



**HAL**  
open science

## In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in XX rainbow

Audrey Dehaullon, Clémence Fraslin, Anastasia Bestin, Charles Poncet, Yann Guiguen, Edwige Quillet, Florence Phocas

### ► To cite this version:

Audrey Dehaullon, Clémence Fraslin, Anastasia Bestin, Charles Poncet, Yann Guiguen, et al.. In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in XX rainbow. 2024. hal-04796738

**HAL Id: hal-04796738**

**<https://hal.science/hal-04796738v1>**

Preprint submitted on 21 Nov 2024

**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.

# PLOS ONE

## In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in XX rainbow trout --Manuscript Draft--

<b>Manuscript Number:</b>	PONE-D-24-47853
<b>Article Type:</b>	Research Article
<b>Full Title:</b>	In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in XX rainbow trout
<b>Short Title:</b>	QTL for spontaneous sex-reversal in XX rainbow trout
<b>Corresponding Author:</b>	Florence Phocas INRAE Centre Île-de-France Jouy-en-Josas Antony: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Ile-de-France Jouy-en-Josas Antony Jouy-en-Josas, FRANCE
<b>Keywords:</b>	fish, sex determination, masculinization, intersex, neomale, genetics
<b>Abstract:</b>	Sex determination is a flexible process in fish, controlled by genetics or environmental factors or a combination of both depending on the species. Revealing the underlying molecular mechanisms may have important implications for research on reproductive development in vertebrates, as well as sex-ratio control and selective breeding in fish. Phenotypic sex in rainbow trout is primarily controlled by a XX/XY male heterogametic sex determination system. Unexpectedly in genetically XX all-female farmed populations, a small proportion of males or intersex individuals are regularly observed. This spontaneous masculinisation is a highly heritable trait, controlled by minor sex-modifier genes that remain unknown, although several QTL regions were detected in previous studies. In this work we used genome-based approaches and various statistical methods to investigate these QTL regions. We validated in six different French farmed populations DNA markers we had previously identified in a different commercial population on chromosomes Omy1, Omy12 and Omy20. We also identified functional candidate genes located that may be involved in spontaneous masculinisation by reducing germ cell proliferation and repressing oogenesis of XX-rainbow trout in the absence of the master sex determining gene. In particular, syndig1, tlx1 and hells on Omy1, as well as khdrbs2 and csmd1 on Omy20 deserve further investigation as potential sex-modifier genes to precise their functional roles as well as their interaction with rearing temperature. Those findings could be used to produce all-female populations that are preferred by farmers due to a delayed maturation of females and higher susceptibility of male trout to diseases.
<b>Order of Authors:</b>	Audrey Dehaullon Clémence Fraslin Anastasia Bestin Charles Poncet Yann Guiguen Edwige Quillet Florence Phocas
<b>Opposed Reviewers:</b>	
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
<b>Financial Disclosure</b>  Enter a financial disclosure statement that describes the sources of funding for the	Yes

work included in this submission. Review the [submission guidelines](#) for detailed requirements. View published research articles from [PLOS ONE](#) for specific examples.

This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.

#### Funded studies

Enter a statement with the following details:

- Initials of the authors who received each award
- Grant numbers awarded to each author
- The full name of each funder
- URL of each funder website
- Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?

Did you receive funding for this work?

Please add funding details as follow-up to "**Financial Disclosure**

Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the [submission guidelines](#) for detailed requirements. View published research articles from [PLOS ONE](#) for specific examples.

This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate.

The European Maritime and Fisheries Fund and the French National Government supported the NeoBio project (R FEA470016FA1000008).

<p><b>Funded studies</b></p> <p>Enter a statement with the following details:</p> <ul style="list-style-type: none"> <li>• Initials of the authors who received each award</li> <li>• Grant numbers awarded to each author</li> <li>• The full name of each funder</li> <li>• URL of each funder website</li> <li>• Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?</li> </ul> <p>Did you receive funding for this work?"</p>	
<p>Please select the country of your main research funder (please select carefully as in some cases this is used in fee calculation).</p> <p>as follow-up to "<b>Financial Disclosure</b></p> <p>Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the <a href="#">submission guidelines</a> for detailed requirements. View published research articles from <a href="#">PLOS ONE</a> for specific examples.</p> <p>This statement is required for submission and <b>will appear in the published article</b> if the submission is accepted. Please make sure it is accurate.</p> <p><b>Funded studies</b></p> <p>Enter a statement with the following details:</p> <ul style="list-style-type: none"> <li>• Initials of the authors who received each award</li> <li>• Grant numbers awarded to each author</li> <li>• The full name of each funder</li> <li>• URL of each funder website</li> <li>• Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?</li> </ul> <p>Did you receive funding for this work?"</p>	<p>FRANCE - FR</p>
<p><b>Competing Interests</b></p>	<p>The authors have declared that no competing interests exist.</p>

Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any [competing interests](#) that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests.

This statement is **required** for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate and that any funding sources listed in your Funding Information later in the submission form are also declared in your Financial Disclosure statement.

View published research articles from [PLOS ONE](#) for specific examples.

**NO authors have competing interests**

Enter: *The authors have declared that no competing interests exist.*

**Authors with competing interests**

Enter competing interest details beginning with this statement:

*I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]*

\* typeset

**Ethics Statement**

Enter an ethics statement for this submission. This statement is required if the study involved:

- Human participants
- Human specimens or tissue
- Vertebrate animals or cephalopods
- Vertebrate embryos or tissues

As part of the standard breeding practices for a commercial breeding program, the handling of fish was not subject to oversight by an institutional ethic committee. Animals were treated in compliance with the European Communities Council Directive (9858/CEC) for the care and use of farm animals.

- Field research

Write "N/A" if the submission does not require an ethics statement.

General guidance is provided below.

Consult the [submission guidelines](#) for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**

**Format for specific study types**

**Human Subject Research (involving human participants and/or tissue)**

- Give the name of the institutional review board or ethics committee that approved the study
- Include the approval number and/or a statement indicating approval of this research
- Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously)

**Animal Research (involving vertebrate animals, embryos or tissues)**

- Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval
- Include an approval number if one was obtained
- If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering
- If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied

**Field Research**

Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting:

- Field permit number
- Name of the institution or relevant body that granted permission

**Data Availability**

Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the [PLOS Data Policy](#) and [FAQ](#) for detailed information.

No - some restrictions will apply

A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and **will be published in the article**, if accepted.

**Important:** Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box.

Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?

**Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details.**

- If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: *All XXX files are available from the XXX database (accession number(s) XXX, XXX).*
- If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following: *All relevant data are within the manuscript and its Supporting Information files.*
- If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example:

*Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.*

*The data underlying the results presented in the study are available from (include the name of the third party*

Raw sequence data that were used in this study are deposited in the ENA (Project PRJEB75960 in <https://www.ebi.ac.uk/ena/browser/view/PRJEB75960>). They will be fully available on the 1st of January 2025. Other relevant data are within Supporting Information files.

<p><i>and contact information or URL).</i></p> <ul style="list-style-type: none"><li>• This text is appropriate if the data are owned by a third party and authors do not have permission to share the data.</li></ul> <p>* typeset</p>	
Additional data availability information:	Tick here if the URLs/accession numbers/DOIs will be available only after acceptance of the manuscript for publication so that we can ensure their inclusion before publication.

Dear Editors,

Please find here our manuscript entitled “ In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in XX rainbow trout” that we are submitting for your consideration for publication as a research article in Plos One.

In this manuscript, we investigated the genetic architecture of spontaneous masculinization in rainbow trout (*Oncorhynchus mykiss*). Although this salmonids species has a heterogametic (XY) sex determining system controlled by a master gene (*sdY*), phenotypically masculinised fish are regularly observed in all-female farmed stocks. Better understanding the factors underlying sex determination in fish is highly interesting both from a fundamental biological perspective and for aquaculture purposes.

We confirmed on six different rainbow trout populations some minor sex-determining regions previously identified as involved in spontaneous maleness on chromosomes Omy1, Omy12 and Omy20 (Fraslin et al., 2000 published in Scientific Reports). We suggest new potential sex-modifier genes based on positional and functional information. Those findings suggest that spontaneous female-to-male sex-reversal could be associated with genetic factors able to reduce germ cell proliferation and arrest oogenesis.

We have no potential competing interests to declare and the content of the manuscript has not been published, or submitted for publication elsewhere. All authors have approved the manuscript for submission.

Best regards,

Florence Phocas

1 **Running title:** QTL for spontaneous sex-reversal in XX rainbow trout

2

3 **In-depth investigation of genome to refine QTL positions for spontaneous sex-reversal in**  
4 **XX rainbow trout**

5 Audrey Dehaullon<sup>a †</sup>, Clémence Fraslin<sup>a,b, †</sup>, Anastasia Bestin<sup>c</sup>, Charles Poncet<sup>d</sup>, Yann  
6 Guiguen<sup>e</sup>, Edwige Quillet<sup>a</sup>, Florence Phocas<sup>a\*</sup>

7

8 <sup>a</sup> Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France

9 <sup>b</sup> The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of  
10 Edinburgh, Midlothian EH25 9RG, United Kingdom

11 <sup>c</sup> SYSAAF, Station INRAE-LPGP, Campus de Beaulieu, 35042 Rennes cedex, France

12 <sup>d</sup> Université Clermont-Auvergne, INRAE, GDEC, 63039 Clermont-Ferrand, France

13 <sup>e</sup> INRAE, LPGP, Campus de Beaulieu, 35042 Rennes cedex, France

14

15 † Co-first authors: Audrey Dehaullon and Clémence Fraslin

16 \*corresponding author's contact:

17 email: [florence.phocas@inrae.fr](mailto:florence.phocas@inrae.fr); orcid: 0000-0003-1161-3665

18

19 **Abstract**

20 Sex determination is a flexible process in fish, controlled by genetics or environmental factors  
21 or a combination of both depending on the species. Revealing the underlying molecular  
22 mechanisms may have important implications for research on reproductive development in  
23 vertebrates, as well as sex-ratio control and selective breeding in fish. Phenotypic sex in  
24 rainbow trout is primarily controlled by a XX/XY male heterogametic sex determination  
25 system. Unexpectedly in genetically XX all-female farmed populations, a small proportion of  
26 males or intersex individuals are regularly observed. This spontaneous masculinisation is a  
27 highly heritable trait, controlled by minor sex-modifier genes that remain unknown, although  
28 several QTL regions were detected in previous studies. In this work we used genome-based  
29 approaches and various statistical methods to investigate these QTL regions. We validated in

30 six different French farmed populations DNA markers we had previously identified in a  
31 different commercial population on chromosomes Omy1, Omy12 and Omy20. We also  
32 identified functional candidate genes located that may be involved in spontaneous  
33 masculinisation by reducing germ cell proliferation and repressing oogenesis of XX-rainbow  
34 trout in the absence of the master sex determining gene. In particular, *syndig1*, *tlx1* and *hells* on  
35 Omy1, as well as *khdrbs2* and *csmd1* on Omy20 deserve further investigation as potential sex-  
36 modifier genes to precise their functional roles as well as their interaction with rearing  
37 temperature. Those findings could be used to produce all-female populations that are preferred  
38 by farmers due to a delayed maturation of females and higher susceptibility of male trout to  
39 diseases.

40

41 **Keywords:** fish, sex determination, masculinization, intersex, neomale, genetics

42

43 **Statements and Declarations**

44 **Competing interests:** The authors declare no competing interests.

45

46

47 **Introduction**

48 Sex reversal has been defined as a mismatch between the phenotypic and the genetic sex of an  
49 individual [1]. In fish, sex reversal can be induced by hormonal treatments (estrogens to  
50 feminize and androgens to masculinize) during critical fry stages [2, 3]. Manipulating the sexual  
51 phenotype of fish has been used to prevent early maturity and reproduction, as well as increase  
52 production in various cultured species or create subsequent sterile generations for fisheries  
53 management [3]. Sex-reversed individuals can be used to produce populations that are  
54 genetically all-male or all-female (also known as ‘monosex’). Rainbow trout (*Oncorhynchus*  
55 *mykiss*) is a worldwide cultured salmonid species whose production is frequently based on all-  
56 female stocks. The production of large rainbow trout over 1 kg body weight has become a  
57 growing commercial interest with the development of a market for fresh and smoked fillets. A  
58 major bottleneck for this commercial market is the early maturation of males (1-2 year) relative  
59 to females (2-3 years) that causes stopped growth, increased sensitivity to pathogens  
60 (*Saprolegnia* spp.), and reduced flesh quality. Therefore, all-female rainbow trout stocks are

61 often preferred in aquaculture. These stocks are currently produced by crossing hormonal sex-  
62 reversed genetically XX-females, known as neomales (XX males), with XX-females. The  
63 growing concerns or potential risks to human and environmental health make the prevalent  
64 approach of producing neomales via the oral application of the hormone 17-alpha  
65 methyltestosterone less sustainable. The finding of a sex control alternative to hormonal  
66 treatment based on a safe, consumer and environmentally friendly method is a major challenge  
67 for the production of all-female populations [4].

68 Sex determination and sex differentiation mechanisms in fish are diverse and complex.  
69 Understanding sex-differentiation processes requires to unravel the origin and developmental  
70 pathways of cells and organs involved in the formation of the primordial gonad. During primary  
71 sex determination in vertebrates, the primordial bipotential gonad commits to either an ovary  
72 or a testis developmental fate. The structure of fish gonads is similar to that of other vertebrates,  
73 with germ cells and associated supporting somatic cells intermixed. Within both ovary and  
74 testis, a clear distinction between germ and somatic cells can be made, with the former having  
75 the potential to mitotically divide and enter meiosis, and the latter differentiating into associated  
76 structural and endocrine cell types. In teleosts, primordial germ cells (PGCs) are established  
77 and specialized at the early blastula stage and migrate to the genital ridge during embryonic  
78 development [5].

79 Most of the teleosts are gonochorists meaning that all individuals within a species develop either  
80 as males or females and remain the same throughout their lives. Various systems of sex  
81 determination have been observed in gonochoristic fishes [1], spanning from a genetic sex  
82 determination system (GSD) with a genetic monofactorial system, XX/XY or ZZ/ZW, to  
83 different multifactorial systems, or environmental sex determination (ESD). During sex  
84 differentiation, the testis and ovary pathways are mutually antagonistic and compete for control  
85 of gonad fate. Sex reversal occurs when the sexual trajectory of the gonad is changed to the  
86 opposing pathway, switching the sexual phenotype of the organism to the opposite sex. The  
87 change in trajectory often occurs because of a failure to maintain the initiated pathway or a  
88 failure to repress the opposite pathway [6]. Once the gonad has committed to a particular fate,  
89 it begins to produce sex-specific hormones that will drive the secondary process of sex  
90 differentiation in which the somatic tissues of the organism differentiate toward one sex or the  
91 other. The developmental pathway leading to steroid production in gonadal somatic cells  
92 requires complex regulation of multiple genes [7] involved in the differentiation of steroid-  
93 producing cells. Sex steroids have local and direct effects on germ-cell development, but also

94 act as endocrine hormones to influence other cell types and organs involved in sex  
95 differentiation.

96 Over the last two decades evidence has accumulated that both somatic and germ cells play  
97 critical roles in fish gonadal differentiation. However, the exact mechanisms and factors  
98 involved remain to be elucidated [8, 9]. As evidenced from studies in mammals [10], somatic  
99 cells may first differentiate in response to the activation of master sex-determining genes, and  
100 subsequent differentiation of PGCs into male or female gametes follow in the differentiating  
101 testis and ovary in response to signals derived from surrounding somatic cells. Alternatively, it  
102 is possible that, first, PGCs interpret internal genetic or external environmental cues and directly  
103 transform into spermatogonia or oogonia; then, the surrounding somatic cells may be induced  
104 by the germ cells to differentiate accordingly to provide an appropriate hormonal environment  
105 for further gonadal differentiation [11]. It is thought that control of estradiol synthesis could  
106 play a key role not only for ovarian, but also for testicular differentiation and sex change in fish  
107 [12]. This working hypothesis states that the gonadal aromatase gene, *cyp19a1a*, up-regulation  
108 would be needed not only for triggering, but also for maintaining ovarian differentiation; and  
109 that *cyp19a1a* down-regulation would be the only necessary step for inducing a testicular  
110 differentiation pathway. In Atlantic salmon as well as in loach and goldfish, it was shown that  
111 germ cells are not essential for gonadal sex differentiation [13]. Germ cell-free female gonads  
112 develop an ovarian somatic structure in these species, unlike in zebrafish or medaka where all  
113 germ cell-free fish develop somatic testes.

114 Evidence from both the zebrafish and medaka models [14, 15, 16, 17] suggests important  
115 feminizing roles for the germ cells in gonadal differentiation and thus corroborates this  
116 alternative hypothesis. In particular, when germ cells are ablated in medaka, XX fish show  
117 female-to-male sex reversal, while XY fish exhibit male-to-female sex reversal [16]. Any  
118 disruption of this germ cell/somatic cell cross-talk would then disrupt ovarian somatic cell  
119 differentiation, leading to an absence of estrogen synthesis and a subsequent masculinization.  
120 More subtle regulations not involving the complete loss of germ cells, but instead some  
121 differential germ cell proliferation rates, could be also suggested as important triggers of  
122 gonadal sex differentiation. In this respect, it is worth to note that the Japanese medaka sex  
123 determining gene, *dmrt1Y*, has been found to be an inhibitor of germ cell proliferation [18] and  
124 is only expressed during testicular differentiation [19].

125 In species with genetic sex determination (GSD), the master sex determining trigger is encoded  
126 by a gene on a sex chromosome, which activates a network of downstream regulators of sex

127 differentiation. Currently, there are more than 20 sex determining genes discovered, and as  
128 many as 13 are derived from the TGF- $\beta$  pathway [20]. This pathway thus plays an important  
129 role in sex determination/differentiation of fishes but potentially regulating the number of germ  
130 cells, and/or inhibit aromatase activity to determine and maintain sex (e.g. Chen et al. [21]).  
131 Rainbow trout is a gonochoric species with an XX/XY GSD system. Interestingly, *sdY* the  
132 master sex-determining gene of rainbow trout and other salmonid species, does not belong to  
133 the usual master sex gene families (the DM domain and Sox protein families), nor does it belong  
134 to TGF- $\beta$  signaling pathway. It is a truncated duplicated copy of the immune response gene *irf9*  
135 [22]. While very specific to salmonids, the gene *sdY* is acting directly on the classical sex  
136 differentiation regulation network in all vertebrates as its protein has been shown to interact  
137 with *foxl2* [23] that is a well-known conserved female differentiation factor [24]. *foxl2* was also  
138 shown to be involved in the polled intersex syndrome in goats, where a deletion of the gene  
139 leads to triggers early testis differentiation and XX female-to-male sex reversal [25; 26].

140 Although phenotypic sex in rainbow trout is primarily determined by a male heterogametic  
141 GSD system, spontaneous masculinization of XX fish is a repeatedly phenomenon observed in  
142 various rainbow trout populations, at generally limited frequencies (~1-2%) with some  
143 individuals being only partially affected (intersex) [27]. Spontaneous sex reversal is a highly  
144 heritable trait [27] and several QTL associated with this trait were detected [28, 27],  
145 highlighting the existence of several minor sex-determining genes that are independent of the  
146 major sex determinant carried by the sex chromosome. In particular, in a previous study [27],  
147 we identified in a French trout population two QTL located on chromosome 1 with a strong  
148 evidence while two suggestive QTL were identified on chromosomes 12 and 20 using the  
149 rainbow trout reference genome of Swanson line [29]. All SNP variants from these four QTL  
150 regions that covered tens of genes were kept for further investigations in the present study.

151 Firstly, we sought to validate the existence of the four QTL in different French rainbow trout  
152 populations, and secondly, we wanted to refine their location in the initial discovery population  
153 by combining machine learning and principal components approaches, and considering the new  
154 reference genome derived for Arlee line [30].

155

## 156 **Material and Methods**

### 157 **Sequenced Samples**

158 The first data set was initially produced by Fraslin et al. [27]. The 4 QTL regions that were  
159 discovered were further studied in depth in this work. The fish samples and records came from  
160 the French rainbow trout farm “Les Fils de Charles Murgat” (Beaufort, France; UE approval  
161 number FR 38 032 001). The 60 female sequenced genomes were initially aligned against the  
162 reference assembly genome Omyk\_1.0 (GenBank assembly accession GCA\_002163495.1) of  
163 the double haploid line Swanson from Washington State University (WSU). Following the  
164 same procedure as described in Fraslin et al. [27], we realigned these 60 sequences against the  
165 reference genome USDA\_OmykA\_1.1 of the WSU double haploid line Arlee (GenBank  
166 assembly accession GCA\_013265735.3). This realignment was performed because French  
167 rainbow trout populations were known to be phylogenetically closer to Arlee’s genome than to  
168 Swanson’s one [31]. A total of 29,229,949 variants were obtained using a series of commands  
169 using three different variant calling tools: GATK 4.2.2.0 [32], Freebayes 1.3.5 [33] and  
170 SAMtools Mpileup 1.11 [34]. Quality controls were done by using vcftools 1.15. First, indels  
171 and SNPs located on un-located contigs or mitochondrial chromosome were removed. Then, a  
172 filtering was performed on variant coverage according to the recommended settings by GATK  
173 [35]. Final numbers of 21,517,540 SNPs and 4,850,647 indels, for a total of 26,368,187  
174 variants, were kept for further analysis. While the percentage of properly paired sequence reads  
175 over all sequenced pairs was in average 87.3% for the alignment against Swanson’s genome  
176 assembly, these statistics went up to 94.1% for the alignment against Arlee’s genome assembly.

177 Among the 60 sequenced females, we considered the set of 23 dams with at least 10 progeny  
178 records for sex phenotype and extreme proportions of sex-reversed offspring, either at least  
179 25% or below 6% of neomales among their XX daughters (see S1 Supplementary Tables, S1  
180 Table). For each of the 3 chromosomes investigated in the present study, all SNPs that exhibited  
181 polymorphism among the genomes of the 23 dams with extreme ratios of offspring sex-reversal  
182 were kept for the final analysis, i.e. 854,157 SNPs for Omy1, 965,320 SNPs for Omy12 and  
183 492,959 SNPs for Omy20.

184

## 185 **Genotyped Samples**

186 To validate the QTL detected by Fraslin et al. [27] as linked to spontaneous sex-reversal of XX  
187 trout, a set of 192 SNPs were selected to design two 96 SNP genotyping arrays using microfluidic  
188 real-time PCR Fluidigm Kasp chemistry. The 192 SNPs (see S1 Supplementary Tables, S2  
189 Table) correspond to 140 SNPs in the 2 main QTL identified on chromosome Omy1, 19 SNPs  
190 in a putative QTL region detected on Omy12 and 33 SNPs in a large 6 Mb region with putative

191 QTL on Omy20. These arrays were used to genotype 315 fish from 6 diverse French  
 192 populations of rainbow trout named by the letters A to F (Table 1) with at least 30 XX fish,  
 193 including a minimum of 8 neomales, per population. Population A was made up of XX sibs  
 194 from the same birth cohort as the parents used to produce the QTL discovery population [27].  
 195 Respectively, fish from populations B and C and those from populations E and F came from  
 196 two other French breeding compagnies, while fish from population D came from a commercial  
 197 site using fry from unknown origin.

198 Pieces of caudal fin sampled from those 315 fish were sent to Gentyane genotyping platform  
 199 (INRAE, Clermont-Ferrand, France) for genotyping after DNA extraction using the DNA  
 200 advance kit from Beckman Coulter following manufacturer instructions.

201 After quality control, 19 SNPs were eliminated (including 14 from Omy1) from the analyzes  
 202 of all the populations because their genotyping rate was below 90%.

203

204 **Table 1** Description of the 6 rainbow trout XX populations used for QTL validation

Population	A	B	C	D	E	F
Number of genotyped fish	49	30	64	77	45	50
Phenotyped as female	34	22	32	29	31	32
Phenotyped as male (including intersex)	15	8	32	48	14	18
Phenotyped as pure male (no intersex)	8	nr	17	32	6	nr

205 nr: values not reported by the breeding compagny

206

### 207 **Statistical analysis for validation of the existence of the QTL**

208 An exact Fisher test was run for each of the 6 validation populations using the `fisher.test`  
 209 function in R version 4.3.1 [36] which implements the method developed by Mehta and Patel  
 210 [37, 38] and improved by Clarkson, Fan and Joe [39] to test the independence of genotypes at  
 211 each SNP and phenotypic sex of fish.

212 This test was applied for the set of 173 validated SNPs across all six populations. In any of  
 213 these populations, ~170 SNPs were polymorph and had a genotyping rate above 90%. To  
 214 account for multiple testing using Bonferroni correction, we considered that SNPs were

215 significant at the genome level in a given population when  $p\_value < 0.0005$  (i.e.  $\log_{10}(p\_value)$   
216  $< 3.3$ , assuming that about 100 SNPs were not highly redundant among  $\sim 170$  tested for each  
217 population) and had only a putative effect when  $p\_value < 0.005$  (i.e.  $\log_{10}(p\_value) < 2.3$ ).

218 To confront the results between the validation and discovery populations, the latter threshold  
219 value was also used to detect, by exact Fisher test, the association of all the sequence variants  
220 observed in the 4 QTL regions with the average progeny sex-ratio of the 23 extreme dams of  
221 the discovery population. A second, more stringent, threshold ( $p\_value < 0.002$ ) was considered  
222 to focus on the main significant associations with sequence variants in the discovery population.  
223 Because of the strong linkage disequilibrium observed across SNPs along the chromosome [40],  
224 the QTL region on Omy20 was strongly expanded to test, one-by-one, all the variants spanning  
225 the chromosome from 27 to 38 Mb.

226

## 227 **Statistical approaches for refinement of QTL locations in the discovery population**

### 228 *Machine learning approach using Random Forests*

229 Random forest (RF) is a machine learning method that aggregates complementary information  
230 from an ensemble of classification or regression trees trained on different bootstrap samples  
231 (animals) drawn with replacement from the original data set [41]. First, every tree is built using  
232 a bootstrap sample of the observations. Second, at each node, a random subset of all predictors  
233 (the size of which is referred to as  $m_{try}$  hereafter) is chosen to determine the best split rather  
234 than the full set. Therefore, all trees in a forest are different. A particularity of the RF is the out-  
235 of-bag data, which corresponds to the animals not included (roughly 1/3) in the bootstrap  
236 sampling for building a specific tree. It can be used as an internal validation set for each tree,  
237 which allows the computation of RF error rate based on misclassified animals in the out-of-bag  
238 data. RF can be applied successfully to “large  $p$ , small  $n$ ” problems, and it is also able to capture  
239 correlation as well as interactions among independent variables [42].

240 We considered the five following genomic zones to define haplotypes of subsequent SNPs  
241 spanning from 67 to 69Mb on Omy1, from 8.5 to 9.5Mb on Omy12, and, on Omy20: 27-29Mb,  
242 33.5-35.5 Mb, and 36-37Mb. In total, we considered 21,310 variants for Omy1, 12,191 for  
243 Omy12, and 56,643 for Omy20, whatever their distance to genes in the QTL regions.

244 We recoded the genotypes depending of the number of reference alleles present in the genotype:  
245 “2” for a homozygous reference genotype, “1” for a heterozygous genotype, “0” for the  
246 alternative homozygote; missing genotypes were encoded “5”.

247 To minimize the RF error rates, we ran a preliminary set of analyses to establish the optimum  
248 SNP window size (swind), the number of sampled trees (ntree), and the number of sampled  
249 predictor variables (mtry) chosen at random to split each node. For any analysis procedure, the  
250 more highly correlated the variables are, the more they can serve as surrogates for each other,  
251 weakening the evidence for association for any single correlated variable to the outcome if all  
252 are included in the same model. Therefore, to avoid redundancy of information and limit the  
253 number of variables to analyze, we built some haplotypic genotypes by retaining one SNP out  
254 of two and gathering them in the same haplotype across a window of 80 initial SNPs  
255 (correspond to swind=40). We varied swind from 20 to 200 subsequent SNPs, ntree from 100  
256 to 1,000 and mtry from 5 to 20. The final parameter set (swind=40, ntree=400 and mtry=7) was  
257 chosen as consistently giving the lower median error rate over 50 runs for any of the three tested  
258 chromosomes. A final set of 246 haplotypes for Omy1 (131 for Omy1\_a and 115 for Omy1\_b),  
259 152 haplotypes for Omy12 and 707 haplotypes for Omy20 (247, 288, and 172, respectively for  
260 Omy20\_a, Omy20\_b and Omy20\_c) were then analyzed by averaging results across 100 RF  
261 runs. We used the package RandomForest 4.18 [43] in R version 4.3.1 [36].

262 Haplotypes were ranked by decreasing order of importance in the prediction model, using the  
263 mean decrease in Gini index. The Gini index measures the probability for a specific feature of  
264 being misclassified when chosen randomly. The higher is the value of mean decrease in Gini  
265 index, the higher is the importance of the variable in the model.

266 A list of 45 positional candidate genes was established based on results of the best ranked  
267 haplotypes as well as the preliminary Fisher exact tests on SNP.

268

### 269 *Discriminant Analysis of Principal Components to identify the most promising variants*

270 The Discriminant Analysis of Principal Components (DAPC) helps explore the genetic  
271 structure of biological populations. This multivariate method consists in a two-step procedure:  
272 1) genetic data are centred and used in a Principal Component Analysis (PCA); 2) a Linear  
273 Discriminant Analysis is applied to the principal components of PCA. A basic matrix operation  
274 is used to express discriminant functions as linear combination of alleles, allowing one to  
275 compute allele contributions, i.e. to identify the most interesting variants [44].

276 We considered the 33,731 SNPs that were in the list of genes identified by RF or Fisher's exact  
277 test. We also studied the SNPs in the close upstream and downstream regions of these genes  
278 ( $\pm 2$  kb). Filtering out SNPs with call rate below 100% using PLINK 1.9 [45], we applied DAPC

279 on a final set of 27,828 variants without any missing genotypes. We used the R adegenet  
280 package [46] to find by DAPC the main variants that best discriminate between dams with high  
281 and dams with low sex-reversal ratios in their offspring.

282

### 283 **Annotation of genes and variants**

284 Genes within QTL regions were annotated using the NCBI *O. mykiss* Arlee genome assembly  
285 USDA\_OmykA\_1.1. (GCA\_013265735.3) (Gao et al. 2021). In addition to NCBI gene  
286 summaries, functional information for genes was extracted from the human gene database  
287 GeneCards® (<https://www.genecards.org/>) that also includes protein summaries from  
288 UniProtKB/Swiss-Prot (<https://www.uniprot.org/uniprotkb/>).

289 SNP annotation was performed for the 33,731 SNPs identified for the list of genes of interest,  
290 applying the software SNPEff4.3T [47] on the NCBI annotation file annotation released for *O.*  
291 *mykiss* Arlee genome reference assembly (GCA\_013265735.3). These annotations permitted to  
292 retrieve any information about the predicted impact of the variants on the genes, and, in some  
293 cases, on the resulting proteins. Interesting variants were further studied using the Genome Data  
294 Viewer of NCBI to indicate their intronic or exonic positions within the genes.

295

## 296 **Results**

### 297 **Validation of the existence of the 4 QTL in various French rainbow trout populations**

298 Among the 173 SNPs that could be tested for their association with spontaneous sex-reversal  
299 in the validation populations, 117 had a significant effect in at least one population. Among the  
300 latter, 50 SNPs on Omy1 had at least an effect in population A that was genetically close to the  
301 QTL discovery population. Most of these 50 SNPs had also an effect observed in populations  
302 B, C and E. On the contrary, none of the SNPs tested on Omy12 or Omy20 had an effect in  
303 population A, probably because of sampling issues for low frequency variants in the population.  
304 However, 7 SNPs on Omy12 (out of 17 that could be tested) had an effect in populations C, D,  
305 E and F; and 13 SNPs on Omy20 (out of 30 that could be tested) had an effect in populations  
306 C, D and F.

307 The list of the 90 SNPs that were at least associated with a putative effect in 2 validations  
308 populations are given in S1 Supplementary Tables, S3 Table with indication of the level of  
309 significance in those populations. Table 2 reported the 63 SNPs that had at least a putative effect

310 in four validation populations, or a clear significant effect in at least two populations. Out of  
 311 these 63 SNPs, 50 were located within genes or in their close vicinity (< 5kb) and NCBI  
 312 annotation for those genes are then reported in Table 2.

313

314 **Table 2** Most significant SNPs in at least 2 out of the 6 validation populations named A to F  
 315 (names in bold correspond to a p\_value < 0.0005, others to a p\_value < 0.005)

Omy	Position	P1	P2	P3	P4	NCBI gene annotation (Genecards symbol)
1	67186650	<b>A</b>	<b>B</b>	<b>C</b>	F	synapse differentiation-inducing gene protein 1 ( <i>syndig1</i> )
1	67187724	<b>A</b>	<b>B</b>	<b>C</b>	F	synapse differentiation-inducing gene protein 1 ( <i>syndig1</i> )
1	67273221	<b>A</b>	<b>B</b>	<b>C</b>	F	1.982 kb upstream acyl-CoA synthetase short chain family member 1 ( <i>acss1</i> )
1	67319795	<b>A</b>	<b>B</b>	<b>C</b>	F	Close to visual homeobox 1 ( <i>vsx1</i> )
1	67340956	<b>A</b>	<b>B</b>	<b>C</b>	F	Close to ectonucleoside triphosphate diphosphohydrolase 6-like ( <i>entpd6</i> )
1	67353129	A	<b>B</b>	<b>C</b>	<b>F</b>	2.941 kb upstream <i>entpd6</i>
1	67387599	<b>B</b>	<b>C</b>	<b>F</b>		barrier-to-autointegration factor ( <i>banf1</i> )
1	67387902	<b>B</b>	<b>C</b>	<b>F</b>		barrier-to-autointegration factor ( <i>banf1</i> )
1	67389450	<b>B</b>	<b>C</b>	<b>F</b>		phosphorylase, glycogen; brain ( <i>pygb</i> )
1	67395895	A	<b>B</b>	<b>C</b>	F	phosphorylase, glycogen; brain ( <i>pygb</i> )
1	67406669	B	<b>C</b>			phosphorylase, glycogen; brain ( <i>pygb</i> )
1	67410770	B	<b>C</b>			phosphorylase, glycogen; brain ( <i>pygb</i> )
1	67413276	B	<b>C</b>	<b>D</b>		phosphorylase, glycogen; brain ( <i>pygb</i> )
1	67434621	A	<b>B</b>	<b>C</b>		abhydrolase domain containing 12, lysophospholipase ( <i>abhd12</i> )
1	67435714	<b>B</b>	<b>C</b>			abhydrolase domain containing 12, lysophospholipase ( <i>abhd12</i> )
1	67461050	<b>B</b>	<b>C</b>			ninein-like ( <i>ninl</i> )
1	67464776	<b>A</b>	<b>B</b>	<b>C</b>	<b>F</b>	ninein- like ( <i>ninl</i> )
1	67494288	<b>B</b>	<b>C</b>			ninein- like ( <i>ninl</i> )
1	67522146	<b>B</b>	<b>C</b>			Closest gene: pleckstrin and Sec7 domain containing a ( <i>psda</i> )
1	67535219	<b>B</b>	<b>C</b>			pleckstrin and Sec7 domain containing a ( <i>psda</i> )
1	67546988	<b>B</b>	<b>C</b>			pleckstrin and Sec7 domain containing a ( <i>psda</i> )
1	67584462	<b>B</b>	<b>C</b>			shieldin complex subunit 2 ( <i>shld2</i> )
1	67626764	<b>B</b>	<b>C</b>			glutamate dehydrogenase 1 ( <i>glud1</i> )
1	68248684	A	B	<b>C</b>	E	Closest gene: solute carrier family 2 member 15a ( <i>slc2a15</i> )
1	68270478	A	B	<b>C</b>	E	4.901 kb upstream of solute carrier family 2 member 15a ( <i>slc2a15</i> )
1	68305255	<b>A</b>	<b>B</b>	<b>C</b>		Closest gene: solute carrier family 2 member 15a ( <i>slc2a15</i> )
1	68310800	<b>A</b>	<b>B</b>	<b>C</b>		Closest gene: fibroblast growth factor 8a ( <i>fgf8</i> )
1	68312110	<b>A</b>	<b>B</b>	<b>C</b>		4.222 kb upstream of fibroblast growth factor 8a ( <i>fgf8</i> )
1	68315508	A	B	<b>C</b>		fibroblast growth factor 8a ( <i>fgf8</i> )
1	68316754	<b>A</b>	<b>B</b>	<b>C</b>		fibroblast growth factor 8a ( <i>fgf8</i> )

1	68329760	A	C		Closest gene: fibroblast growth factor 8a ( <i>fgf8</i> )
1	68330279	A	B	C	Closest gene: F-box and WD repeat domain containing 4 ( <i>fbxw4</i> )
1	68339634	A	C		F-box and WD repeat domain containing 4 ( <i>fbxw4</i> )
1	68466812	C	F		0.138 kb downstream of helicase, lymphoid specific ( <i>hells</i> )
1	68467934	A	B	C	helicase, lymphoid specific ( <i>hells</i> )
1	68470380	A	B	C	helicase, lymphoid specific ( <i>hells</i> )
1	68495916	A	B	C	0.826 kb downstream of uncharacterized LOC118965305
1	68496653	A	B	C	1.563 kb downstream of uncharacterized LOC118965305
1	68501373	A	B	C	Closest gene : uncharacterized LOC118965305
1	68503159	A	B	C	Closest gene : uncharacterized LOC118965305
1	68504970	A	B	C	Closest gene : uncharacterized LOC118965305
1	68505133	A	B	C	Closest gene : uncharacterized LOC118965305
1	68505220	A	B	C	Closest gene : uncharacterized LOC118965305
1	68505500	A	B	C	Closest gene : uncharacterized LOC118965305
1	68509786	A	B	C	Closest gene : LOC110527930
1	68511322	A	B	C	Closest gene : LOC110527930
1	68519589	A	B	C	0.205 kb upstream of LOC110527930
1	68521100	A	B	C	uncharacterized protein LOC110527930
1	68522155	A	B	C	uncharacterized protein LOC110527930
1	68524318	A	B	C	uncharacterized protein LOC110527930
1	68529520	A	B	C E	uncharacterized protein LOC110527930
1	68529535	A	B	C E	uncharacterized protein LOC110527930
1	68529584	A	B	C E	uncharacterized protein LOC110527930
1	68529621	A	B	C E	uncharacterized protein LOC110527930
1	68529794	A	B	C E	uncharacterized protein LOC110527930
1	68538278	A	B	C	uncharacterized protein LOC110527930
1	68546628	A	B	C	uncharacterized protein LOC110527930
1	68605451	A	B	C	collagen alpha-1(XIII) chain ( <i>coll3a1</i> )
12	9164517	D	F		hyperpolarization-activated cyclic nucleotide-gated cation channel 1 ( <i>hcn1</i> )
20	34350986	D	F		3-hydroxymethyl-3-methylglutaryl-CoA lyase-like 1 ( <i>hmgcll1</i> )
20	34353785	D	F		3-hydroxymethyl-3-methylglutaryl-CoA lyase-like 1 ( <i>hmgcll1</i> )
20	37496331	C	F		ENSOMYG00000068776
20	37497790	D	F		ENSOMYG00000068776

316

### 317 Extension of the QTL region boundaries in the discovery population

318 Because most of the SNPs located at the boundaries of the QTL regions reported by Fraslin et  
319 al. [27] had a significant effect in some of the validation populations, we applied Fisher's exact

320 test on the dams' average progeny phenotypes in the discovery population for all sequence  
 321 variants located in QTL regions enlarged by 0.5 Mb on each side. Based on these tests, we  
 322 defined the new extended boundaries for the QTL that are given in Table 3. In particular, we  
 323 enlarged the search of QTL in 3 different regions of Omy20, spanning from 27 to 29 Mb, 33.5  
 324 to 35.5 Mb, and 36 to 38 Mb, respectively for Omy20\_a, Omy20\_b and Omy20\_c.

325

326 **Table 3** Extended boundaries of sex-reversal QTL regions in the current study (location on  
 327 Arlee reference genome) compared to QTL intervals reported by Fraslin et al. [27]

Chromosome	Region name	QTL name	Start of QTL (bp)	End of QTL (bp)
Omy1	Z1	Omy1_a	67 000 061	67 960 401
Omy1	Z1	Omy1_b	67 960 402	68 999 919
Omy12	Z2	Omy12	8 500 067	9 499 932
Omy20	Z3	Omy20_a	27 000 149	28 998 983
Omy20	Z4	Omy20_b	33 508 998	35 499 146
Omy20	Z5	Omy20_c	36 000 252	36 999 928

328

329 **Table 4** Most significant SNPs by Fisher's exact test ( $p\_value \leq 0.002$ ) considering the 23  
 330 dams with extreme progeny sex-ratios in the QTL discovery population.

Omy	Position	REF	ALT	p-value	NCBI gene annotation (Genecards symbol)
1	67931300	A	C	0,0017	opsin 4a ( <i>ops4a</i> )
1	68330279	C	T	0,0017	8.758 kb upstream of F-box and WD repeat domain containing 4 ( <i>fbxw4</i> )
1	68832468	G	A	0,0013	Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1 ( <i>gbf1</i> )
20	27841889	A	G	0,0020	ARFGEF family member 3 ( <i>arfgef3</i> )
20	27841897	T	A	0,0020	ARFGEF family member 3 ( <i>arfgef3</i> )
20	28044820	G	C	0,0017	v-akt murine thymoma viral oncogene homolog 3a ( <i>akt3a</i> )
20	28206089	A	G	0,0017	SHH signaling and ciliogenesis regulator sdccag8 ( <i>sdccag8</i> )
20	28224422	C	*,T	0,0020	SHH signaling and ciliogenesis regulator sdccag8 ( <i>sdccag8</i> )
20	28270362	A	G	0,0017	centrosomal protein 170aa ( <i>cep170aa</i> )
20	28333748	G	T	0,0017	0.117 kb upstream leucine-rich pentatricopeptide repeat containing ( <i>lrpprc</i> )

20	28695801	A	G	0,0006	calmodulin-lysine N-methyltransferase ( <i>camkmt</i> )
20	28775897	A	T	0,0013	30.316 kb downstream calmodulin-lysine N-methyltransferase ( <i>camkmt</i> )
20	28803467	A	C,T	0,0005	45.584 kb upstream SIX homeobox 3a ( <i>six3a</i> )
20	33670075	G	A,T	0,0017	KH domain-containing, RNA-binding, signal transduction-associated protein 2 ( <i>khdrbs2</i> )
20	34018207	C	T	0,0006	Dystonin ( <i>dst</i> )
20	35051101	G	T	0,0017	CUB and sushi domain-containing protein 1 ( <i>csmd1</i> )
20	36030963	G	T	0,0013	0.468 kb upstream uncharacterized ncRNA (LOC110499423)
20	37557112	C	G	0,0006	ENSOMYG00000068776*
20	37995993	T	G,*	0,0017	79.980 kb upstream charged multivesicular body protein 6 ( <i>chmp6</i> )

331 \* ENSOMYG00000068776 on Omy20 (37,056,423 ... 37,561,145) is a paralog of *rbfox3* located on Omy23

332 In total, 246 SNPs had a  $p\_value \leq 0.005$  (see S1 Supplementary Tables, S4 Table), and the 19  
333 SNPs with a more stringent  $p\_value \leq 0.002$  were mainly located within genes or in their close  
334 vicinity (Table 4).

335

### 336 **Identification of the most relevant haplotypes and genes explaining sex-reversal in the** 337 **discovery population**

338 Initially, we ran a RF analysis for each of the 3 chromosomes with all their QTL regions jointly  
339 analyzed. Respectively for the RF on Omy1, Omy12 and Omy20, we presented the 30, 15 and  
340 30 best ranked haplotypes based on Gini index as well as the corresponding genes, in S5, S6  
341 and S7 Tables. For the 2-Mb region on Omy1 (246 haplotypes spanning from 67 to 69 Mb), the  
342 median error rate was 28% with a minimum error rate of 15% and a maximum rate of 37% of  
343 fish phenotypes wrongly classified as the opposite sex across 100 RF simulations. While the  
344 RF analysis of error rate indicated that the 2 Mb tested on Omy1 were really useful to  
345 discriminate the dams with high and low progeny sex-reversal ratios, the median error rate  
346 derived from the RF analysis on Omy12 (152 haplotypes spanning 8.5 to 9.5 Mb) was 42%  
347 (ranging from 29% to 51% across the 100 RF runs), indicating that the best haplotypes and  
348 genes identified on this chromosome were not so strongly associated with sex-reversal. In terms  
349 of error rate, the RF analysis on the 707 haplotypes covering the QTL regions on Omy20 (247  
350 haplotypes from 27 to 29 Mb, 288 haplotypes from 33.5 to 35.5 Mb, and 172 haplotypes from  
351 36 to 37 Mb) gave intermediate results in-between the two previous analyses, misclassifying

352 one third of the dams on average with a minimum error rate of 29% and a maximum rate of  
353 38% across 100 RF simulations.

354 These results are consistent with the higher proportions of genetic variance explained by QTL  
355 we estimated on Omy1 than on Omy12 or Omy20 [27]. In addition, we considered all together  
356 the 1,105 haplotypes covering the QTL regions along the three chromosomes in a last RF  
357 analysis. The empirical distribution of error rates for the full RF analysis gave median,  
358 minimum and maximum values similar to the ones of the RF on Omy20 (33%, 29%, and 38%,  
359 respectively) while we were expecting that these values will be at least as low as for the RF on  
360 Omy1. Due to the very small size of our dataset, it is likely that we had too many variables in  
361 the full RF analysis, creating some « noise » by considering over 1,100 haplotypes together in  
362 the analysis to classify our 23 individuals.

363 Looking at the locations (Fig.1) of the best ten RF ranked haplotypes on the 2 Mb evaluated on  
364 Omy1 (see S1 Supplementary Tables, S5 Table), they fell within the nine following genes, six  
365 of them being identified with significant effects on sex-reversal in several validation  
366 populations (Table 2) or within some significant SNPs in the discovery population (Table 4):  
367 *sdhaf4*, *tlx1*, *syndig1* and *vsx1* genes located in QTL Omy1\_a; *cep68*, *fbxw4*, *hells*,  
368 LOC110527930, and *gbf1* in QTL Omy1\_b. The first best-ranked haplotype included two genes  
369 from QTL Omy1\_b (*sdhaf4* and *tlx1*), while the second and third best-ranked haplotypes were  
370 in the sequence of the *syndig1* gene on Omy1\_a.

371

372 **Fig. 1** Positions of genes within QTL regions a and b on Omy1 with significant SNPs positions  
373 in discovery or validation population as well as best ranked haplotypes

374

375 These results are not fully consistent with our first GWAS [27] using progeny imputed  
376 genotypes and Bayesian mixed model methodologies. In this previous analysis, the *syndig1*  
377 gene was not included in the credibility interval for Omy1\_a. Although there was no SNP within  
378 this gene among the most significant SNPs ( $p\_value \leq 0.002$ ) in the discovery population, 2  
379 SNPs had, however, a  $p\_value < 0.005$  (see S1 Supplementary Tables, S4 Table). In addition,  
380 *syndig1* exhibited highly significant SNPs in four validation populations, including population  
381 A. All these results make *syndig1* a very plausible candidate gene for QTL Omy1\_a. As regards  
382 to the initial QTL Omy1\_b, there is now an extended credibility interval, including genes before  
383 *slc2a15a* such as *tlx1* and *cep68*, and genes after the uncharacterized LOC110527930 such as

384 *gbf1*. While LOC110527930 initially was our favourite causative gene for Omy1\_b, genes such  
385 as *tlx1*, *cep68*, *fbxw4* and *hells* are now preferred within Omy1\_b region (Fig. 3).

386 On the 1 Mb tested on Omy12 (see S1 Supplementary Tables, S6 Table), the first three best  
387 haplotypes to explain sex-reversal were within the same two *hcn1* and *ccbe1* genes that were  
388 also identified in three populations unrelated to the discovery population (see S1 Supplementary  
389 Tables, S3 Table).

390 On the 5 Mb tested on Omy20 (Fig.2), unfortunately, there was no variant tested in the  
391 validation populations for the eight genes corresponding to the ten best ranked haplotypes under  
392 RF analysis (see S1 Supplementary Tables, S7 Table) that were in decreasing order of  
393 importance: *lrrc59*, *csmd1*, *khdrbs2*, *abcg5*, *akt3*, *rgs17*, *casin2* and *dst*. However, some highly  
394 significant SNPs are detected in the discovery population for four of these genes (*akt3*,  
395 *khdrbs2*, *dst* and *csmd1*, see Table 4).

396

397 **Fig. 2** Positions of genes within QTL regions a, b and c on Omy20 with significant SNPs  
398 positions in discovery or validation population as well as best ranked haplotypes

399

400 As regards to the ranking of the best haplotypes to explain sex-reversal across all the QTL  
401 regions spanning the 3 chromosomes (see S1 Supplementary Tables, S8 Table), the first one  
402 was positioned within *lrrc59* gene on Omy20\_c, the second and third ones included *sdhaf4*  
403 and *tlx1* on Omy1\_b, and the fourth one was located within the *csmd1* gene on Omy20\_b. The  
404 next four best-ranked haplotypes were all on Omy20, located within *abcg5*, *casin2*, *khdrbs2*,  
405 and *dst* genes. Some genes identified on Omy1 in validation populations (*syndig1*, *hells*, *fbxw4*  
406 and LOC110527930) appeared then as the subsequent haplotypes ranked with very similar Gini  
407 indices (Fig. 1). Noticeably, most of the best-ranked haplotypes concerned sequences within  
408 genes, and were consistently ranked between the full RF analysis and the RF limited to their  
409 chromosomal location.

410 Therefore, based on the results of the 60 best ranked haplotypes on the Gini index of this full  
411 RF analysis, as well as the additional results from Fisher's exact tests (Tables 2 and 4), we  
412 identified a list of 45 genes (see S1 Supplementary Tables, S9 Table) whose variants in their  
413 sequence or in their close vicinity (2kb) were the most likely to explain a significant part of sex-  
414 reversal phenotypes.

415

## 416 **Identification of the potential causative variants combining DAPC and SNP annotations**

417 We considered the corresponding sequence variants of those genes ( $\pm 2$  kb) in a DAPC in order  
418 to find the best 100 SNPs to discriminate among the 23 dams the ones with high progeny sex-  
419 reversal ratios from those with low ratios. Running a simple ACP on all those variants do not  
420 allow to correctly discriminate the two groups of dams as shown by Fig.3A. On the contrary,  
421 the DAPC allowed us to perfectly discriminate the two groups as demonstrated in Fig.3B.

422

## 423 **Fig. 3 PCA of variants between “Fem” and “Mal” dam groups on 45 candidate genes**

424 **A. PCA of the variants in the 45 candidate genes ( $\pm 2$  kb) and B. PCA of the 100 best**  
425 **variants selected by DAPC to discriminate the “Fem” group of the seven dams with low**  
426 **progeny sex-reversal ratios and the “Mal” group of the 16 dams with high progeny sex-**  
427 **reversal ratios.**

428

429 In S1 Supplementary Tables, S10 Table are given the genotypes for the 100 best SNPs that  
430 were genotyped for the 23 dams among 27, 828 SNPs (without missing genotypes out of 33,731  
431 SNPs), as well as for 5 additional dams with phenotypic sex recorded on at least 6 offspring in  
432 the discovery population (see S1 Supplementary Tables, S1 Table). We considered these 5 dams  
433 as an internal validation: 3 dams (noted AS, AT and AU) had a high progeny sex-reversal ratio  
434 while no sex-reversed offspring were observed for the last 2 dams (noted BS and BT). Those  
435 best 100 SNPs were located within or nearby the downstream or upstream regions of 16 genes  
436 belonging to the list of the 45 genes previously identified by RF and Fisher’s exact test: on  
437 *Omy1*, *syndig1*, *vsx1*, and *myoz1*; on *Omy12*, *hcn1* and *ccbe1*; on *Omy20*, *arfgef3*, *cep170aa*,  
438 *abcg5*, *khdrbs2*, *dst*, *hmgcll1*, *cilk1*, *rps6ka2*, LOC110499423, *lrrc59* and *casin2*. Most of  
439 these 100 SNPs were located in intronic regions of the genes, two were synonymous mutations  
440 and none were splice donor variants in introns or missense variants in exons. When focusing  
441 on the 50 best SNPs to discriminate between the two groups of dams, only *hcn1* remained on  
442 *Omy12*; *syndig1* and *myoz1* were kept for *Omy1\_a* and *Omy1\_b*, respectively; and, all genes,  
443 except LOC110499423, remained for QTL regions on *Omy20*.

444

## 445 **Discussion**

446

#### 447 **Significance of the different QTL regions across French rainbow trout populations**

448 Considering the list of the 100 best-ranked SNPs under DAPC in the discovery population  
449 confirms the hypothesis of two QTL on Omy1, one QTL on Omy12, and lead us to suggest the  
450 existence of three distinct QTL on Omy20. All the QTL regions previously identified [27] are  
451 confirmed as playing a role on sex-reversal in at least two populations different from the  
452 discovery one. The most significant SNPs across several populations are located within genes  
453 or in their close vicinity.

454 Regarding QTL Omy1\_a, the candidate genes that appear as the most relevant to explain sex-  
455 reversal across populations were *syndig1*, *acss1*, *vsx1*, *entpd6*, *pygb* and *ninl*. In the initial study  
456 [27], the peak SNPs for this QTL were located either within the *pygb* or *ninl* gene, depending  
457 on the statistical approach. In particular, Fraslin et al. [27] mentioned a SNP located at  
458 63,543,061 bp (on Swanson reference genome) within the *ninl* gene and annotated has a  
459 missense variant. This SNP was not genotyped in the validation population, but in this study, a  
460 very close SNP located at 63,530,416 bp on Swanson reference genome (corresponding to  
461 67,434,621 bp on Arlee reference genome) has a putative effect in population A and a  
462 significant effect at the genome level in populations B and C (see S1 Supplementary Tables, S3  
463 Table).

464 As regards to the main QTL Omy1\_b, Fraslin et al. [27] identified it with all peak SNPs located  
465 in the intergenic region between the *hells* gene and the uncharacterized protein LOC110527930,  
466 the latter being proposed as the most relevant positional candidate for explaining 14% of the  
467 genetic variance of sex-reversal in the discovery population. However, in the present study,  
468 other genes should also be considered as candidate genes in several validation populations.  
469 Some were located before the *hells* gene (*slc2a15a*, *fgf8*, and *fbxw4*), while the most distal one,  
470 *coll3a1*, was located after LOC110527930. This observation, as well as the identification of  
471 *syndig1* in QTL Omy1\_a that was positioned before the first QTL interval defined in Fraslin et  
472 al. [27] were the motivation to further study the sequence variants in larger genomic regions.

473 In the previous study [27], the QTL on Omy12 was detected using the 57K chip, but not  
474 considering imputed sequence variants in the discovery population. In the present study, this  
475 QTL is not validated in population A although this population is closely related to the discovery  
476 one. Nevertheless, the QTL is validated in population D and F with a significant SNP identified

477 within *hcn1* gene. We can hypothesize that the QTL is then segregating in the discovery  
478 population, but either at a too low frequency or too small effect to be statistically detected.

479 The QTL region on Omy20 is not validated in population A, but two QTL are segregated on  
480 Omy20 in populations C, D and F, with significant SNPs located in two distant genes: *hmgcll1*  
481 located around 34.4 Mb, and ENSOMYG00000068776 gene located around 37.5 Mb. The same  
482 hypothesis as for the QTL on Omy12 can thus be proposed for the absence of statistical  
483 validation in population A. In addition, in an INRAE experimental rainbow trout line, a QTL  
484 for spontaneous sex-reversal was also identified on Omy20 (previously named RT17) but with  
485 a very imprecise position [28].

486 In the next paragraphs, we describe the putative functional roles of our best positional candidate  
487 genes for the two QTL on Omy1 as well as for the most promising genes for the three different  
488 QTL on Omy20. Other positional candidate genes on Omy12, Omy20\_a and Omy20\_b with  
489 less functional evidence are presented in an additional section [see S1 Appendix].

490

#### 491 ***syndig1* as the best positional and functional candidate gene for QTL Omy1\_a**

492 It is very likely that one or a few variants in the *syndig1* (synapse differentiation-inducing gene  
493 protein 1) sequence or in its close downstream region (< 1kb) are the causative variants  
494 explaining a part of the sex-reversal phenotype. First of all, 4 SNPs were significant in two to  
495 four validation populations (see S1 Supplementary Tables, S3 Table). In the discovery  
496 population, the RF analysis highlighted 3 haplotypes in the gene (see S1 Supplementary Tables,  
497 Supplementary table 5), among which 3 missense variants were annotated. The DAPC ranked 3  
498 SNPs within or downstream *syndig1* (see S1 Supplementary Tables, S10 Table) among the 100  
499 best SNPs to discriminate the low and high sex-reversed progeny dams. However, among the  
500 50 best SNPs, the only 2 retained were located in downstream intergenic positions. We  
501 speculate that these latter positions were kept to the detriment of SNPs within *syndig1* due to  
502 linkage disequilibrium between SNPs of the two close genes *syndig1* and *vsx1* (visual system  
503 homeobox 1). Indeed, a unique SNP was retained from *vsx1* in the list of the 100 best SNPs and  
504 was filtered out in the list of the 50 best ones. We may speculate that *vsx1* may play a minor  
505 additional role to explain sex-reversal as significant SNPs were detected both within gene in  
506 the discovery population and in an upstream region in 4 validation populations. It is worth  
507 mentioning the existence of 5 missense variants in *vsx1*. This gene is expressed during early  
508 embryogenesis as well as in the adult retina. It participates in regulating retinal progenitor

509 proliferation, differentiation and functional maintenance of bipolar cells [48, 49, 50]. As far as  
510 we know, no direct link has been suggested between *vsx1* and sex determination in any species.

511 As regards to *syndig1*, this gene encodes a protein that belongs to the interferon-induced  
512 transmembrane family of proteins. It plays a critical role during synapse development to  
513 regulate *ampar* ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) content and  
514 have a role in postsynaptic development and maturation [51, 52]. Although *syndig1* expression  
515 is mainly restricted in the embryonic retina in mice [53], its expression has been detected also  
516 in granulosa cells at the small antral stage in sheep [54] suggesting a potential role in follicular  
517 development in vertebrates.

518

### 519 **Multiple candidate genes for QTL Omy1\_b, with two favourites: *tlx1* and *hells***

520 While this QTL was identified as the one explaining the largest proportion of genetic variance  
521 of spontaneous sex-reversal [27], the previous study pointed on different candidate genes than  
522 in the current one. In the first study, the three putative functional candidate genes were *fgf8a*,  
523 *cyp17a1* and LOC110527930 based on progeny imputed genotypes, while the current analysis  
524 of dams' sequences pointed to genes located upstream, prioritizing *tlx1* (T cell leukemia  
525 homeobox 1) for RF analysis and *myoz1* (myozenin-1) for DAPC. It seems, in fact, that two  
526 different QTL should be distinguished within Omy1\_b, Omy1\_b1 and Omy1\_b2, on each side  
527 of *cyp17a1*. Although *cyp17a1* is known to be involved in gonad masculinization in the  
528 common carp and zebrafish [55, 56], no variant in this gene was significantly associated with  
529 sex-reversal in rainbow trout, neither in the discovery population, nor in the validation ones.  
530 The new analyses highlighted genes, *myoz1* and *tlx1*, that are located in region Omy1\_b1 before  
531 the start of the QTL region previously defined at *slc2a15a* gene [27]. However, it is possible  
532 that a cis-regulatory element within the QTL region could control at long-distance the  
533 expression of *cyp17a1*, influencing sex differentiation in rainbow trout. Such long-distance cis-  
534 regulatory elements have been found to control sex like in the Sry-deficient Amami spiny rat  
535 where the male-specific upregulation of *sox9* is controlled by an enhancer region 0.5 Mb  
536 upstream of *sox9* [57].

537 The initial Omy1\_b is now defined as a second genomic region Omy1\_b2, spanning from *fbxw4*  
538 to *gbf1* (including *hells* and LOC110527930). While the region Omy1\_b2 was clearly  
539 associated with sex-reversal in four validation populations, there was no SNP among the 100  
540 best ranked by DAPC. The RF analysis, however, identified some haplotypes among the 10

541 best ones to explain sex-reversal on Omy1, putting similar emphasis to four genes: *fbxw4*, *hells*,  
542 LOC110527930 and *gbfl*. In addition, Fisher's exact test gave a p\_value  $\leq 0.002$  for a SNP  
543 within *gbfl* (Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1) in the  
544 discovery population, while SNPs in this gene were unfortunately not tested in the validation  
545 populations as outside the limits of the initial credibility interval for the QTL. As far as *fbxw4*  
546 (F-box and WD repeat domain containing 4) gene is concerned, no direct link with gonad  
547 development, oogonia or spermatogonia could be established. This gene is a member of the  
548 F-box/WD-40 gene family that may participate in Wnt signaling which is a pathway involved  
549 in gonadal differentiation (see S1 Supplementary Tables, S9 Table). As regards to *gbfl*, it is  
550 one of the guanine-nucleotide exchange factors (GEF) for members of the Arf family of small  
551 GTPases involved in vesicular trafficking in the early secretory pathway (see S1 Supplementary  
552 Tables, S9 Table). A *gbfl* expression was shown throughout spermatocyte development in the  
553 Golgi apparatus with expression extending into early spermatids where it localizes also to  
554 acrosome [58]. The positional candidate gene *hells* (helicase lymphoid specific) in Omy1\_b2  
555 is a very convincing functional candidate gene for sex-reversal. *hells* gene (also known as *lsh*)  
556 is a member of the *snf2* helicase family that is involved in cellular proliferation, and required  
557 for normal development and survival in the mouse [59]. *snf2*-like helicases play a role in  
558 chromatin remodeling, and the *hells* protein has been implicated in the control of genome-wide  
559 DNA methylation [60]. Analysis of ovarian explants obtained from *hells* mutant females  
560 demonstrates that lack of *hells* function is associated with severe oocyte loss and lack of ovarian  
561 follicle formation [61]. Although *hells* is expressed in undifferentiated embryonic stem cells, it  
562 is significantly downregulated upon differentiation [62]. Because of the neonatal lethality of  
563 the *hells* *-/-* mice, Zeng et al. [63] used allografting of testis tissue from *hells* *-/-* mice to study  
564 postnatal male germ cell differentiation. They showed that proliferation of spermatogonia was  
565 reduced in the absence of *hells*, and germ cell differentiation arrested at the midpachytene stage,  
566 implicating an essential role for HELLS during male meiosis [63].

567 As regards to Omy1\_b1, there is no direct evidence for a potential role of *myoz1* gene in sex-  
568 reversal (see S1 Supplementary Tables, S9 Table). The second gene pointed out for Omy1\_b1,  
569 *tlx1* (also known as *hox11*), encodes a nuclear transcription factor that belongs to the NK-linked  
570 or NK-like (*nkl*) subfamily of homeobox genes. The encoded protein is essential for cell  
571 survival during spleen development in fish [64] as in mammals [65, 66], and is also involved  
572 in specification of neuronal cell fates during embryogenesis (see S1 Supplementary Tables, S9  
573 Table). At the molecular level, depending on the cellular context and its interaction with

574 transcriptional cofactors, *tlx1* can either activate or repress gene transcription [67]. *tlx1*-  
 575 dependent regulation of retinoic acid (RA) metabolism is critical for spleen organogenesis [68].  
 576 RA is the active metabolite of vitamin A that is required for vertebrate embryogenesis [69, 70].  
 577 In mice, the transcription factor steroidogenic factor 1 (*nr5a1* also known as SF1) regulates RA  
 578 metabolism during germ cell development [71]. *nr5a1* is a transcriptional activator that is  
 579 essential for sexual differentiation and formation of the primary steroidogenic tissues.  
 580 Mutations within *nr5a1* underlie different disorders of sexual development, including sex  
 581 reversal. In patients with disorders of sexual development, a mutant form of *nr5a1* was shown  
 582 to be defective in activating *tlx1* transcription [72]. In consequence, we can hypothesize that  
 583 genetic variants in *tlx1* may also directly deregulate RA metabolism, inducing sex-reversal.

584 Gathering all the information presented here above for the QTL on *Omy1*, we can conclude that  
 585 there is a strong statistical and biological evidence of significant involvement in the sex  
 586 determination and differentiation cascade of *syndig1*, *tlx1*, *hells* and *gbf1*, and, putatively also  
 587 of *vsx1*, and *myoz1*. Notably, at the only exception of *myoz1*, all these proteins are relatively  
 588 well conserved in terms of protein identity across various fishes and mammals (Table 5).

589

590 **Table 5** Protein sequence identities of genes associated to sex-reversal in rainbow trout across  
 591 six species in Vertebrates

592

<b>rainbow trout / target (% id AA)</b>	<b>Atlantic salmon</b>	<b>Medaka</b>	<b>Zebrafish</b>	<b>Mouse</b>	<b>Sheep</b>	<b>Human</b>
<i>syndig1</i>	98%	58%	69%	65%	67%	59%
<i>vsx1</i>	96%	69%	84%	62%	47%	52%
<i>fbxw4</i>	99%	68%	71%	58%	58%	58%
<i>hells</i>	98%	79%	79%	65%	64%	64%
<i>myoz1a</i>	92%	56%	63%	39%	37%	40%
<i>tlx1</i>	95%	75%	82%	61%	58%	61%
<i>gbf1</i>	99%	82%	83%	72%	72%	72%

<i>hcn1</i>	99%	84%	84%	74%	68%	68%
<i>arfgef3</i>	99%	76%	84%	71%	72%	72%
<i>cep170aa</i>	98%	63%	68%	53%	54%	55%
<i>abcg5</i>	100%	91%	94%	87%	90%	88%
<i>khdrbs2</i>	98%	76%	82%	70%	69%	69%
<i>dst</i>	96%	72%	71%	60%	59%	62%
<i>cilk1</i>	99%	70%	NA	32%	55%	55%
<i>csmd1</i>	99%	79%	89%	76%	76%	77%
<i>lrrc59</i>	98%	71%	62%	58%	58%	58%
<i>caskin2</i>	91%	54%	54%	59%	59%	59%

593

594 ***arfgef3* as the best functional candidate for Omy20\_a due to its activation of the**  
595 **estrogen/ER $\alpha$  signaling pathway**

596 *arfgef3* gene (named also *big3*) is a member of the *big1/sec7p* subfamily of ADP ribosylation  
597 factor-GTP exchange factors (ARF-GEFs). The protein is very well conserved across  
598 vertebrates (Table 5). *arfgef3* overexpression has been pointed out as one of the important  
599 mechanisms causing the activation of the estrogen/ER $\alpha$  signaling pathway in the hormone-  
600 related growth of breast cancer cells [73]. As summed up in the introduction, the germ cells  
601 seem to play a critical role in sex determination, especially in the feminization of gonads. PGCs  
602 may first interpret internal genetic or external environmental cues and directly transform into  
603 oogonia; then, the surrounding somatic cells may be induced by the germ cells to differentiate  
604 accordingly to provide an appropriate hormonal environment for further gonadal differentiation  
605 [11]. In support of this hypothesis, high temperatures were reported to decrease the number of  
606 germ cells, which often accompanies female-to-male sex reversal [17]. Guiguen et al. [12]  
607 suggested that over-expression of germ cell proliferation or anti-apoptotic germ cell factors in  
608 the female gonads could promote ovarian differentiation.

609

610 ***khdrbs2*, *dst*, *hmgcll1* and *csmd1* as functional candidates for QTL Omy20\_b**

611 ***Implication of khdrbs genes in sex determination and regulation of alternative splicing***

612 In nematodes, *gld-1*, the ortholog of *khdrbs* (KH domain-containing, RNA-binding, signal  
613 transduction-associated protein 2), is localized in the germ cell cytoplasm. This gene is  
614 indispensable for oogenesis and meiotic prophase progression and it can stimulate sex  
615 determination of males in the hermaphrodite germ line [74, 75, 76]. In fruit fly however, *nsr*,  
616 the ortholog of *khdrbs* regulates some male fertility genes in the primary [77].

617 In vertebrates, *khdrbs* (*khdrbs1*, *khdrbs2* and *khdrbs3*) genes encode respectively for the  
618 SAM68, SLM1 and SLM2 proteins that have RNA binding and signal transduction activities  
619 [78, 79]. In fish, the *khdrbs1* was further duplicated through the teleost-specific whole genome  
620 duplication, forming *khdrbs1a* and *khdrbs1b*. In zebrafish, all the *khdrbs* genes were found to  
621 be primarily expressed in the brain during early development and at the adult stage. *khdrbs1a*  
622 was also found to be expressed in the gonad primordium as well as abundantly expressed in the  
623 gonads of adult fish along with *khdrbs1b*, and *khdrbs3* [80]. In our study, eight intronic *khdrbs2*  
624 variants were ranked among the 100 best SNPs in DAPC (see S1 Supplementary Tables, S10  
625 Table), and an indel/SNP variant was identified with a p-value < 0.002 in the discovery  
626 population (Table 4). In mouse, *khdrbs1* and *khdrbs3* are predominantly expressed in the brain  
627 and testis, whether *khdrbs2* is exclusively expressed in the brain [81]. *khdrbs2* knockout mice  
628 are fertile fertile whereas *khdrbs1* knockout mice show impaired fertility in males [81]. *khdrbs3*  
629 is involved in spermatogenesis via directly binding to RNA [81]. In addition, *khdrbs1 knockout*  
630 females displayed a reduction in the number of developing ovarian follicles, alteration of  
631 estrous cycles, and impaired fertility (e.g. [82]). It seems that the invertebrate genes *gld-1* and  
632 *nsr* that are orthologs of *khdrbs1* and *khdrbs3*, respectively, have similar roles in the regulation  
633 of gametogenesis [80].

634

### 635 ***dst* and *hmgcll1* as novel players in sex determination and differentiation in fish?**

636 *dst* (dystonin) encodes a member of the plakin protein family of adhesion junction plaque  
637 proteins. *dst* isoforms are expressed in neural tissue where it regulates the organization and  
638 stability of the microtubule network of sensory neurons to allow axonal transport. Low  
639 temperatures significantly impact *dst* splicing in killifish, stickleback and zebrafish [83]. Gene  
640 ontology annotations related to this gene include calcium ion binding and actin binding (see S1  
641 Supplementary Tables, S9 Table). As far as we know, *dst* has not yet been demonstrated as  
642 being involved in sex determination or differentiation. Nevertheless, there is no statistical doubt  
643 that this gene is involved in sex-reversal in rainbow trout. A SNP was highly significant (p-value  
644 = 0.0006) in the discovery population (Table 4) with homozygous TT dams exhibiting high

645 ratios of progeny sex-reversal (see S1 Supplementary Tables, S10 Table). In addition, 4  
646 haplotypes within the gene ranked among the 60 best in the full RF analysis (see S1  
647 Supplementary Tables, S8 Table), and 15 intronic SNPs were ranked among the 100 best in  
648 DAPC, 10 remaining in the 50 best SNPs (see S1 Supplementary Tables, S10 Table).

649 *hmgcll1* (3-hydroxymethyl-3-methylglutaryl-CoA lyase-like 1) was first characterized as a  
650 lyase activity enzyme localized in extra-mitochondrial region involved in ketogenesis for  
651 energy production in nonhepatic animal tissues (see S1 Supplementary Tables, S9 Table). A  
652 novel role of *hmgcll1* in cell cycle regulation was recently suggested [84] that could be linked  
653 with a putative functional role on sex-reversal. Statistically, there is no doubt about this effect.  
654 Two SNPs within the gene were significant in two validation population (Table 3). In the  
655 discovery population, an upstream intergenic variant at position 34,345,558 bp was among the  
656 50 best ranked in DAPC, with homozygous TT dams exhibited high ratios of progeny sex-  
657 reversal (see S1 Supplementary Tables, S10 Table).

658

#### 659 ***csmdl* is associated to gonadal failure in mammals**

660 The *csmdl* (CUB and sushi multiple domains protein 1) gene was one of the best positional  
661 candidates in our RF and DAPC analyses with 8 SNPs being among the 50 best ones to  
662 discriminate the two groups of extreme phenotypes. This gene encodes for a large (>3,000  
663 amino acids) transmembrane protein. The family of proteins *csmd* (CUB and sushi multiple  
664 domains) is one of the three types of the complement system-related proteins involved in the  
665 recognition of molecules in innate immune-system and in the central nervous system (reviewed  
666 in [85]). *csmdl* is predicted to act upstream of, or within, several processes, including learning  
667 or memory, and reproductive structure development. The protein encoded by *csmdl* is highly  
668 conserved between vertebrates with 89% amino acid sequence identity between trout and  
669 medaka and still 77% between trout and human or mouse (Table 5). In humans and mice, *csmdl*  
670 was associated with gonadal failure in both sexes [86]. Significant association between early  
671 menopause and rare 5'-deletions of *csmdl* was detected in humans. In mice, knockout females  
672 show significant reduction in ovarian quality and breeding success, and *csmdl*-knockout males  
673 show increased rates of infertility and severe histological degeneration of the testes. We did not  
674 observe any variants predicted with moderate or high impacts on protein expression in this gene.  
675 But, as indicated by Lee et al. [86], intronic variants mainly located in introns 1 and 2 likely  
676 harbored functional elements that can influence *csmdl* expression in testis.

677

678 ***lrrc59* and *caskin2* as the best positional candidates for QTL Omy20\_c.**

679 As far as we know, no direct effect of *lrrc59* (leucine-rich repeat-containing protein 59) on sex  
680 determination is known. This gene enables RNA binding activity and cadherin binding activity  
681 (see S1 Supplementary Tables, S9 Table). A key role for *lrrc59* in the organization and  
682 regulation of local translation in the endoplasmic reticulum has been suggested [87].  
683 Statistically, it was the best-ranked haplotype on the full RF analysis (see S1 Supplementary  
684 Tables, S8 Table), indicating a strong importance of *lrrc59* to predict sex-reversal.

685 *caskin2* (CASK Interacting Protein 2) encodes a large protein that may be involved in protein-  
686 protein interactions (see S1 Supplementary Tables, S9 Table). *caskin2* downregulates genes  
687 associated with endothelial cell activation and upregulates genes associated with endothelial  
688 cell quiescence [88]. Morpholino knockdown of *caskin2* in zebrafish results in abnormal  
689 vascular development characterized by overly branched intersegmental vessels and failure to  
690 form the dorsal longitudinal vessels, while *caskin2* knockout mice are viable and fertile [88].  
691 *caskin2* is not well conserved across vertebrates, including only 54% AA identity between  
692 rainbow trout and either Medaka or Zebrafish (Table 5). Statistically, two haplotypes including  
693 parts of *caskin2* were among the best ranked haplotypes in the full RF analysis (see S1  
694 Supplementary Tables, S8 Table). In addition, 3 SNPs (2 intronic and an upstream intergenic  
695 variant) were among the 50 best ranked in DAPC (see S1 Supplementary Tables, S10 Table).  
696 The expression of *caskin2* was down-regulated in gonads of *Gobiocypris rarus* of both sexes,  
697 when exposed to a 7-day exposure to 17 $\alpha$ -methyltestosterone which had a sex-reversal effect  
698 [89]. In Coho salmon, androgens were revealed to play a major role in stimulating primary  
699 ovarian follicle development and the transition into secondary follicles [90].

700

701 **Putative role of *foxl3* in germ cell fate decision**

702 Intriguingly, *foxl3*, a co-paralog gene of *foxl2*, is located on Omy20 spanning from 22,942,515  
703 to 22,948,650 bp, upstream the QTL regions we studied on the same chromosome. As *foxl3* is  
704 expressed in the gonads of teleosts, we checked if any variant within *foxl3* could be associated  
705 to sex-reversal in our study, and could not find any. However, we can speculate that *foxl3* is  
706 involved in germ cell fate decision in rainbow trout. Indeed, *foxl3* was already suggested to be  
707 involved in the onset of oocyte meiosis, and/or in the regulation of male specific genes during  
708 late testis development and testis maturation in salmonids [91, 92, 93]. It was also demonstrated

709 in medaka that *foxl3* expression is only maintained in female germ cells, and that functional  
710 sperm is produced in the ovary of a *foxl3* loss-of-function mutant [94]. In addition, *foxl3* was  
711 shown to initiate oogenesis in medaka through two genetically independent pathways, meiosis  
712 and follicular development, and that a third pathway independent of *foxl3* may also be  
713 hypothesized [95]. Last, *foxl3* expression levels were strongly increased in gonads during the  
714 natural female-to-male sex reversal in the rice field eel (*Monopterus albus*). *foxl3* and miR-9  
715 seemed to be involved in physiological processes that promote oocyte degeneration in the  
716 ovotestis and stimulating spermatogenesis in spermatogonia [96]. Therefore, we can  
717 hypothesize that expression of *foxl3* may be regulated by some other genes, presumably located  
718 in the QTL regions identified on Omy20 in our study, similarly to the long-distance cis-  
719 regulation observed for *sox9* expression [57]. Further investigation should try to validate such  
720 speculations, as well as the causative variants and precise functional roles of the identified genes  
721 in our QTL regions.

722

## 723 **Conclusions**

724 While considerable scientific effort has been made over the last fifty years to unravel  
725 mechanisms for sex-determination and differentiation, the nature of genetic variations that  
726 regulate these mechanisms in fish is still not well-known. We confirmed on several French  
727 commercial rainbow trout populations, the importance of some genomic regions from  
728 chromosomes Omy1, Omy12 and Omy20 as minor sex-determining genes involved in maleness  
729 in the absence of the master sex-determining gene *sdY* in rainbow trout. Together, our results are  
730 consistent with a model in which spontaneous female-to-male sex-reversal in rainbow trout is  
731 associated with genetic factors able to reduce germ cell proliferation and arrest oogenesis. The  
732 main candidate genes that we suggested based on positional and functional information are  
733 *syndig1*, *tlx1*, *hells* and *gbf1* on Omy1, and *arfgef3*, *khdrbs2*, *dst*, *hmgcll1*, *csmd1*, *lrrc59* and  
734 *caskin2* on Omy20. Most of these genes were previously indicated as playing roles in  
735 developmental pathways in vertebrates.

736 From a fish farmer perspective, further investigation is still needed before any operational use  
737 of the genotypes at minor sex-modifier genes either to eradicate spontaneous sex-reversal in  
738 all-female production stocks or to produce progeny from free-hormone neomales in rainbow  
739 trout. Indeed, the high heritability of maleness [27] suggests that the use of spontaneously  
740 masculinised individuals as progenitors of all-female stocks would increase the frequency of

741 undesirable masculinised progeny. Improve knowledge about the joint effects of genetic basis  
742 and environmental factors determining spontaneous maleness in all-female stocks is then  
743 needed to search for a trade-off combining together genetic and environmental control of gonad  
744 masculinisation according to the destination of the fish (broodstock vs all-female production  
745 stock). High temperature during development masculinizes many fish species [97]. In rainbow  
746 trout, high rearing temperature during the early phase of fry development after hatching was  
747 shown to increase the frequency of sex reversal in XX females to a greater or lesser extent  
748 depending on the genetic background of the trout lines [98, 99]. Further investigation should  
749 therefore focus on the interactions between genetics and rearing temperatures to determine the  
750 key combination of factors to produce all-female progeny from free-hormone neomales in  
751 rainbow trout.

752

### 753 **Supplementary Information**

754

#### 755 **S1 Appendix (S1-Appendix.doc)**

756 **Positional candidate genes on Omy12 and Omy20 without known evidence of a functional**  
757 **role linked to sex-reversal**

758

759 **S1 Supplementary Tables (S1-Supplementary Tables.xls). This excel file includes 10**  
760 **supplementary tables whose titles are listed below.**

761 **S1 Table Number of progeny (female, intersex and male) and proportion of neomales per**  
762 **sequenced dam**

763 **S2 Table Positions on the Swanson reference genome and flanking sequences of the 192**  
764 **SNPs used in validation**

765 **S3 Table List of the 192 SNPs detected in 6 validation populations (named A to F) with**  
766 **their positions on Swanson (SW) and Arlee (AR) reference genomes**

767 **S.4 Table List of most significant SNPs found in Omy1, Omy12 and Omy20 QTL**

768 **S5 Table Mean, Minimum and Maximum values of the 30 haplotypes with the highest**  
769 **Gini index over 100 runs of Random Forests for Omy1 QTL**

770 **S6 Table Mean, Minimum and Maximum values of the 10 haplotypes with the highest**  
771 **Gini index over 100 runs of Random Forests for Omy12 QTL**

772 **S7 Table Mean, Minimum and Maximum values of the 30 haplotypes with the highest**  
773 **Gini index over 100 runs of Random Forests for the 3 Omy 20 QTL (Z3, Z4, Z5)**

774 **S8 Table Mean values of the 60 haplotypes with the highest Gini index over 100 runs of**  
775 **Random Forests for the QTL regions of Omy1 (Z1), Omy12 (Z2) and Omy20 (Z3 to Z5)**

776 **S9 Table List of potential candidate genes for QTL on Omy1, Omy12 and Omy20 with**  
777 **their descriptions on GeneCards, UniProtKB/Swiss-Prot and NCBI databases.**

778 **S10 Table Location, annotation, alleles and genotypes for sequenced dams of the 100 most**  
779 **important SNPs in DAPC analysis. Genotypes in blue are considered as male genotypes,**  
780 **genotypes in orange as female genotypes and genotypes in light orange are heterozygous.**

781

782

783 **Acknowledgements** The European Maritime and Fisheries Fund and the French National  
784 Government supported this work (R FEA470016FA1000008). The authors are grateful to the  
785 companies “Les Fils de Charles Murgat”, “Aqualande” “Bretagne Truite”, “Font Rome” and  
786 “Viviers de Sarrance” for fish samples and data collection.

787

788 **Author contributions** AD: Data curation, Formal analysis, Preparation of tables and figures,  
789 Writing—review and editing. CF: Data curation, Formal analysis, Preparation of figures,  
790 Writing—review and editing. AB: Data collection, Writing—review and editing. CP: Data  
791 collection. YG: Data collection, Interpretation, Writing—review and editing. EQ:  
792 Conceptualization, Data collection, Funding acquisition, Project administration, Writing—  
793 review and editing. FP: Conceptualization, Methodology, Supervision, Statistical analysis,  
794 Interpretation, Project administration, Writing—original draft, review, and editing. All authors  
795 read and approved the final version of the manuscript.

796

797 **Data availability** Raw sequence data that were used in this study are deposited in the ENA  
798 (Project PRJEB75960 in <https://www.ebi.ac.uk/ena/browser/view/PRJEB75960>).

799

800 **Declarations**

801 **Animal and human rights statement** We declare that all applicable international, national,  
802 and/or institutional guidelines for sampling, care, and experimental use of organisms for the  
803 study have been followed and all necessary approvals have been obtained. As part of the  
804 standard breeding practices for a commercial breeding program, the handling of fish was not  
805 subject to oversight by an institutional ethic committee. Animals were treated in compliance  
806 with the European Communities Council Directive (9858/CEC) for the care and use of farm  
807 animals.

808

809 **References**

- 810 1. Baroiller JF, D'Cotta H. The Reversible Sex of Gonochoristic Fish: Insights and  
811 Consequences. *Sex Dev.* 2016;10(5-6):242-266. doi: 10.1159/000452362.
- 812 2. Pandian TJ, Sheela SG. Hormonal induction of sex reversal in fish. *Aquaculture.*  
813 1995;138:1–22.
- 814 3. Piferrer F. Endocrine sex control strategies for the feminization of teleost fish.  
815 *Aquaculture.* 2001;197:229–281.
- 816 4. Azuma T, Takeda K, Doi T, Muto K, Akutsu M, Sawada M, et al. The influence of  
817 temperature on sex determination in sockeye salmon *Oncorhynchus nerka*. *Aquaculture.*  
818 2004;234:461-473. doi: 10.1016/j.aquaculture.2003.11.023.
- 819 5. Robles V, Riesco MF, Psenicka M, Saito T, Valcarce DG, Cabrita E, et al. Biology of  
820 teleost primordial germ cells (PGCs) and spermatogonia: biotechnological applications.  
821 *Aquaculture.* 2017;472:4-20.
- 822 6. Weber C, Capel B. Sex reversal. *Curr Biol.* 2018 Nov 5;28(21):R1234-R1236. doi:  
823 10.1016/j.cub.2018.09.043.
- 824 7. Swain A, Lovell-Badge R. A molecular approach to sex determination in mammals.  
825 *Acta Paediatr Suppl.* 1997 Nov;423:46-9. doi: 10.1111/j.1651-2227.1997.tb18368.x.
- 826 8. Nagahama Y, Chakraborty T, Paul-Prasanth B, Ohta K, Nakamura M. Sex  
827 determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol*  
828 *Rev.* 2021 Jul 1;101(3):1237-1308. doi: 10.1152/physrev.00044.2019.
- 829 9. Herpin A, Schartl M. Plasticity of gene-regulatory networks controlling sex  
830 determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO*  
831 *Rep.* 2015 Oct;16(10):1260-74. doi: 10.15252/embr.201540667.

- 832 10. Lin YT, Capel B. Cell fate commitment during mammalian sex determination. *Curr*  
833 *Opin Genet Dev.* 2015 Jun;32:144-52. doi: 10.1016/j.gde.2015.03.003.
- 834 11. Devlin RH, Nagahama Y. Sex determination and sex differentiation in fish: an overview  
835 of genetic, physiological, and environmental influences. *Aquaculture.* 2002;208:191-  
836 364. doi: 10.1016/S0044-8486(02)00057-1.
- 837 12. Guiguen Y, Fostier A, Piferrer F, Chang CF. Ovarian aromatase and estrogens: a pivotal  
838 role for gonadal sex differentiation and sex change in fish. *Gen Comp Endocrinol.* 2010  
839 Feb 1;165(3):352-66. doi: 10.1016/j.ygcen.2009.03.002.
- 840 13. Wargelius A, Leininger S, Skaftnesmo KO, Kleppe L, Andersson E, Taranger GL, et al.  
841 Dnd knockout ablates germ cells and demonstrates germ cell independent sex  
842 differentiation in Atlantic salmon. *Sci Rep.* 2016 Feb 18;6:21284. doi:  
843 10.1038/srep21284.
- 844 14. Slanchev K, Stebler J, de la Cueva-Méndez G, Raz E. Development without germ cells:  
845 the role of the germ line in zebrafish sex differentiation. *Proc Natl Acad Sci U S A.* 2005  
846 Mar 15;102(11):4074-9. doi: 10.1073/pnas.0407475102.
- 847 15. Siegfried KR, Nüsslein-Volhard C. Germ line control of female sex determination in  
848 zebrafish. *Dev Biol.* 2008 Dec 15;324(2):277-87. doi: 10.1016/j.ydbio.2008.09.025.
- 849 16. Kurokawa H, Saito D, Nakamura S, Katoh-Fukui Y, Ohta K, Baba T, et al. Germ cells  
850 are essential for sexual dimorphism in the medaka gonad. *Proc Natl Acad Sci U S A.*  
851 2007 Oct 23;104(43):16958-63. doi: 10.1073/pnas.0609932104.
- 852 17. Nishimura T, Yamada K, Fujimori C, Kikuchi M, Kawasaki T, Siegfried KR, et al.  
853 Germ cells in the teleost fish medaka have an inherent feminizing effect. *PLoS Genet.*  
854 2018 Mar 29;14(3):e1007259. doi: 10.1371/journal.pgen.1007259.
- 855 18. Herpin A, Schindler D, Kraiss A, Hornung U, Winkler C, Schartl M. Inhibition of  
856 primordial germ cell proliferation by the medaka male determining gene *Dmrt 1* bY.  
857 *BMC Dev Biol.* 2007 Aug 30;7:99. doi: 10.1186/1471-213X-7-99.
- 858 19. Kobayashi T, Matsuda M, Kajiura-Kobayashi H, Suzuki A, Saito N, Nakamoto M, et  
859 al. Two DM domain genes, *DMY* and *DMRT1*, involved in testicular differentiation  
860 and development in the medaka, *Oryzias latipes*. *Dev Dyn.* 2004 Nov;231(3):518-26.  
861 doi: 10.1002/dvdy.20158.
- 862 20. Pan Q, Kay T, Depincé A, Adolphi M, Schartl M, Guiguen Y, Herpin A. Evolution of  
863 master sex determiners: TGF- $\beta$  signalling pathways at regulatory crossroads. *Philos*  
864 *Trans R Soc Lond B Biol Sci.* 2021 Aug 30;376(1832):20200091. doi:  
865 10.1098/rstb.2020.0091.

- 866 21. Chen J, Zhu Z, Hu W. Progress in research on fish sex determining genes. *Water*  
867 *Biology and Security* 1. 2022;100008. <https://doi.org/10.1016/j.watbs.2022.100008>
- 868 22. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, et al. The sexually  
869 dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-  
870 chromosome sequence in many salmonids. *Evol Appl*. 2013 Apr;6(3):486-96. doi:  
871 10.1111/eva.12032.
- 872 23. Bertho S, Herpin A, Branthonne A, Jouanno E, Yano A, Nicol B, et al. The unusual  
873 rainbow trout sex determination gene hijacked the canonical vertebrate gonadal  
874 differentiation pathway. *Proc Natl Acad Sci U S A*. 2018 Dec 11;115(50):12781-12786.  
875 doi: 10.1073/pnas.1803826115.
- 876 24. Bertho S, Pasquier J, Pan Q, Le Trionnaire G, Bobe J, Postlethwait JH, et al. Foxl2 and  
877 Its Relatives Are Evolutionary Conserved Players in Gonadal Sex Differentiation. *Sex*  
878 *Dev*. 2016;10(3):111-29. doi: 10.1159/000447611.
- 879 25. Pailhoux E, Vigier B, Chaffaux S, Servel N, Taourit S, Furet JP, et al. A 11.7-kb deletion  
880 triggers intersexuality and polledness in goats. *Nat Genet*. 2001 Dec;29(4):453-8. doi:  
881 10.1038/ng769.
- 882 26. Pannetier M, Elzaïat M, Thépot D, Pailhoux E. Telling the story of XX sex reversal in  
883 the goat: highlighting the sex-crossroad in domestic mammals. *Sex Dev*. 2012;6(1-  
884 3):33-45. doi: 10.1159/000334056.
- 885 27. Fraslin C, Phocas F, Bestin A, Charles M, Bernard M, Krieg F, et al. Genetic  
886 determinism of spontaneous masculinisation in XX female rainbow trout: new insights  
887 using medium throughput genotyping and whole-genome sequencing. *Sci Rep*. 2020  
888 Oct 19;10(1):17693. doi: 10.1038/s41598-020-74757-8.
- 889 28. Guyomard R, Guiguen Y, Bernard M, Charlet A, Dechamp N, Hervet C, et al. RAD-  
890 seq mapping of spontaneous masculinization in XX doubled haploid rainbow trout lines.  
891 Conference, 10th WCGALP, Vancouver, BC, Canada. 2014.
- 892 29. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, et al. Sex-  
893 dependent dominance maintains migration supergene in rainbow trout. *Nat Ecol Evol*.  
894 2019 Dec;3(12):1731-1742. doi: 10.1038/s41559-019-1044-6. Erratum in: *Nat Ecol*  
895 *Evol*. 2020 Jan;4(1):170. doi: 10.1038/s41559-019-1076-y.
- 896 30. Gao G, Magadan S, Waldbieser GC, Youngblood RC, Wheeler PA, Scheffler BE, et al.  
897 A long reads-based de-novo assembly of the genome of the Arlee homozygous line  
898 reveals chromosomal rearrangements in rainbow trout. *G3 (Bethesda)*. 2021 Apr  
899 15;11(4):jkab052. doi: 10.1093/g3journal/jkab052.

- 900 31. Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E, et al. A resource of  
901 single-nucleotide polymorphisms for rainbow trout generated by restriction-site  
902 associated DNA sequencing of doubled haploids. *Mol Ecol Resour.* 2014  
903 May;14(3):588-96. doi: 10.1111/1755-0998.12204.
- 904 32. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera  
905 GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples.  
906 *bioRxiv.* 2017;201178. doi: 10.1101/201178.
- 907 33. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing.  
908 2012. doi: 10.48550/arXiv.1207.3907.
- 909 34. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years  
910 of SAMtools and BCFtools. *Gigascience.* 2021 Feb 16;10(2):giab008. doi:  
911 10.1093/gigascience/giab008.
- 912 35. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A  
913 framework for variation discovery and genotyping using next-generation DNA  
914 sequencing data. *Nat Genet.* 2011 May;43(5):491-8. doi: 10.1038/ng.806.
- 915 36. R Development Core Team. R: A Language and Environment for Statistical Computing.  
916 Conference, R Foundation for Statistical Computing, Vienna, Austria. 2023
- 917 37. Mehta CR, Patel NR. A network algorithm for performing Fisher's exact test in  $r \times c$   
918 contingency tables. *J Am Stat Assoc.* 1983;78:427–434
- 919 38. Mehta CR, Patel NR. Algorithm 643: FEXACT, a FORTRAN subroutine for Fisher's  
920 exact test on unordered  $r \times c$  contingency tables. *ACM Trans Math Softw.* 1986;12,  
921 154–161
- 922 39. Clarkson D B, Fan Y, Joe H. A remark on algorithm 643: FEXACT: an algorithm for  
923 performing Fisher's exact test in  $r \times c$  contingency tables. *ACM Trans. Math. Softw.*  
924 1993;19,4:484–488. doi: 10.1145/168173.168412.
- 925 40. Bernard M, Dehaullon A, Gao G, Paul K, Lagarde H, Charles M, et al. Development of  
926 a High-Density 665 K SNP Array for Rainbow Trout Genome-Wide Genotyping. *Front*  
927 *Genet.* 2022 Jul 18;13:941340. doi: 10.3389/fgene.2022.941340.
- 928 41. Breiman L. Random Forests. *Machine Learning.* 2001;45:5–32. doi:  
929 10.1023/A:1010933404324.
- 930 42. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics.* 2012  
931 Jun;99(6):323-9. doi: 10.1016/j.ygeno.2012.04.003.

- 932 43. Liaw A, Wiener M. Classification and Regression by randomForest. 2002 [cited 6 Feb  
933 2024]. In: R News, 2(3), 18-22. Available from: [https://CRAN.R-](https://CRAN.R-project.org/doc/Rnews/)  
934 [project.org/doc/Rnews/](https://CRAN.R-project.org/doc/Rnews/).
- 935 44. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a  
936 new method for the analysis of genetically structured populations. BMC Genet. 2010  
937 Oct 15;11:94. doi: 10.1186/1471-2156-11-94.
- 938 45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar  
939 P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association  
940 and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75. doi:  
941 10.1086/519795.
- 942 46. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers.  
943 Bioinformatics. 2008 Jun 1;24(11):1403-5. doi: 10.1093/bioinformatics/btn129.
- 944 47. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for  
945 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs  
946 in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012  
947 Apr-Jun;6(2):80-92. doi: 10.4161/fly.19695.
- 948 48. Hayashi T, Huang J, Deeb SS. RINX(VSX1), a novel homeobox gene expressed in the  
949 inner nuclear layer of the adult retina. Genomics. 2000 Jul 15;67(2):128-39. doi:  
950 10.1006/geno.2000.6248.
- 951 49. Ohtoshi A, Justice MJ, Behringer RR. Isolation and characterization of *Vsx1*, a novel  
952 mouse CVC paired-like homeobox gene expressed during embryogenesis and in the  
953 retina. Biochem Biophys Res Commun. 2001 Aug 10;286(1):133-40. doi:  
954 10.1006/bbrc.2001.5372.
- 955 50. Clark AM, Yun S, Veien ES, Wu YY, Chow RL, Dorsky RI, et al. Negative regulation  
956 of *Vsx1* by its paralog *Chx10/Vsx2* is conserved in the vertebrate retina. Brain Res.  
957 2008 Feb 4;1192:99-113. doi: 10.1016/j.brainres.2007.06.007.
- 958 51. Kalashnikova E, Lorca RA, Kaur I, Barisone GA, Li B, Ishimaru T, et al. SynDIG1: an  
959 activity-regulated, AMPA- receptor-interacting transmembrane protein that regulates  
960 excitatory synapse development. Neuron. 2010 Jan 14;65(1):80-93. doi:  
961 10.1016/j.neuron.2009.12.021.
- 962 52. Lovero KL, Blankenship SM, Shi Y, Nicoll RA. SynDIG1 promotes excitatory  
963 synaptogenesis independent of AMPA receptor trafficking and biophysical regulation.  
964 PLoS One. 2013 Jun 13;8(6):e66171. doi: 10.1371/journal.pone.0066171.

- 965 53. Chenux G, Matt L, Hill TC, Kaur I, Liu XB, Kirk LM, et al. Loss of SynDIG1 Reduces  
966 Excitatory Synapse Maturation But Not Formation In Vivo. *eNeuro*. 2016 Oct  
967 21;3(5):ENEURO.0130-16.2016. doi: 10.1523/ENEURO.0130-16.2016.
- 968 54. Bonnet A, Servin B, Mulsant P, Mandon-Pepin B. Spatio-Temporal Gene Expression  
969 Profiling during In Vivo Early Ovarian Folliculogenesis: Integrated Transcriptomic  
970 Study and Molecular Signature of Early Follicular Growth. *PLoS One*. 2015 Nov  
971 5;10(11):e0141482. doi: 10.1371/journal.pone.0141482.
- 972 55. Ruane NM, Lambert JG, Goos HJ, Komen J. Hypocorticism and interrenal hyperplasia  
973 are not directly related to masculinization in XX mas(-1)/mas(-1) carp, *Cyprinus carpio*.  
974 *Gen Comp Endocrinol*. 2005 Aug;143(1):66-74. doi: 10.1016/j.ygcen.2005.02.024.
- 975 56. Zhai G, Shu T, Xia Y, Lu Y, Shang G, Jin X, et al. Characterization of Sexual Trait  
976 Development in cyp17a1-Deficient Zebrafish. *Endocrinology*. 2018 Oct  
977 1;159(10):3549-3562. doi: 10.1210/en.2018-00551.
- 978 57. Terao M, Ogawa Y, Takada S, Kajitani R, Okuno M, Mochimaru Y, et al. Turnover of  
979 mammal sex chromosomes in the Sry-deficient Amami spiny rat is due to male-specific  
980 upregulation of Sox9. *Proc Natl Acad Sci U S A*. 2022 Dec 6;119(49):e2211574119.  
981 doi: 10.1073/pnas.2211574119.
- 982 58. Au CE, Hermo L, Byrne E, Smirle J, Fazel A, Simon PH, et al. Expression, sorting, and  
983 segregation of Golgi proteins during germ cell differentiation in the testis. *Mol Biol  
984 Cell*. 2015 Nov 5;26(22):4015-32. doi: 10.1091/mbc.E14-12-1632.
- 985 59. Geiman TM, Tessarollo L, Anver MR, Kopp JB, Ward JM, Muegge K. Lsh, a SNF2  
986 family member, is required for normal murine development. *Biochim Biophys Acta*.  
987 2001 May 3;1526(2):211-20. doi: 10.1016/s0304-4165(01)00129-5.
- 988 60. Dennis K, Fan T, Geiman T, Yan Q, Muegge K. Lsh, a member of the SNF2 family, is  
989 required for genome-wide methylation. *Genes Dev*. 2001 Nov 15;15(22):2940-4. doi:  
990 10.1101/gad.929101.
- 991 61. De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, Muegge K. Lsh is  
992 required for meiotic chromosome synapsis and retrotransposon silencing in female germ  
993 cells. *Nat Cell Biol*. 2006 Dec;8(12):1448-54. doi: 10.1038/ncb1513.
- 994 62. Xi S, Geiman TM, Briones V, Guang Tao Y, Xu H, Muegge K. Lsh participates in DNA  
995 methylation and silencing of stem cell genes. *Stem Cells*. 2009 Nov;27(11):2691-702.  
996 doi: 10.1002/stem.183.
- 997 63. Zeng W, Baumann C, Schmidtmann A, Honaramooz A, Tang L, Bondareva A, et al.  
998 Lymphoid-specific helicase (HELLS) is essential for meiotic progression in mouse

999 spermatocytes. Biol Reprod. 2011 Jun;84(6):1235-41. doi:  
1000 10.1095/biolreprod.110.085720.

1001 64. Xie L, Tao Y, Wu R, Ye Q, Xu H, Li Y. Congenital asplenia due to a *tlx1* mutation  
1002 reduces resistance to *Aeromonas hydrophila* infection in zebrafish. Fish Shellfish  
1003 Immunol. 2019 Dec;95:538-545. doi: 10.1016/j.fsi.2019.10.065.

1004 65. Roberts CW, Shutter JR, Korsmeyer SJ. Hox11 controls the genesis of the spleen.  
1005 Nature. 1994 Apr 21;368(6473):747-9. doi: 10.1038/368747a0.

1006 66. Dear TN, Colledge WH, Carlton MB, Lavenir I, Larson T, Smith AJ, et al. The Hox11  
1007 gene is essential for cell survival during spleen development. Development. 1995  
1008 Sep;121(9):2909-15. doi: 10.1242/dev.121.9.2909.

1009 67. Rice KL, Izon DJ, Ford J, Boodhoo A, Kees UR, Greene WK. Overexpression of stem  
1010 cell associated ALDH1A1, a target of the leukemogenic transcription factor  
1011 TLX1/HOX11, inhibits lymphopoiesis and promotes myelopoiesis in murine  
1012 hematopoietic progenitors. Leuk Res. 2008 Jun;32(6):873-83. doi:  
1013 10.1016/j.leukres.2007.11.001.

1014 68. Lenti E, Farinello D, Yokoyama KK, Penkov D, Castagnaro L, Lavorgna G, et al.  
1015 Transcription factor TLX1 controls retinoic acid signaling to ensure spleen  
1016 development. J Clin Invest. 2016 Jul 1;126(7):2452-64. doi: 10.1172/JCI82956.

1017 69. Kashimada K, Svingen T, Feng CW, Pelosi E, Bagheri-Fam S, Harley VR, et al.  
1018 Antagonistic regulation of *Cyp26b1* by transcription factors SOX9/SF1 and FOXL2  
1019 during gonadal development in mice. FASEB J. 2011 Oct;25(10):3561-9. doi:  
1020 10.1096/fj.11-184333.

1021 70. Rhinn M, Dollé P. Retinoic acid signalling during development. Development. 2012  
1022 Mar;139(5):843-58. doi: 10.1242/dev.065938.

1023 71. Bowles J, Knight D, Smith C, Wilhelm D, Richman J, Mamiya S, et al. Retinoid  
1024 signaling determines germ cell fate in mice. Science. 2006 Apr 28;312(5773):596-600.  
1025 doi: 10.1126/science.1125691.

1026 72. Zangen D, Kaufman Y, Banne E, Weinberg-Shukron A, Abulibdeh A, Garfinkel BP, et  
1027 al. Testicular differentiation factor SF-1 is required for human spleen development. J  
1028 Clin Invest. 2014 May;124(5):2071-5. doi: 10.1172/JCI73186.

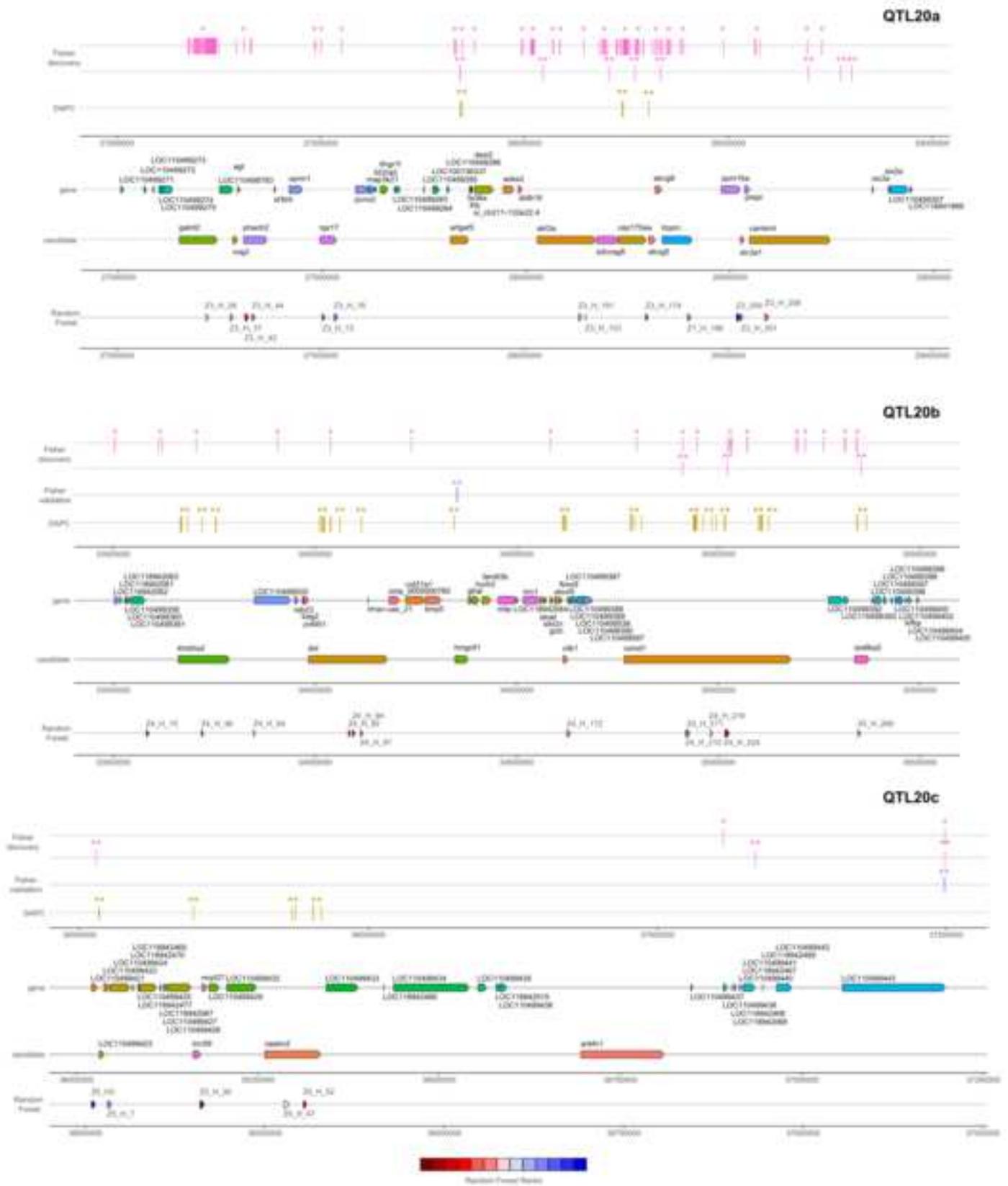
1029 73. Kim JW, Akiyama M, Park JH, Lin ML, Shimo A, Ueki T, et al. Activation of an  
1030 estrogen/estrogen receptor signaling by BIG3 through its inhibitory effect on nuclear  
1031 transport of PHB2/REA in breast cancer. Cancer Sci. 2009 Aug;100(8):1468-78. doi:  
1032 10.1111/j.1349-7006.2009.01209.x.

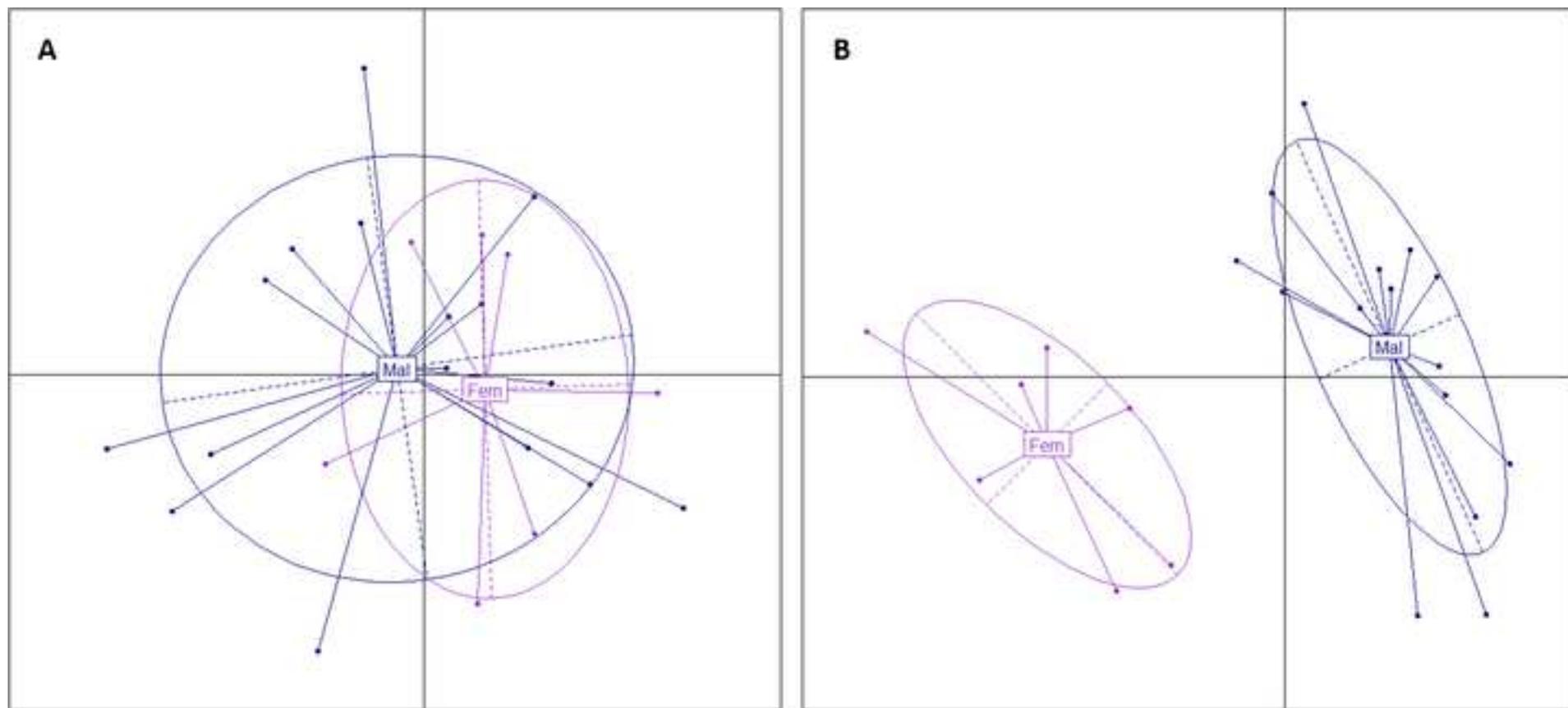
- 1033 74. Jones AR, Schedl T. Mutations in *gld-1*, a female germ cell-specific tumor suppressor  
1034 gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated  
1035 protein Sam68. *Genes Dev.* 1995 Jun 15;9(12):1491-504. doi: 10.1101/gad.9.12.1491.
- 1036 75. Lee MH, Schedl T. Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif  
1037 containing protein required for *C. elegans* germ cell development. *Genes Dev.* 2001 Sep  
1038 15;15(18):2408-20. doi: 10.1101/gad.915901.
- 1039 76. Lee MH, Schedl T. *C. elegans* star proteins, GLD-1 and ASD-2, regulate specific RNA  
1040 targets to control development. *Adv Exp Med Biol.* 2010;693:106-22. doi: 10.1007/978-  
1041 1-4419-7005-3\_8.
- 1042 77. Ding Y, Zhao L, Yang S, Jiang Y, Chen Y, Zhao R, et al. A young *Drosophila* duplicate  
1043 gene plays essential roles in spermatogenesis by regulating several Y-linked male  
1044 fertility genes. *PLoS Genet.* 2010 Dec 23;6(12):e1001255. doi:  
1045 10.1371/journal.pgen.1001255.
- 1046 78. Frisone P, Pradella D, Di Matteo A, Belloni E, Ghigna C, Paronetto MP. SAM68: Signal  
1047 Transduction and RNA Metabolism in Human Cancer. *Biomed Res Int.*  
1048 2015;2015:528954. doi: 10.1155/2015/528954.
- 1049 79. Ehrmann I, Fort P, Elliott DJ. STARS in the CNS. *Biochem Soc Trans.* 2016 Aug  
1050 15;44(4):1066-72. doi: 10.1042/BST20160084.
- 1051 80. Wang S, Yang Q, Wang Z, Feng S, Li H, Ji D, et al. Evolutionary and Expression  
1052 Analyses Show Co-option of *khdrbs* Genes for Origin of Vertebrate Brain. *Front Genet.*  
1053 2018 Jan 4;8:225. doi: 10.3389/fgene.2017.00225.
- 1054 81. Zhang LY, Zeng M, Chen P, Sun HQ, Tao DC, Liu YQ, et al. Identification of  
1055 messenger RNA substrates for mouse T-STAR. *Biochemistry (Mosc).* 2009  
1056 Nov;74(11):1270-7. doi: 10.1134/s0006297909110145.
- 1057 82. Ehrmann I, Dalgliesh C, Liu Y, Danilenko M, Crosier M, Overman L, et al. The tissue-  
1058 specific RNA binding protein T-STAR controls regional splicing patterns of neurexin  
1059 pre-mRNAs in the brain. *PLoS Genet.* 2013 Apr;9(4):e1003474. doi:  
1060 10.1371/journal.pgen.1003474.
- 1061 83. Healy TM, Schulte PM. Patterns of alternative splicing in response to cold acclimation  
1062 in fish. *J Exp Biol.* 2019 Mar 7;222(Pt 5):jeb193516. doi: 10.1242/jeb.193516.
- 1063 84. Park JH, Woo YM, Youm EM, Hamad N, Won HH, Naka K, et al. HMGCLL1 is a  
1064 predictive biomarker for deep molecular response to imatinib therapy in chronic  
1065 myeloid leukemia. *Leukemia.* 2019 Jun;33(6):1439-1450. doi: 10.1038/s41375-018-  
1066 0321-8.

- 1067 85. Yuzaki M. Synapse formation and maintenance by C1q family proteins: a new class of  
1068 secreted synapse organizers. *Eur J Neurosci*. 2010 Jul;32(2):191-7. doi: 10.1111/j.1460-  
1069 9568.2010.07346.x.
- 1070 86. Lee AS, Rusch J, Lima AC, Usmani A, Huang N, Lepamets M, et al. Rare mutations in  
1071 the complement regulatory gene CSMD1 are associated with male and female  
1072 infertility. *Nat Commun*. 2019 Oct 11;10(1):4626. doi: 10.1038/s41467-019-12522-w.
- 1073 87. Hannigan MM, Hoffman AM, Thompson JW, Zheng T, Nicchitta CV. Quantitative  
1074 Proteomics Links the LRRC59 Interactome to mRNA Translation on the ER Membrane.  
1075 *Mol Cell Proteomics*. 2020 Nov;19(11):1826-1849. doi: 10.1074/mcp.RA120.002228.
- 1076 88. Mueller SB, Arora S, Mattocks N, Gurley SB, Kontos CD. Caskin2 is a novel regulator  
1077 of endothelial cell quiescence. *Arterioscler Thromb Vasc Biol*. 2014;A14:34.  
1078 [https://doi.org/10.1161/atvb.34.suppl\\_1.14](https://doi.org/10.1161/atvb.34.suppl_1.14)
- 1079 89. Gao J, Liu S, Zhang Y, Yang Y, Yuan C, Chen S, et al. Effects of 17  $\alpha$ -  
1080 methyltestosterone on transcriptome, gonadal histology and sex steroid hormones in  
1081 rare minnow *Gobiocypris rarus*. *Comp Biochem Physiol Part D Genomics Proteomics*.  
1082 2015 Sep;15:20-7. doi: 10.1016/j.cbd.2015.05.001.
- 1083 90. Monson C, Forsgren K, Goetz G, Harding L, Swanson P, Young G. A teleost androgen  
1084 promotes development of primary ovarian follicles in coho salmon and rapidly alters  
1085 the ovarian transcriptome. *Biol Reprod*. 2017 Nov 1;97(5):731-745. doi:  
1086 10.1093/biolre/iox124.
- 1087 91. Baron D, Cocquet J, Xia X, Fellous M, Guiguen Y, Veitia RA. An evolutionary and  
1088 functional analysis of FoxL2 in rainbow trout gonad differentiation. *J Mol Endocrinol*.  
1089 2004 Dec;33(3):705-15. doi: 10.1677/jme.1.01566.
- 1090 92. von Schalburg KR, Yasuike M, Davidson WS, Koop BF. Regulation, expression and  
1091 characterization of aromatase (*cyp19b1*) transcripts in ovary and testis of rainbow trout  
1092 (*Oncorhynchus mykiss*). *Comp Biochem Physiol B Biochem Mol Biol*. 2010  
1093 Feb;155(2):118-25. doi: 10.1016/j.cbpb.2009.10.015.
- 1094 93. von Schalburg KR, Yasuike M, Yazawa R, de Boer JG, Reid L, So S, et al. Regulation  
1095 and expression of sexual differentiation factors in embryonic and extragonadal tissues  
1096 of Atlantic salmon. *BMC Genomics*. 2011 Jan 13;12:31. doi: 10.1186/1471-2164-12-  
1097 31.
- 1098 94. Nishimura T, Sato T, Yamamoto Y, Watakabe I, Ohkawa Y, Suyama M, et al. Sex  
1099 determination. *foxl3* is a germ cell-intrinsic factor involved in sperm-egg fate decision  
1100 in medaka. *Science*. 2015 Jul 17;349(6245):328-31. doi: 10.1126/science.aaa2657.

- 1101 95. Kikuchi M, Nishimura T, Ishishita S, Matsuda Y, Tanaka M. foxl3, a sexual switch in  
1102 germ cells, initiates two independent molecular pathways for commitment to oogenesis  
1103 in medaka. Proc Natl Acad Sci U S A. 2020 Jun 2;117(22):12174-12181. doi:  
1104 10.1073/pnas.1918556117.
- 1105 96. Gao Y, Jia D, Hu Q, Li D. Foxl3, a Target of miR-9, Stimulates Spermatogenesis in  
1106 Spermatogonia During Natural Sex Change in *Monopterus albus*. Endocrinology. 2016  
1107 Nov;157(11):4388-4399. doi: 10.1210/en.2016-1256.
- 1108 97. Hattori RS, Castañeda-Cortés DC, Arias Padilla LF, Strobl-Mazzulla PH, Fernandino  
1109 JI. Activation of stress response axis as a key process in environment-induced sex  
1110 plasticity in fish. Cell Mol Life Sci. 2020 Nov;77(21):4223-4236. doi: 10.1007/s00018-  
1111 020-03532-9.
- 1112 98. Magerhans A, Müller-Belecke A, Hörstgen-Schwark G. Effect of rearing temperatures  
1113 post hatching on sex ratios of rainbow trout (*Oncorhynchus mykiss*) populations.  
1114 Aquaculture. 2009;294:25–29
- 1115 99. Valdivia K, Jouanno E, Volff JN, Galiana-Arnoux D, Guyomard R, Helary L, et al. High  
1116 temperature increases the masculinization rate of the all-female (XX) rainbow trout  
1117 "Mal" population. PLoS One. 2014 Dec 12;9(12):e113355. doi:  
1118 10.1371/journal.pone.0113355.

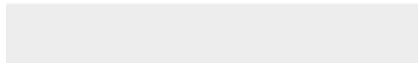








Click here to access/download  
**Supporting Information**  
S1-Appendix.docx





Click here to access/download  
**Supporting Information**  
S1-Supplementary Tables.xls

