm induction in *Xenopus*. *Development* 2002, **129**(13):3089-3103.
46. Nishimatsu S, Thomsen GH: Ventral mesoderm induction and patterning by bone morphogenetic protein heterodimers in *Xenopus* embryos. *Mech Dev* 1998, **74**(1-2):75-88.
47. Ohe K, Tamai KT, Parvinen M, Sassone-Corsi P: DAX-1 and SOX6 molecular interplay results in an antagonistic effect in pre-mRNA splicing. *Dev Dyn* 2009, **238**(6):1595-1604.
48. Niakan KK, Davis EC, Clipsham RC, Jiang M, Dehart DB, Sulik KK, McCabe ER: Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol Genet Metab* 2006, **88**(3):261-271.
49. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, **126**(4):663-676.
50. Lalli E, Ohe K, Hindelang C, Sassone-Corsi P: Orphan receptor DAX-1 is a shuttling RNA binding protein associated with polyribosomes via mRNA. *Mol Cell Biol* 2000, **20**(13):4910-4921.
51. Lalli E, Ohe K, Latorre E, Bianchi ME, Sassone-Corsi P: Sexy splicing: regulatory interplays governing sex determination from *Drosophila* to mammals. *J Cell Sci* 2003, **116**(3):441-445.
52. Xu B, Yang WH, Gerin I, Hu CD, Hammer GD, Koenig RJ: Dax-1 and steroid receptor RNA activator (SRA) function as transcriptional coactivators for steroidogenic factor 1 in steroidogenesis. *Mol Cell Biol* 2009, **29**(7):1719-1734.
53. Callard GV, Tchoudakova AV, Kishida M, Wood E: Differential tissue distribution developmental programming, estrogen regulation and promoter characteristics of *cyp19* genes in teleost fish. *J Steroid Biochem Mol Biol* 2001, **79**(1-5):305-314.

54. Kanda H, Okubo T, Omori N, Niihara H, Matsumoto N, Yamada K, Yoshimoto S, Ito M, Yamashita S, Shiba T, Takamatsu N: **Transcriptional regulation of the rainbow trout CYP19a gene by FTZ-F1 homologue.** *J Steroid Biochem Mol Biol* 2006, **99**(2-3):85-92.
55. Wong TT, Ijiri S, Zohar Y: **Molecular biology of ovarian aromatase in sex reversal: complementary DNA and 5'-flanking region isolation and differential expression of ovarian aromatase in the gilthead seabream (*Sparus aurata*).** *Biol Reprod* 2006, **74**(5):857-864.
56. von Schalburg KR, Yazawa R, de Boer J, Lubieniecki KP, Goh B, Straub CA, Beetz-Sargent MR, Robb A, Davidson WS, Devlin RH, Koop BF: **Isolation, characterization and comparison of Atlantic and Chinook salmon growth hormone 1 and 2.** *BMC Genomics* 2008, **9**:522-534.
57. Image J. [<http://rsbweb.nih.gov/ij/>].
58. Primer3. [<http://frodo.wi.mit.edu/>].
59. Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Høyheim B: **A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects.** *BMC Genomics* 2005, **6**(1):50-57.
60. CHORI - Children's Hospital Oakland Research Institute. [<http://bacpac.chori.org/highdensity.htm>].
61. iCE - internet Contig Explorer. [<http://www.bcgsc.ca/platform/bioinfo/software/ice>].
62. Ewing B, Green P: **Base-calling of automated sequencer traces using PHRED. II Error probabilities.** *Genome Res* 1998, **8**:186-194.
63. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment.** *Genome Res* 1998, **8**:175-185.
64. PHRAP. [<http://bozeman.mbt.washington.edu>].
65. Gordon D, Abajan C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**:195-202.
66. EMBOSS. [<http://emboss.sourceforge.net/>].
67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
68. Sonnhammer EL, Durbin R: **A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis.** *Gene* 1995, **167**:GC1-10.
69. Phillips RB, Nichols KM, DeKoning JJ, Morasch MR, Keatley KA, Rexroad C, Gahr SA, Danzmann RG, Drew RE, Thorgaard GH: **Assignment of rainbow trout linkage groups to specific chromosomes.** *Genetics* 2006, **174**(3):1661-1670.
70. Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Koop BF, Danzmann RG, Davidson WS: **Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*).** *BMC Genet* 2009, **10**:46-56.
71. Sun C, Nakatake Y, Akagi T, Ura H, Matsuda T, Nishiyama A, Koide H, Ko MS, Niwa H, Yokota T: **Dax1 binds to Oct3/4 and inhibits its transcriptional activity in embryonic stem cells.** *Mol Cell Biol* 2009, **29**(16):4574-4583.
72. Wilson M, Koopman P: **Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators.** *Curr Opin Genet Dev* 2002, **12**(4):441-446.
73. Tremblay JJ, Drouin J: **Egr-1 is a downstream effector of GnRH and synergizes by direct interaction with Ptx1 and SF-1 to enhance luteinizing hormone beta gene transcription.** *Mol Cell Biol* 1999, **19**(4):2567-2576.
74. Yang BY, Chan KM, Lin CM, Chen TT: **Characterization of rainbow trout (*Oncorhynchus mykiss*) growth hormone 1 gene and the promoter region of growth hormone 2 gene.** *Arch Biochem Biophys* 1997, **340**(2):359-368.

doi:10.1186/1471-2164-12-31

Cite this article as: von Schalburg *et al.*: Regulation and expression of sexual differentiation factors in embryonic and extragonadal tissues of Atlantic salmon. *BMC Genomics* 2011 **12**:31.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

