



HAL
open science

Characterization of rainbow trout gonad, brain and gill deep cDNA repertoires using a Roche 454-Titanium sequencing approach

Aurélie Le Cam, Julien Bobe, Olivier Bouchez, Cédric Cabeau, Olivier Kah, Christophe C. Klopp, Jean-Jacques Lareyre, Isabelle Leguen, Jérôme Lluch, Jérôme Montfort, et al.

► To cite this version:

Aurélie Le Cam, Julien Bobe, Olivier Bouchez, Cédric Cabeau, Olivier Kah, et al.. Characterization of rainbow trout gonad, brain and gill deep cDNA repertoires using a Roche 454-Titanium sequencing approach. *Gene*, 2012, 500 (1), pp.32-39. 10.1016/j.gene.2012.03.053 . hal-01205069

HAL Id: hal-01205069

<https://hal.science/hal-01205069>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Post-Print: Gene 2012 500 (1), 32-39. DOI : 10.1016/j.gene.2012.03.053**

2

3 **Characterization of rainbow trout gonad, brain and gill deep cDNA**
4 **repertoires using a Roche 454-Titanium sequencing approach**

5

6 **Aurélie Le Cam^{a,1}, Julien Bobe^{a,4}, Olivier Bouchez^{d,2}, Cédric Cabau^{b,3}, Olivier Kah^{c,4},**
7 **Christophe Klopp^{b,3}, Jean-Jacques Lareyre^{a,4}, Isabelle Le Guen^{a,4}, Jérôme Lluch^{d,2},**
8 **Jérôme Montfort^{a,3}, Francois Moreews^{b,3}, Barbara Nicol^{a,4}, Patrick Prunet^{a,4}, Pierre-**
9 **Yves Rescan^{a,4}, Arianna Servili^{c,4}, and Yann Guiguen^{a,§,5}**

10

11 ^a INRA, UR1037 LPGP Fish Physiology and Genomics, F-35000 Rennes, France.

12 ^b INRA, Sigenae UR875 BIA, F-31326 Castanet-Tolosan Cedex, France.

13 ^c UMR CNRS 6026, IFR 140, Université de Rennes 1, France.

14 ^d INRA, Plateforme Génomique, Génopole Toulouse / Midi-Pyrénées, Auzeville, F-31326
15 Castanet-Tolosan, France.

16

17 [§]Corresponding author: Tel.: 00 33 2 02 23 48 50 09, Fax: 00 33 2 02 23 48 50 20

18 E-mail address: yann.guiguen@rennes.inra.fr

19

20 **Keywords:** transcriptome; next-generation sequencing (NGS); fish; salmonids;
21 rainbow trout.

22

23 ¹ A.L.C. produced the double-stranded cDNA, supervised the construction and
24 sequencing of the 454 libraries, and participated in the writing of the manuscript.

25 ²O.B. and J.L. aided in the construction of the 454 libraries and performed the
26 sequencing reactions.

27 ³J.M., C.C., F.M. and C.K. developed the rainbow trout contig browser and performed
28 the bioinformatic analyses.

29 ⁴J.B., B.N., O.K., A.S., I.L.G., P.P., J.J.L., and P.Y.R. analyzed the tissue-specific
30 repertoires, performed QPCR analysis and drafted parts of the manuscript.

31 ⁵Y.G. designed and supervised the study, participated in data analysis and drafted the
32 final manuscript.

33

34 *Abbreviations:* ACTH, adrenocorticotrophic hormone; *amh*, anti-müllerian hormone
35 gene; AMP, adenosine monophosphate; bHLH, basic-helix-loop-helix; Blbp, Brain
36 Lipid Binding protein; bp, base pair(s); *Btg1*, B Cell Translocation gene 1; cDNA,
37 DNA complementary to RNA; CE, cornified envelope; *dazl*, deleted in azoospermia
38 gene; DDD, Digital Differential Display; *dmrt1*, double-sex and mab-3 related
39 transcription factor gene 1; EST, Expressed Sequenced Tag; *fabp7a*, fatty acid
40 binding protein gene 7a; *gap43*, growth associated protein 43 gene; Gb, giga base;
41 *gdf9*, growth and differentiation factor gene 9; *glast*, glutamate aspartate transporter
42 gene; *gsdf*, gonadal soma-derived factor gene; Na/K-ATPase, sodium-potassium
43 adenosine triphosphatase; *nrrn1*, neuritin gene 1; *neurod1*, Neurogenic differentiation
44 1 gene; NGS, Next-Generation Sequencing; *rec8*, Meiotic recombination protein
45 REC8 homolog; RNA-seq, RNA sequencing; *ocpi*, oocyte cysteine protease inhibitor
46 gene; POMC, pro-opiomelanocortin; *pou5f1*, POU domain class V transcription factor
47 1 gene; RyR1, Ryanodine receptor 1; SAGE, Serial analysis of gene expression;
48 *Sox9a*, SRY-box containing gene 9a; Smyd1, SRA, NCBI Sequence Read Archive;
49 *sycp 1, 2 and 3*, synaptonemal complex protein; *rhcg2*, Rh type C glycoprotein2a

50 gene; tudor1, tudor domain containing 1; TBX19, T-box transcription factor; *tbx19*,
51 T-box transcription factor gene 19; *tuba1*, tubulin alpha1 gene; vasa, vasa homolog;
52 *wnt11*, Wingless-type MMTV integration site family member 11; *zar1*, zygote arrest
53 gene 1; ZP, zona pellucida; *zpc1* and 5; *zpax*, egg envelope component ZPAX; *zar1*,
54 zygote arrest gene 1.

55 **Abstract**

56 Rainbow trout, *Oncorhynchus mykiss*, is an important aquaculture species worldwide
57 and, in addition to being of commercial interest, it is also a research model organism
58 of considerable scientific importance. Because of the lack of a whole genome
59 sequence in that species, transcriptomic analyses of this species have often been
60 hindered. Using next-generation sequencing (NGS) technologies, we sought to fill
61 these informational gaps. Here, using Roche 454-Titanium technology, we provide
62 new tissue-specific cDNA repertoires from several rainbow trout tissues. Non-
63 normalized cDNA libraries were constructed from testis, ovary, brain and gill rainbow
64 trout tissue samples, and these different libraries were sequenced in 10 separate half-
65 runs of 454-Titanium. Overall, we produced a total of 3 million quality sequences
66 with an average size of 328 bp, representing more than 1 Gb of expressed sequence
67 information. These sequences have been combined with all publicly available rainbow
68 trout sequences, resulting in a total of 242,187 clusters of putative transcript groups
69 and 22,373 singletons. To identify the predominantly expressed genes in different
70 tissues of interest, we developed a Digital Differential Display (DDD) approach. This
71 approach allowed us to characterize the genes that are predominantly expressed
72 within each tissue of interest. Of these genes, some were already known to be tissue-
73 specific, thereby validating our approach. Many others, however, were novel

74 candidates, demonstrating the usefulness of our strategy and of such tissue-specific
75 resources. This new sequence information, acquired using NGS 454-Titanium
76 technology, deeply enriched our current knowledge of the expressed genes in rainbow
77 trout through the identification of an increased number of tissue-specific sequences.
78 This identification allowed a precise cDNA tissue repertoire to be characterized in
79 several important rainbow trout tissues. The rainbow trout contig browser can be
80 accessed at the following publicly available web site (<http://www.sigenae.org/>).

81

82 **1. Introduction**

83 Rainbow trout, *Oncorhynchus mykiss*, is an important aquaculture species worldwide
84 and, in addition to being of commercial interest, it is also a model research organism
85 of considerable scientific importance (reviewed by Thorgaard et al. (Thorgaard et al.,
86 2002)). Many genomic resources are available for this species, including extensive
87 classical Expressed Sequenced Tag (EST) repertoires (Rexroad et al., 2003; Govoroun
88 et al., 2006; Koop et al., 2008); however, because of the lack of a currently available
89 whole genome sequence in that species, genomic analysis has been inhibited by the
90 limited knowledge of the rainbow trout transcriptome. The recent development of
91 next-generation sequencing (NGS) has been a major technological breakthrough
92 (Metzker, 2010). This technique provides new opportunities to better characterize
93 gene repertoires, even in non-model species for which a whole genome sequence is
94 unavailable (Crawford et al., 2010; Mizrachi et al., 2010; Fraser et al., 2011). These
95 NGS technologies are now widely used for transcript expression profiling, and this
96 approach, also referred to as RNA-seq, is progressively replacing classical techniques
97 such as SAGE analysis or DNA microarrays (Costa et al., 2010).

98 Using NGS and 454-Titanium technology, this study sought to complement the
99 current knowledge of the rainbow trout transcriptome by providing new tissue-
100 specific cDNA repertoires from different rainbow trout tissues. 454-Titanium
101 technology has already been used to characterize rainbow trout transcripts (Salem et
102 al., 2010); however, tissue-specific information is currently lacking. We thus
103 developed sequencing resources and bioinformatics tools to initiate expression
104 profiling analyses on the vast new collection of tissue-specific sequence information
105 generated by this study. These first analyses demonstrate that our approach is reliable.
106 These new resources may now be utilized to search for more subtle genetic regulation
107 events, allowing the identification of genes that are predominantly expressed under
108 different physiological conditions.

109 **2. Materials and Methods**

110 **2.1 Tissue sampling and RNA extraction**

111 Total RNA was extracted from the collected tissues (gill, testis, brain and ovary) with
112 TRIzol™ reagent (Invitrogen, Cergy-Pontoise, France) according to the
113 manufacturer's instructions. The total RNA yield was estimated using a Nanodrop
114 1000 spectrophotometer (Labtech, Palaiseau, France), and RNA quality was examined
115 with an Agilent Bioanalyzer (Agilent Technologies, Massy, France). Only good
116 quality total RNA samples (RNA Integrity Number, RIN > 8) were selected for
117 further use. The total RNA from several extractions was pooled to minimize the
118 impact of biological variations and to obtain the amount of RNA required for library
119 preparation. Because of the presence of mucus, the gill sample pools were purified
120 with the NucleoSpin® RNA II kit (Macherey-Nagel, Hoerd, France) according to the
121 "Clean-up of RNA from reaction mixtures" protocol. This purification was made from

122 200 µg of the total RNA mix, and the elution was performed twice with 70 µl of H₂O
123 (2 successive elutions). mRNA was then isolated from 250 µg of total RNA using the
124 Nucleotrap® mRNA kit (Macherey-Nagel, Hoerd, France) as described in the
125 manufacturer's manual, with the exception of the elution procedure. Here, 40 µl of
126 pre-warmed (68°C) H₂O were used per 40 µl of Oligo(dT) latex beads. Three rounds
127 of elution were performed. The quality of the mRNA and the depletion of ribosomal
128 RNA were determined with the Agilent Bioanalyzer (Agilent Technologies, Massy,
129 France).

130 **2.2 Double stranded-cDNA preparation**

131 Double-stranded cDNA was prepared as previously described (Leroux et al., 2010)
132 using modifications of the SuperScript Double-Stranded cDNA Synthesis Kit
133 (Invitrogen, Cergy-Pontoise, France) protocol. Briefly, we used a modified primer (5'
134 GAGAGAGAGACTGGAG(T)16VN 3') containing the GsuI restriction site and
135 methylated dCTP (10 mM dm5CTP, Fermentas, Saint-Remy-les-Chevreuse, France)
136 during the first-strand cDNA synthesis. The first strand was synthesized from 3 µg of
137 mRNA. The next steps (second strand synthesis, RNase I digestion,
138 phenol/chloroform extraction and ethanol precipitation) were performed according to
139 the manufacturer's instructions. The pellets were resuspended in 9 µl of water. Three
140 to five independent syntheses were made for each sample and pooled to obtain a total
141 of 3.5 to 5 µg of total cDNA. Next, the double-stranded cDNA was digested by GsuI
142 (5 U) for 90 min at 30°C to remove the polyA tail. The cDNA samples were
143 quantified using the Quant-iT PicoGreen kit (Invitrogen, Cergy-Pontoise, France).
144

145 **2.3 454 The Roche sequencing method**

146 After cDNA nebulization, the libraries were prepared using the GS DNA Library
147 Preparation Kit according to the manufacturer's recommendations (Roche Applied
148 Science, Meylan, France); 3.5 to 5 µg of each cDNA sample were used to generate
149 the libraries. Sheared DNA was ligated to the linker for emPCR and sequencing. The
150 final libraries were quantified with the SlingShot™ kit using the Fluidigm® Digital
151 Array according to the manufacturer's instructions. Pyrosequencing using 454/Roche
152 GS FLX Titanium chemistry was performed with the GS Titanium LV emPCR and
153 sequencing kits according to the manufacturer's instructions. Each sample was
154 sequenced on one region of a two-region Picotiterplate.

155

156 **2.4 Real-time PCR analysis**

157 Real-time PCR was performed using the Step One Plus system (Applied Biosystems,
158 Foster City, USA) as previously described (Desvignes et al., 2009). For each tissue,
159 three separate reverse transcription reactions were conducted using 2 µg of DNase-
160 treated RNA that originated from three different fishes. Control reactions were
161 performed in the absence of reverse transcriptase. RT products and control reaction
162 samples were diluted 1/25, and 4 µl was used for each real-time PCR reaction. Real-
163 time PCR was performed with the Fast SYBR Green Master Mix (Applied
164 Biosystem) and either 300 or 600 nm of each primer. All real-time PCR reactions
165 were performed in duplicate. After amplification, a melting curve was performed
166 according to the manufacturer recommendations to check the specificity of
167 amplification. The relative expression of each gene was normalized to the level of 18S
168 gene amplification and was calculated as the percentage of the highest expression

169 level recorded for each gene. The sequences of the primers used are listed in
170 Supplementary file 2.

171 **2.5 *In situ* hybridization**

172 Digoxigenin-labeled antisense RNA probes were synthesized from a PCR-amplified
173 template (EST clone: BX085208 containing the rainbow trout β -enolase sequence)
174 using T3 RNA polymerase. *In situ* hybridizations were performed on transverse
175 sections of rainbow trout embryos as previously described (Gabillard et al., 2003),
176 with minor modifications.

177 **2.6 Bioinformatic analysis of the sequence data**

178 The initial sequence set used for this transcriptome assembly was composed of
179 publicly available mRNA sequences from Genbank, Sanger ESTs from dbEST, GS-
180 Flx 454 RNA-Seq from SRA (including a complete rainbow trout muscle 454 dataset
181 publicly available under the SRA accession number SRA029716) and the locally
182 produced 454 Titanium RNA-Seq reads that are available under the SRA accession
183 number SRA026493 (Table 1). The reads were cleaned to remove sequences
184 composed of less than 100 base pairs, polyA tails, and *E. coli* and yeast contaminating
185 sequences using SeqClean (<http://compbio.dfci.harvard.edu/tgi/software/>). To ease the
186 assembly process, our next step was to remove redundancy using the cd-hit-est
187 software (Li and Godzik, 2006). This software discarded those reads that were
188 included in other reads. A threshold of 0.98 was chosen, producing a non-redundant
189 representative read set (parameters -M 30000 -d 0 -n 10 -l 11 -r 1 -p 1 -g 1 -G 1 -c
190 0.98 -aS 0.5 -T 8). The discarded reads were repositioned in the resultant contigs after
191 assembly using the inclusion coordinates. The assembly was performed using TGICL
192 (Pertea et al., 2003), which first clusters the reads using megablast, given a similarity

193 threshold (-p 96) and a similarity length (-l 60), and then, using the mRNAs as
194 backbones, splits the clusters larger than a given size (-s 100000). The program then
195 produces alignment files and consensus sequences with CAP3. All of the 454 singlets
196 were removed from the final results, which contained 219,814 contigs and 22,373
197 singlets. The contigs and singlets were annotated by similarity searches using a large
198 set of nucleotide and protein reference databases, including UniProtKB/Swiss-Prot,
199 RefSeq protein, RefSeq RNA, PFAM, DFCI gene indices, UniGene clusters and
200 Ensembl transcripts for numerous species. The full list of databases is available on the
201 home page of the assembly ([http://public-](http://public-contigbrowser.siginae.org:9090/Oncorhynchus_mykiss/index.html)
202 [contigbrowser.siginae.org:9090/Oncorhynchus_mykiss/index.html](http://public-contigbrowser.siginae.org:9090/Oncorhynchus_mykiss/index.html)). The e-value
203 thresholds that were used to determine similarities were dependent upon the database
204 that was used; e^{-10} was used for transcript databases, e^{-2} was used for species-specific
205 nucleic databases, and e^{-5} was used for all others. The UniProtKB/Swiss-Prot
206 similarity search was performed using a bi-directional best hit approach to identify
207 putative orthologs, and the resulting annotation character string had a [BBH] tag.
208 Variation detection was also performed using the contigs. The in-house detection
209 script screens the CAP3 alignment files to find substitutions that occur at least twice
210 at a position with a depth of at least 6 sequences and that have no other substitution in
211 a twenty base pair window around this position. All of the resulting files from the
212 assembly, the annotation and the variation detection steps were then reformatted to
213 populate the contig browser database, which was used to show the contig layout, and
214 the BioMart (Haider et al., 2009) database, which enables a large variety of queries.
215 The contig browser was built using the Ensembl 40 web code (Birney et al., 2006)
216 and provides such new features as the Digital Differential Display view and the Venn

217 diagram view. The contig and singlet FASTA files and the BioMart tabular files can
218 be downloaded from the web server.

219

220 **3. Results and Discussion**

221 **3.1 The generation of tissue-specific transcriptomic resources using a RNA-** 222 **Seq approach**

223 Eight non-normalized cDNA libraries were constructed from rainbow trout testis,
224 ovary, brain, and gill tissue samples and were subsequently sequenced in 10 separate
225 half-runs using 454-Titanium (Table 1). Overall, we produced a total of 3 million
226 good-quality sequences (testis, 1,234,260; ovary, 738,510; brain, 481,363; and gill,
227 613,223) with an average sequence size of 328 bp. Collectively, these sequences
228 constitute more than 1 Gb of expressed sequence information. This raised the total
229 number of expressed sequences that are publicly available for rainbow trout to 4.8
230 millions. In contrast to previous large NGS datasets produced in the rainbow trout
231 (Salem et al., 2010), our strategy also provided tissue-specific information that was
232 not previously available for this species. All of the acquired sequences were deposited
233 in the NCBI Sequence Read Archive (SRA) and were assembled along with the
234 publicly available rainbow trout sequences. This resulted in a total of 242,187 clusters
235 of putative transcripts (contigs) and 22,373 singletons. These sequences and their
236 assembly can be accessed at the following publicly available web site
237 (<http://www.sigenae.org/>). This browser is based on an Ensembl architecture
238 (Hubbard et al., 2002; Stabenau et al., 2004). One can utilize many classical features
239 to browse all of the computed information on each contig and can extract large user-
240 specific datasets using the Biomart generic interface (Smedley et al., 2009). *In silico*

241 strategies have often been used to mine sequence databases and discover new genes
242 with a predominant expression pattern in one tissue or a group of tissues (Murray et
243 al., 2007); however, the use of such strategies has been restricted to species with large
244 available sequence datasets. In non-model species, this approach has, in contrast,
245 often been limited by the rather low number of sequences available. With the recent
246 advent of NGS, this limitation is no longer critical. Information on many non-model
247 species is now more accessible due to RNA-Seq approaches (Wilhelm and Landry,
248 2009), allowing for the characterization of genes with differential expression patterns
249 (Crawford et al., 2010; Mizrachi et al., 2010; Fraser et al., 2011). Therefore, we
250 applied this approach to our own dataset to validate its utility.

251 **3.2 The use of the 454 tissue-specific transcriptomic resource for the** 252 **identification of tissue-predominant transcripts**

253 A Digital Differential Display (DDD) analysis tool was implemented within the
254 rainbow trout sequence browser to characterize contigs with a predominant expression
255 pattern. DDD analysis was subsequently performed using the 454 sequencing data
256 that was obtained from this study. For each tissue of interest (ovary, testis, brain, gill
257 and muscle), we compared the number of reads contained in each of the contigs in the
258 tissue of interest to all of the remaining tissue libraries. Some of these examples are
259 discussed below, and the DDD software is available within the rainbow trout contig
260 browser to allow for further user-specific queries. To confirm that this strategy is
261 effective for the identification of genes that are predominantly expressed in a given
262 tissue, we carried out qPCR validation for 9 genes (see Fig 1 and Tables 2 to 6).

263 3.2.1 The identification of genes that are preferentially expressed in the rainbow trout 264 ovary

265 When performing a DDD analysis on the trout ovarian samples using a 10-fold cut-
266 off, 614 contigs were found to be over abundant in the ovarian/oocyte libraries in
267 comparison to the non-ovarian libraries. Among those 614 contigs, 438 contigs were
268 found only 1 or 2 times in the non-ovarian libraries (see additional file 1). In addition,
269 432 contigs were found at least 10 times in the ovarian libraries but were never
270 detected in the non-ovarian libraries. Among the most overrepresented genes found in
271 the ovarian libraries, we identified several ZP domain-containing genes. Our results
272 show that, in agreement with existing data from medaka (*Oryzias latipes*) (Kanamori
273 et al., 2003), the *zpc1*, *zpc5*, and *zpac* genes are predominantly expressed in the
274 ovaries in the rainbow trout (Table 2). Similarly, several genes that are known to be
275 specifically or predominantly expressed in rainbow trout oocytes or ovaries were
276 identified. In agreement with existing data on rainbow trout (Bobe et al., 2008),
277 zygote arrest 1 (*zar1*) and growth and differentiation factor 9 (*gdf9*) exhibited a very
278 strict ovarian-predominant profile. Similarly, oocyte cysteine protease inhibitor (*ocpi*)
279 was dramatically overexpressed in the ovarian libraries, again in agreement with
280 existing data (Bobe and Goetz, 2001). In addition to these previously documented
281 patterns, which validate our approach, the DDD analysis has led to the identification
282 of many genes that also putatively exhibit predominant or specific ovarian and/or
283 oocyte expression. Interestingly, our data also revealed the predominant expression of
284 several ovarian genes in the rainbow trout that are known to be strongly expressed in
285 the zebrafish (*Danio rerio*) oocyte, such as bucky ball (Marlow and Mullins, 2008), or
286 in the medaka, such as the transcription factor *pou5f1* (Sanchez-Sanchez et al., 2010)
287 (Fig. 1 and Table 2). Furthermore, we have shown the ovarian-predominant
288 expression of *wnt11* (Fig. 1 and Table 2). In sexually mature rainbow trout, *wnt11*
289 mRNA was detected in all assayed tissues. In comparison to skin, which had the

290 lowest level of *wnt11* expression, we observed a 15,000-fold higher expression of
291 *wnt11* in post-ovulatory ovaries and a 1,700-fold higher expression in oocytes (Fig.
292 1). Our results are consistent with previous studies reporting the expression of *wnt11*
293 in mouse granulosa cells and oocytes (Harwood et al., 2008). Interestingly, *wnt11* is a
294 maternal transcript. Maternal *wnt11* mRNA has been shown to accumulate in
295 *Xenopus* oocytes (Ku and Melton, 1993), where it plays a role in axis formation in the
296 embryo (Tao et al., 2005). Together, these results suggest the participation of
297 maternally inherited *wnt11* mRNA in teleost early development. The DDD analysis
298 also allowed the identification of novel oocyte-predominant genes in rainbow trout
299 (see Additional File 1). Such genes deserve further attention.

300 3.2.2 The identification of genes that are preferentially expressed in trout testes

301 When performing a DDD analysis of the trout testicular samples, we characterized
302 503 contigs that were only expressed in the testes, with a minimum of 10 reads used
303 as a threshold. In addition, 590 contigs were significantly overexpressed in the testes,
304 with a 10-fold minimum change and a threshold of 10 reads (see additional file 1). As
305 expected, numerous contigs corresponded to transcripts that are known to be
306 preferentially expressed in the testes of several vertebrate species, including the
307 rainbow trout (Rolland et al., 2009). These transcripts are involved in the sexual
308 differentiation of the male gonad (*Sox9a*, *dmrt1*, *amh*) and in specific testicular
309 functions such as the regulation of germ cell proliferation (*amh*, *gsdf*), meiosis (*syce1*,
310 *syce2* and *syce3*, *rec8*, *dazl*, *vasa*, *tudor1*), and spermiogenesis (*Dynein intermediate*
311 *chain 1*, *microtubule associated protein 6*). These observations illustrate the reliability
312 of the DDD method in the analysis of the 454 sequencing data. Interestingly, our
313 analysis revealed several transcripts that were unexpectedly overexpressed in trout
314 testes. For instance, one contig sequence encodes the trout counterpart of the

315 evolutionarily conserved TBX19 transcription factor (Table 3). TBX19 is a member
316 of a multigene superfamily that harbors a conserved DNA binding domain called the
317 T-box (see (Plageman and Yutzey, 2005) for review). The TBX transcription factors
318 are involved in the regulation of developmental processes. In humans and mice,
319 TBX19 expression is restricted to two pituitary cell types that express pro-
320 opiomelanocortin (POMC). In trout, enhanced TBX19 expression was also detected in
321 the pituitary along with the testes (Fig. 1). A mutation of the *tbx19* gene in humans
322 results in the impaired differentiation of the POMC cell lineages as well as ACTH
323 deficiency (Vallette-Kasic et al., 2007). An understanding of the biological relevance
324 of the enhanced expression of the *tbx19* gene in trout testes will require further
325 investigation. We also noted that several transcripts that originated from transposable
326 elements were among the most frequently detected transcripts in trout testes. These
327 transcripts encode reverse transcriptases, transposases and partial envelope proteins
328 (Table 3). Transposable elements have long been considered “junk” DNA, but there is
329 accumulating evidence that they may have been recruited by the progenote to fulfill
330 important biological functions (Volf, 2006, Kleene et al., 1998). For instance,
331 retroviral envelope proteins are required for the cell contact and fusion of human
332 trophoblast cells (Vargas et al., 2009). Moreover, the transposition of active mobile
333 elements in the vicinity of any gene can modulate gene expression, and it has been
334 proposed that these events may participate in the epigenetic regulation of genes in
335 somatic cells (see (Muotri et al., 2007, Collier and Largaespada, 2007) for reviews).
336 In addition, transposition is likely in germ cells, and this could be considered a major
337 source of genetic variation that can be passed on to further generations. The
338 identification of mobile transposable elements in the testes contributes to our
339 understanding of the mechanisms that lead to dynamic changes in the genome

340 structure and may provide an explanation for the observed differences in the
341 reproductive capacity of individuals.

342 3.2.3 The identification of genes that are preferentially expressed in muscle

343 When performing a DDD analysis for the trout muscle samples using a 10-fold cut-
344 off, 393 contigs were found to be overabundant in the white muscle libraries in
345 comparison to the non-muscle libraries (see Additional File 1). Among these, 266
346 contigs were found only 1 or 2 times in the non-muscle libraries. Of the top ten
347 overabundant contigs (exhibiting overexpression ranging from 2517- to 793-fold), six
348 were found to encode sarcomeric proteins such as myosins and troponins (Table 4).
349 Other sarcomeric protein-encoding genes were found to be overexpressed in the
350 muscle libraries including Lim domain-binding protein 3, nebulin and myozenin-1.
351 Additionally, sarcosin (*kbtbd10*, for Kelch repeat and BTB domain-containing protein
352 10, see Fig. 1), which is involved in the assembly of myofibrils (Greenberg et al.,
353 2008), and Stars, an actin-binding Rho-activating protein that promotes actin
354 polymerization (Arai et al., 2002), were predominantly expressed in the muscle
355 library. The DDD analysis is also strongly validated by the observation that most of
356 the enzymes identified in the muscle libraries were known to be muscle-specific
357 isoforms or enzymes. Among these enzymes were creatine kinase, Phosphoglycerate
358 mutase (*pgam2*) (see Fig. 1), Adenylate kinase isoenzyme 1, L-lactate dehydrogenase
359 A, Fructose-bisphosphate aldolase A, 6-phosphofructokinase, Glycogen
360 phosphorylase, AMP deaminase 1 and β -enolase (Table 4). Consistently, the *in situ*
361 hybridization of β -enolase revealed a strong hybridization signal specifically in
362 glycolytic white fast muscle fibers (Fig. 2). In agreement with existing data, the
363 histone methyl transferase Smyd1 (Tan et al., 2006) and the antiproliferative protein
364 Btg1 (Busson et al., 2005), which both regulate muscle differentiation, were clearly

365 overexpressed in the muscle libraries. In addition, DDD analysis has led to the
366 identification of the genes encoding Ryanodine receptor 1 (RyR1) and calsequestrin,
367 which are known to participate in the formation of a complex that is required for Ca²⁺
368 ion release from the sarcoplasmic reticulum. This is an essential step in muscle
369 contraction (Zhang et al., 1997).

370 3.2.4 The identification of genes that are preferentially expressed in the brain

371 When performing a DDD analysis with a 10-fold cut-off in the trout brain samples
372 that were taken 40 days post-fertilization, 313 contigs were found to be overabundant
373 in the brain libraries in comparison to the non-brain libraries. In addition, 365 contigs
374 were found at least 10 times in the brain libraries but were never detected in the non-
375 brain libraries. At the stage examined, the brain is organized and regionalized, but it is
376 still actively growing. Accordingly, many of the genes that were identified in the
377 brain samples are known to be involved in neurogenesis and/or neuronal
378 differentiation (Table 5). Some of these genes (*glast* and *fabp7a*) are likely to be
379 expressed in the radial glial cells that are the main progenitors at this stage, whereas
380 others are involved in neuronal growth or neuron differentiation (synapsin,
381 neuromodulin, neuritin, and neuroD). Among the genes expected to be strongly
382 expressed in the developing brain is *fabp7a*, and Brain Lipid Binding Protein (Blbp,
383 see Fig. 1). *fabp7a*, is a member of the fatty acid-binding protein family that was
384 retrieved several times. This finding is of particular interest because, in mice, Blbp is
385 a recognized marker of radial glial cells, which are known to be progenitor cells in the
386 developing mammalian brain (Anthony and Heintz, 2008). Recently, it was shown
387 that *blbp* is also expressed in the brains of larval and adult fish (Diotel et al., 2010;
388 Marz et al., 2010). Ependymin (*epd1*) is a meningeal-derived extracellular matrix
389 glycoprotein that was also overrepresented in the brain in comparison to other tissues.

390 Little is known about the function of such proteins; however, they are thought to
391 increase in abundance during periods of neuronal plasticity in mammals (Rother et al.,
392 1995). The hypothesis that ependymin is also involved in neurogenesis in fish is of
393 particular interest and will need to be further examined. The gene *slc1a2b* encodes a
394 glial high-affinity glutamate transporter that was previously identified in a screen of
395 the zebrafish hypothalamus (Toro et al., 2009). In mammals, these glutamate
396 transporters are strongly expressed in astrocytes. In these cells, the excess of
397 glutamate is quickly removed from the extracellular space, where the glutamate level
398 is kept low. Such transporters are also expressed in the radial glial cells of rodents
399 and, likely, of fishes, although the latter has not been documented. Neuromodulin
400 (*gap43*) was also found to be expressed only in the brain (see Fig. 1), in agreement
401 with its recognized functions as a 'plasticity' protein that is highly expressed in
402 neuronal growth cones during development (Udvardia et al., 2001). Neurogenic
403 differentiation 1 (*neurod1*) is a member of the NeuroD family of basic-helix-loop-
404 helix (bHLH) transcription factors that are involved in brain differentiation and
405 patterning. Tubulin alpha chains (*tuba1*) were also highly represented, which was
406 expected based on their known roles in neurotubule formation (Hieber et al., 1998).
407 Neuritin and synapsin, respectively, are involved in neurite growth and synapse
408 formation. Neuritin (*nrn1*) is expressed in the postmitotic neurons of the developing
409 nervous system and in neuronal structures that are undergoing plasticity in adults
410 (Naeve et al., 1997). Interestingly, neuritin was also represented in the libraries
411 constructed from non-nervous tissues. Synapsins are a family of proteins that are
412 implicated in the regulation of neurotransmitter release at synapses. Recent data
413 suggest that synapsins are not essential for synaptic function, but instead serve
414 important modulatory functions at the synapse (Kile et al., 2010).

415 3.2.5 The identification of genes that are preferentially expressed in the gills
416 When comparing gill with other tissues using a DDD analysis with a 10 fold cut-off,
417 we found 313 contigs that were overrepresented in the gill libraries. Moreover, 365
418 contigs were only expressed in the gill libraries (see Additional File 1). As expected,
419 several of these genes are known to be specifically or predominantly expressed in the
420 gills (Table 6). Included in the gill-specific genes is a new calpain, which was
421 reported to be expressed in the trout gill by Salem and coworkers (Salem et al., 2006).
422 Additionally, we identified several ion transporters such as Na/K-ATPase α 1a, which
423 was previously shown to be overexpressed in gill tissue by Richards et al. (Richards et
424 al., 2003). Proteins that are known to be regulators of cellular ion transport, such as
425 the FXYP proteins that are known to regulate the substrate affinity of the Na/K-
426 ATPase enzyme (Garty and Karlish, 2006), were also overexpressed in gill tissue. In
427 the Atlantic salmon (*Salmo salar*), the expression of FXYP11 isoform was reported
428 to be 1,000-fold higher in gill than in other tissues (Tipsmark, 2008). In agreement
429 with this finding, our data revealed that this FXYP11 isoform is overexpressed in gill
430 tissue. Other interesting candidate genes that were found are known to be linked to
431 ammonia excretion. Ammonia is predominantly excreted in the gills in teleost fish.
432 Consequently, it is not surprising to find a higher expression of genes encoding
433 ammonia transporters (Rhesus (Rh) proteins) in gill tissue than in other organs. For
434 instance, we observed the predominant expression of Rh type C glycoprotein (*rhcg2*)
435 in the trout gill, in agreement with data reported by (Nawata et al., 2007). DDD
436 analysis also allowed us to identify new putative genes that are predominantly
437 expressed in gill tissue in comparison to other tissues. These include the zymogen
438 granule protein 16 (*zg16*) (see Fig. 1), which is considered to be a secretory lectin that
439 is involved in the regulation of granule secretion (Zhou et al., 2007), and the

440 envoplakin (*evpl*) and corniferin genes, which are known in mammals to be involved
441 in maintaining the structure of the epithelium. Envoplakin (see Fig. 1) is a component
442 of desmosomes and cornified envelope (CE) and is expressed in keratinizing and non-
443 keratinizing squamous epithelia (Ruhrberg et al., 1996). Cornifelin is also a
444 component of the CE (Michibata et al., 2004). The cornified envelope provides a
445 protective barrier between the environment and the living layers of the skin and is
446 thought to play an important role in the maintenance of the integrity of the epidermis
447 in mammals. The presence of several CE proteins in gill tissue could suggest their
448 importance in maintaining the barrier that protects the gill epithelium from the water
449 environment.

450 **4. Conclusions**

451 The present work utilized 454-Titanium technology to uncover tissue-specific cDNA
452 repertoires from various rainbow trout tissues. This important information has already
453 allowed us to develop an *in silico* data mining program that searches for genes that are
454 differentially expressed between tissues; however, these analyses are still difficult
455 because of the inherent complexity of the *de novo* assembly of these large transcript
456 sequencing datasets. This complexity has been already described (Costa et al., 2010)
457 and involves, for instance, the presence of different splice forms of genes, as well as
458 of allele-specific variants, which can split different contigs after the assembly process.
459 The current lack of a whole genome reference sequence in rainbow trout is a major
460 problem; however, ongoing projects should soon provide this genome resource,
461 allowing researchers to directly map these short read sequences onto the genome,
462 partially solving the difficulties inherent in *de novo* transcriptome assembly.
463

464 **References**

- 465 Anthony, T.E. and Heintz, N. Genetic lineage tracing defines distinct neurogenic and
466 gliogenic stages of ventral telencephalic radial glial development. *Neural*
467 *development* **3** (2008), p. 30.
- 468 Arai, A., Spencer, J.A. and Olson, E.N. STARS, a striated muscle activator of Rho
469 signaling and serum response factor-dependent transcription. *The Journal of*
470 *biological chemistry* **277** (2002), pp. 24453-9.
- 471 Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cox, T.,
472 Cunningham, F., Curwen, V., Cutts, T., Down, T., Durbin, R., Fernandez-
473 Suarez, X.M., Flicek, P., Graf, S., Hammond, M., Herrero, J., Howe, K., Iyer,
474 V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Kokocinski, F., Kulesha,
475 E., London, D., Longden, I., Melsopp, C., Meidl, P., Overduin, B., Parker, A.,
476 Proctor, G., Prlic, A., Rae, M., Rios, D., Redmond, S., Schuster, M., Sealy, I.,
477 Searle, S., Severin, J., Slater, G., Smedley, D., Smith, J., Stabenau, A., Stalker,
478 J., Trevanion, S., Ureta-Vidal, A., Vogel, J., White, S., Woodward, C. and
479 Hubbard, T.J. Ensembl 2006. *Nucleic Acids Research* **34** (2006), pp. D556-61.
- 480 Bobe, J. and Goetz, F.W. Cysteine protease inhibitor is specifically expressed in pre-
481 and early-vitellogenic oocytes from the brook trout periovulatory ovary. *Mol*
482 *Reprod Dev* **60** (2001), pp. 312-8.
- 483 Bobe, J., Nguyen, T., Mahe, S. and Monget, P. In silico identification and molecular
484 characterization of genes predominantly expressed in the fish oocyte. *BMC*
485 *Genomics* **9** (2008), p. 499.
- 486 Busson, M., Carazo, A., Seyer, P., Grandemange, S., Casas, F., Pessemesse, L.,
487 Rouault, J.P., Wrutniak-Cabello, C. and Cabello, G. Coactivation of nuclear

488 receptors and myogenic factors induces the major BTG1 influence on muscle
489 differentiation. *Oncogene* **24** (2005), pp. 1698-710.

490 Collier, L.S. and Largaespada, D.A. Transposable elements and the dynamic somatic
491 genome. *Genome biology* **8 Suppl 1** (2007), p. S5.

492 Costa, V., Angelini, C., De, F.I. and Ciccodicola, A. Uncovering the complexity of
493 transcriptomes with RNA-Seq. *J Biomed Biotechnol* **2010** (2010), p. 853916.

494 Crawford, J.E., Guelbeogo, W.M., Sanou, A., Traore, A., Vernick, K.D., Sagnon, N.
495 and Lazzaro, B.P. De novo transcriptome sequencing in *Anopheles funestus*
496 using Illumina RNA-seq technology. *PLoS One* **5** (2010), p. e14202.

497 Desvignes, T., Pontarotti, P., Fauvel, C. and Bobe, J. Nme protein family evolutionary
498 history, a vertebrate perspective. *BMC Evol Biol* **9** (2009), p. 256.

499 Diotel, N., Le Page, Y., Mouriec, K., Tong, S.K., Pellegrini, E., Vaillant, C., Anglade,
500 I., Brion, F., Pakdel, F., Chung, B.C. and Kah, O. Aromatase in the brain of
501 teleost fish: expression, regulation and putative functions. *Frontiers in*
502 *neuroendocrinology* **31** (2010), pp. 172-92.

503 Fraser, B.A., Weadick, C.J., Janowitz, I., Rodd, H. and Hughes, K.A. Sequencing and
504 characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC*
505 *Genomics* **12** (2011), p. 202.

506 Gabillard, J.C., Duval, H., Cauty, C., Rescan, P.Y., Weil, C. and Le Bail, P.Y.
507 Differential expression of the two GH genes during embryonic development of
508 rainbow trout *Oncorhynchus mykiss* in relation with the IGFs system.
509 *Molecular Reproduction and Development* **64** (2003), pp. 32-40.

510 Garty, H. and Karlish, S.J. Role of FXYD proteins in ion transport. *Annual review of*
511 *physiology* **68** (2006), pp. 431-59.

512 Govoroun, M., Le Gac, F. and Guiguen, Y. Generation of a large scale repertoire of
513 Expressed Sequence Tags (ESTs) from normalised rainbow trout cDNA
514 libraries. *BMC.Genomics* **7** (2006), p. 196.

515 Greenberg, C.C., Connelly, P.S., Daniels, M.P. and Horowitz, R. Krp1 (Sarcosin)
516 promotes lateral fusion of myofibril assembly intermediates in cultured mouse
517 cardiomyocytes. *Experimental Cell Research* **314** (2008), pp. 1177-91.

518 Haider, S., Ballester, B., Smedley, D., Zhang, J., Rice, P. and Kasprzyk, A. BioMart
519 Central Portal--unified access to biological data. *Nucleic Acids Research* **37**
520 (2009), pp. W23-7.

521 Harwood, B.N., Cross, S.K., Radford, E.E., Haac, B.E. and De Vries, W.N. Members
522 of the WNT signaling pathways are widely expressed in mouse ovaries,
523 oocytes, and cleavage stage embryos. *Developmental dynamics : an official*
524 *publication of the American Association of Anatomists* **237** (2008), pp. 1099-
525 111.

526 Hieber, V., Dai, X., Foreman, M. and Goldman, D. Induction of alpha1-tubulin gene
527 expression during development and regeneration of the fish central nervous
528 system. *Journal of neurobiology* **37** (1998), pp. 429-40.

529 Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff,
530 J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M.,
531 Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C.,
532 Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle,
533 S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E.,
534 Ureta-Vidal, A., Vastrik, I. and Clamp, M. The Ensembl genome database
535 project. *Nucleic Acids Research* **30** (2002), pp. 38-41.

536 Kanamori, A., Naruse, K., Mitani, H., Shima, A. and Hori, H. Genomic organization
537 of ZP domain containing egg envelope genes in medaka (*Oryzias latipes*).
538 *Gene* **305** (2003), pp. 35-45.

539 Kile, B.M., Guillot, T.S., Venton, B.J., Wetsel, W.C., Augustine, G.J. and Wightman,
540 R.M. Synapsins differentially control dopamine and serotonin release. *The*
541 *Journal of neuroscience : the official journal of the Society for Neuroscience*
542 **30** (2010), pp. 9762-70.

543 Kleene, K.C., Mulligan, E., Steiger, D., Donohue, K. and Mastrangelo, M.A. The
544 mouse gene encoding the testis-specific isoform of Poly(A) binding protein
545 (Pabp2) is an expressed retroposon: intimations that gene expression in
546 spermatogenic cells facilitates the creation of new genes. *Journal of Molecular*
547 *Evolution* **47** (1998), pp. 275-81.

548 Koop, B.F., von Schalburg, K.R., Leong, J., Walker, N., Lieph, R., Cooper, G.A.,
549 Robb, A., Beetz-Sargent, M., Holt, R.A., Moore, R., Brahmabhatt, S., Rosner,
550 J., Rexroad, C.E., 3rd, McGowan, C.R. and Davidson, W.S. A salmonid EST
551 genomic study: genes, duplications, phylogeny and microarrays. *BMC*
552 *Genomics* **9** (2008), p. 545.

553 Ku, M. and Melton, D.A. Xwnt-11: a maternally expressed *Xenopus* wnt gene.
554 *Development* **119** (1993), pp. 1161-1173.

555 Leroux, S., Feve, K., Vignoles, F., Bouchez, O., Klopp, C., Noiro, C., Gourichon, D.,
556 Richard, S., Leterrier, C., Beaumont, C., Minvielle, F., Vignal, A. and Pitel, F.
557 Non PCR-amplified Transcripts and AFLP fragments as reduced
558 representations of the quail genome for 454 Titanium sequencing. *BMC*
559 *research notes* **3** (2010), p. 214.

- 560 Marlow, F.L. and Mullins, M.C. Bucky ball functions in Balbiani body assembly and
561 animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish.
562 *Developmental Biology* **321** (2008), pp. 40-50.
- 563 Marz, M., Chapouton, P., Diotel, N., Vaillant, C., Hesl, B., Takamiya, M., Lam, C.S.,
564 Kah, O., Bally-Cuif, L. and Strahle, U. Heterogeneity in progenitor cell
565 subtypes in the ventricular zone of the zebrafish adult telencephalon. *Glia* **58**
566 (2010), pp. 870-88.
- 567 Metzker, M.L. Sequencing technologies - the next generation. *Nature reviews.*
568 *Genetics* **11** (2010), pp. 31-46.
- 569 Michibata, H., Chiba, H., Wakimoto, K., Seishima, M., Kawasaki, S., Okubo, K.,
570 Mitsui, H., Torii, H. and Imai, Y. Identification and characterization of a novel
571 component of the cornified envelope, cornifelin. *Biochemical and Biophysical*
572 *Research Communications* **318** (2004), pp. 803-13.
- 573 Mizrachi, E., Hefer, C.A., Ranik, M., Joubert, F. and Myburg, A.A. De novo
574 assembled expressed gene catalog of a fast-growing Eucalyptus tree produced
575 by Illumina mRNA-Seq. *BMC Genomics* **11** (2010), p. 681.
- 576 Muotri, A.R., Marchetto, M.C., Coufal, N.G. and Gage, F.H. The necessary junk: new
577 functions for transposable elements. *Human Molecular Genetics* **16 Spec No.**
578 **2** (2007), pp. R159-67.
- 579 Murray, D., Doran, P., MacMathuna, P. and Moss, A.C. In silico gene expression
580 analysis--an overview. *Molecular cancer* **6** (2007), p. 50.
- 581 Naeve, G.S., Ramakrishnan, M., Kramer, R., Hevroni, D., Citri, Y. and Theill, L.E.
582 Neuritin: a gene induced by neural activity and neurotrophins that promotes
583 neuritogenesis. *Proceedings of the National Academy of Sciences of the*
584 *United States of America* **94** (1997), pp. 2648-53.

585 Nawata, C.M., Hung, C.C., Tsui, T.K., Wilson, J.M., Wright, P.A. and Wood, C.M.
586 Ammonia excretion in rainbow trout (*Oncorhynchus mykiss*): evidence for Rh
587 glycoprotein and H⁺-ATPase involvement. *Physiological genomics* **31** (2007),
588 pp. 463-74.

589 Perteau, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee,
590 Y., White, J., Cheung, F., Parvizi, B., Tsai, J. and Quackenbush, J. TIGR Gene
591 Indices clustering tools (TGICL): a software system for fast clustering of large
592 EST datasets. *Bioinformatics* **19** (2003), pp. 651-2.

593 Plageman, T.F., Jr. and Yutzey, K.E. T-box genes and heart development: putting the
594 "T" in heart. *Developmental dynamics : an official publication of the*
595 *American Association of Anatomists* **232** (2005), pp. 11-20.

596 Rexroad, C.E., Lee, Y., Keele, J.W., Karamycheva, S., Brown, G., Koop, B., Gahr,
597 S.A., Palti, Y. and Quackenbush, J. Sequence analysis of a rainbow trout
598 cDNA library and creation of a gene index. *Cytogenetic and Genome*
599 *Research* **102** (2003), pp. 347-354.

600 Richards, J.G., Semple, J.W., Bystriansky, J.S. and Schulte, P.M. Na⁺/K⁺-ATPase
601 alpha-isoform switching in gills of rainbow trout (*Oncorhynchus mykiss*)
602 during salinity transfer. *The Journal of experimental biology* **206** (2003), pp.
603 4475-86.

604 Rolland, A.D., Lareyre, J.J., Goupil, A.S., Montfort, J., Ricordel, M.J., Esquerre, D.,
605 Hugot, K., Houlgatte, R., Chalmel, F. and Le Gac, F. Expression profiling of
606 rainbow trout testis development identifies evolutionary conserved genes
607 involved in spermatogenesis. *BMC Genomics* **10** (2009), p. 546.

608 Rother, S., Schmidt, R., Brysch, W. and Schlingensiepen, K.H. Learning-induced
609 expression of meningeal ependymin mRNA and demonstration of ependymin
610 in neurons and glial cells. *Journal of neurochemistry* **65** (1995), pp. 1456-64.

611 Ruhrberg, C., Hajibagheri, M.A., Simon, M., Dooley, T.P. and Watt, F.M.
612 Envoplakin, a novel precursor of the cornified envelope that has homology to
613 desmoplakin. *The Journal of cell biology* **134** (1996), pp. 715-29.

614 Salem, M., Rexroad, C.E., 3rd, Wang, J., Thorgaard, G.H. and Yao, J.
615 Characterization of the rainbow trout transcriptome using Sanger and 454-
616 pyrosequencing approaches. *BMC Genomics* **11** (2010), p. 564.

617 Salem, M., Rexroad, C.E. and Yao, J. Identification of a novel gill-specific calpain
618 from rainbow trout (*Oncorhynchus mykiss*). *Fish Physiology and*
619 *Biochemistry* **32** (2006), pp. 1-6.

620 Sanchez-Sanchez, A.V., Camp, E., Garcia-Espana, A., Leal-Tassias, A. and Mullor,
621 J.L. Medaka Oct4 is expressed during early embryo development, and in
622 primordial germ cells and adult gonads. *Developmental dynamics : an official*
623 *publication of the American Association of Anatomists* **239** (2010), pp. 672-9.

624 Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. and
625 Kasprzyk, A. BioMart--biological queries made easy. *BMC Genomics* **10**
626 (2009), p. 22.

627 Stabenau, A., McVicker, G., Melsopp, C., Proctor, G., Clamp, M. and Birney, E. The
628 Ensembl core software libraries. *Genome Research* **14** (2004), pp. 929-33.

629 Tan, X., Rotllant, J., Li, H., De Deyne, P. and Du, S.J. SmyD1, a histone
630 methyltransferase, is required for myofibril organization and muscle
631 contraction in zebrafish embryos. *Proceedings of the National Academy of*
632 *Sciences of the United States of America* **103** (2006), pp. 2713-8.

633 Tao, Q., Yokota, C., Puck, H., Kofron, M., Birsoy, B., Yan, D., Asashima, M., Wylie,
634 C.C., Lin, X. and Heasman, J. Maternal wnt11 activates the canonical wnt
635 signaling pathway required for axis formation in *Xenopus* embryos. *Cell* **120**
636 (2005), pp. 857-71.

637 Thorgaard, G.H., Bailey, G.S., Williams, D., Buhler, D.R., Kaattari, S.L., Ristow,
638 S.S., Hansen, J.D., Winton, J.R., Bartholomew, J.L., Nagler, J.J., Walsh, P.J.,
639 Vijayan, M.M., Devlin, R.H., Hardy, R.W., Overturf, K.E., Young, W.P.,
640 Robison, B.D., Rexroad, C. and Palti, Y. Status and opportunities for
641 genomics research with rainbow trout. *Comp Biochem.Physiol B*
642 *Biochem.Mol.Biol.* **133** (2002), pp. 609-646.

643 Tipsmark, C.K. Identification of FXVD protein genes in a teleost: tissue-specific
644 expression and response to salinity change. *American journal of physiology.*
645 *Regulatory, integrative and comparative physiology* **294** (2008), pp. R1367-
646 78.

647 Toro, S., Wegner, J., Muller, M., Westerfield, M. and Varga, Z.M. Identification of
648 differentially expressed genes in the zebrafish hypothalamic-pituitary axis.
649 *Gene expression patterns : GEP* **9** (2009), pp. 200-8.

650 Udvadia, A.J., Koster, R.W. and Skene, J.H. GAP-43 promoter elements in transgenic
651 zebrafish reveal a difference in signals for axon growth during CNS
652 development and regeneration. *Development* **128** (2001), pp. 1175-82.

653 Vallette-Kasic, S., Couture, C., Balsalobre, A., Gauthier, Y., Metherell, L., Dattani,
654 M. and Drouin, J. The TPIT gene mutation M86R associated with isolated
655 adrenocorticotropin deficiency interferes with protein: protein interactions.
656 *The Journal of clinical endocrinology and metabolism* **92** (2007), pp. 3991-9.

657 Vargas, A., Moreau, J., Landry, S., LeBellego, F., Toufaily, C., Rassart, E., Lafond, J.
658 and Barbeau, B. Syncytin-2 plays an important role in the fusion of human
659 trophoblast cells. *Journal of Molecular Biology* **392** (2009), pp. 301-18.

660 Volff, J.N. Turning junk into gold: domestication of transposable elements and the
661 creation of new genes in eukaryotes. *BioEssays : news and reviews in*
662 *molecular, cellular and developmental biology* **28** (2006), pp. 913-22.

663 Wilhelm, B.T. and Landry, J.R. RNA-Seq-quantitative measurement of expression
664 through massively parallel RNA-sequencing. *Methods* **48** (2009), pp. 249-57.

665 Zhang, L., Kelley, J., Schmeisser, G., Kobayashi, Y.M. and Jones, L.R. Complex
666 formation between junctin, triadin, calsequestrin, and the ryanodine receptor.
667 Proteins of the cardiac junctional sarcoplasmic reticulum membrane. *The*
668 *Journal of biological chemistry* **272** (1997), pp. 23389-97.

669 Zhou, Y.B., Cao, J.B., Yang, H.M., Zhu, H., Xu, Z.G., Wang, K.S., Zhang, X., Wang,
670 Z.Q. and Han, Z.G. hZG16, a novel human secreted protein expressed in liver,
671 was down-regulated in hepatocellular carcinoma. *Biochemical and*
672 *Biophysical Research Communications* **355** (2007), pp. 679-86.

673

674

675 **Acknowledgments**

676 This work was supported by funds from the "Institut National de la Recherche
677 Agronomique" (AIP bioressources INRA 2009) and by an EC's Seventh Framework
678 Program (222719-LIFECYCLE). B.N. was supported by a joined Région Bretagne
679 and INRA PhD fellowship.

680

681 **Figures**

682 **Figure 1 - The identification of rainbow trout genes that are predominantly** 683 **expressed in brain, testis, ovary, gill or muscle tissues.**

684 Real-time PCR analysis was conducted using total RNA from the following tissues:
685 brain, pituitary, gill, heart, stomach, liver, intestine, muscle, skin, spleen, head kidney
686 (HK), trunk kidney (TK), oocyte, post-ovulatory ovary, and stage II testis. For each
687 tissue, three separate RT reactions were carried out using separate RNA samples that
688 originated from three different fish. The relative expression of each gene was
689 normalized to that of the 18S gene and was calculated as the percentage of the highest
690 expression level recorded for each gene.

691 **Figure 2 – The expression of β -enolase in muscle**

692 Transverse sections through adult muscle showing a strong signal for the β -enolase
693 transcript in rainbow trout white fast muscle (WM). No signal is detected in red slow
694 muscle (RM) or in the adipose tissue separating these two muscle regions.

695

696 **Tables**

697 **Table 1 - Summary of the sequencing run results.**

Tissues	Number of reads	Number of base pairs	Average sequence length	Number of 454 runs
Gill	613.223	232.066.637	378 bp	1
Brain	481.363	159.232.104	331 bp	1.5
Ovary	738.510	234.665.107	318 bp	1
Testis	1.234.260	381.485.303	309 bp	1.5
Total	3.067.356	1.007.449.151	328 bp	5 runs of 454

698

699 **Table 2 – Selected genes that were over-expressed in the ovarian libraries.**

700 The expression scores (representing the number of reads in the corresponding contig)
 701 for the genes that are discussed within the text are shown for the ovarian and non-
 702 ovarian libraries. Genes with expression profiles that were confirmed by qPCR are
 703 identified in the table with an asterisk (*) following the contig name.

Contig name	Best Blast Hit	Ovarian	Non-ovarian
BX082756.p.om.8	ZPC domain containing protein 5	1042	1
CU072510.p.om.8	ZPC domain containing protein 1	132	0
GAY7CUQ01AN5KC.p.om.8	ZPA domain containing protein	68	0
CB486811.p.om.8*	Protein Wnt-11	80	4
BX079798.p.om.8*	bucky ball	87	0
BX074171.p.om.8	oocyte protease inhibitor-2	248	2
CR373079.p.om.8	oocyte protease inhibitor-2 (opi-2)	170	2
BX303838.p.om.8	serpin peptidase inhibitor	169	0
CB490333.p.om.8*	POU domain, class 5, transcription factor 1	40	0
BX088048.p.om.8	zygote arrest 1 (zar1)	87	3
CX041779.p.om.8	growth/differentiation factor 9	32	1

704

705 **Table 3 – Selected genes that were over-expressed in the testicular libraries.**

706 The expression scores (representing the number of reads in the corresponding contig)
 707 for the genes that are discussed within the text are shown for the testicular and non-
 708 testicular libraries. Genes with expression profiles that were confirmed by qPCR are
 709 identified in the table with an asterisk (*) following the contig name.

Contig name	Best Blast Hit	Testicular	Non-testicular
-------------	----------------	------------	----------------

DQ489287.p.om.8	GSDF	2445	20
BX857318.p.om.8	SYCP1	198	2
GBPNQUD01A6W4I.p.om.8*	Myelin associated glycoprotein	372	1
CU063452.p.om.8	AMH	176	1
BX080625.p.om.8	Cdk6	124	1
BX080735.p.om.8	Cohesin subunit	119	2
F0DLFKL01BENKG.p.om.8	Unknown	105	1
FYV3OTN02HHPOI.p.om.8	Rec8	106	2
BX314487.p.om.8	L-asparaginase	85	1
FYV3OTN01B8PA8.p.om.8	ERV-FRD provirus ancestral Env polyprotein	68	2
F6VG15V02FTMKM.p.om.8	Histone-lysine N-methyltransferase	61	1
FYV3OTN01A1QIV.p.om.8	Zinc finger CW-type PWWP domain protein 1	53	1
CB487402.p.om.8	LINE-1 reverse transcriptase	62	5
FYV3OTN01E3SYW.p.om.8	Transposable element Tc1 transposase	47	1
F6VG15V02HO88H.p.om.8	Pro-Pol polyprotein	88	2
CX036404.p.om.8	DAZL	24	1
GBPNQUD01BLTP4.p.om.8*	TBX19	23	0
GBPNQUD01AIF47.p.om.8	Retrotransposon protein 1	138	0
BX076816.p.om.8	Envelope protein	51	0

710

711 **Table 4 – Selected genes that were over-expressed in the muscle libraries.**

712 The expression scores (representing the number of reads in the corresponding contig)

713 for the genes discussed within the text are shown for the muscle and non-muscle

714 libraries. Genes with expression profiles that were confirmed by qPCR are identified

715 in the table with an asterisk (*) following the contig name.

Contig name	Best Blast Hit	Muscle	Non-muscle
CA343693.p.om.8*	Phosphoglycerate mutase 2-1 (muscle)	5035	2
CR372691.p.om.8	myosin, light polypeptide 3-3	2141	1
CA365356.p.om.8	fast skeletal troponin C	21213	10
CA345074.p.om.8*	Kelch repeat and BTB domain-containing protein 10	224	2
CA361268.p.om.8	SET and MYND domain-containing protein 1	103	2
CX252008.p.om.8	Calsequestrin-1	83	2

716

717 **Table 5 – Selected genes that were over-expressed in the brain libraries.**

718 The expression scores (representing the number of reads in the corresponding contig)
719 for the genes that are discussed within the text are shown for the brain and non-brain
720 libraries. Genes with expression profiles that were confirmed by qPCR are identified
721 in the table with an asterisk (*) following the contig name.

Contig name	Best Blast Hit	Brain	Non-brain
BX082627.p.om.8*	BLBP	1125	95
BX316810.p.om.8	glial high affinity glutamate transporter	261	1
CA368669.p.om.8	Neuritin	12	6
CA381682.p.om.8*	Neuromodulin	176	2
CX153668.p.om.8	Brain-specific polypeptide PEP-19	51	13
CA343423.p.om.8	Synapsin	47	7

722

723 **Table 6 – Selected genes that were over-expressed in the gill libraries.**

724 The expression scores (representing the number of reads in the corresponding contig)
725 for the genes that are discussed within the text are shown for the gill and non-gill
726 libraries. Genes with expression profiles that were confirmed by qPCR are identified
727 in the table with an asterisk (*) following the contig name.

Contig name	Best Blast Hit	Gill	Non-gill
CA345680.p.om.8	gill-specific calpain	43	1
CA362824.p.om.8	Rh type C glycoprotein (rhcg2)	132	0
CA345079.p.om.8	Na/K ATPase alpha subunit isoform 1a	319	21
CA349792.p.om.8	FXVD domain containing ion transport regulator 11	294	0
CA365840.p.om.8*	Envoplakin	58	1
BX321817.p.om.8	cornifelin-like protein b	57	1
CA345392.p.om.8*	Zymogen granule membrane protein 16	2075	10

729 Additional files

730 **Additional File 1 – Genes that were detected by Digital Differential Display**
731 **analysis to be predominantly expressed in rainbow trout ovary, testis, brain,**
732 **muscle and gill tissues.**

733 This file contains the number of reads per condition for the contigs that were
734 identified to be predominantly expressed in rainbow trout ovary, testis, brain, muscle
735 and gill tissues (one folder page per tissue of interest) using the digital differential
736 display analysis. For each contig containing 10 or more reads in the tissue of interest
737 (FP in contig) and no hits in the other tissues (SP in contig), or for each contig having
738 a fold change of more than 10 (FP/SP), this file provides information on the
739 significance of the enrichment using Fisher's Exact Test (no value is given for contigs
740 with no hit in the SP condition). The file also provides information regarding the top
741 blast hit results for all of the contig sequences (best hit description, best hit database,

742 the percentage of sequence coverage with the best hit, and the percentage of sequence
743 conservation with the best hit). For each contig, a dynamic internet link with the
744 corresponding contig view webpage is available (click directly on the contig name;
745 Excel file, .xls).

746 **Additional File 2 – Primer sequences used for qPCR validation of some**
747 **differentially expressed genes.**

748