

### Double maternal effect: duplicated nucleoplasmin 2 genes, npm2a and npm2b, are shared by fish and tetrapods, and have distinct and essential roles in early embryogenesis

Caroline Cheung, Jérémy Pasquier, Aurélien Bouleau, Thuy Thao Vi Nguyen, Franck Chesnel, Yann Guiguen, Julien Bobe

### ▶ To cite this version:

Caroline Cheung, Jérémy Pasquier, Aurélien Bouleau, Thuy Thao Vi Nguyen, Franck Chesnel, et al.. Double maternal effect: duplicated nucleoplasmin 2 genes, npm2a and npm2b, are shared by fish and tetrapods, and have distinct and essential roles in early embryogenesis. 2020. hal-01608935

### HAL Id: hal-01608935 https://hal.science/hal-01608935

Preprint submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

1	Double maternal effect: duplicated nucleoplasmin 2 genes, <i>npm2a</i> and <i>npm2b</i> , are
2	shared by fish and tetrapods, and have distinct and essential roles in early
3	embryogenesis
4	Caroline T. Cheung <sup>(1, #)</sup> , Jérémy Pasquier <sup>(1, #)</sup> , Aurélien Bouleau <sup>(1)</sup> , Thao-Vi Nguyen <sup>(1)</sup> , Franck
5	Chesnel <sup>(2)</sup> , Yann Guiguen <sup>(1)</sup> , and Julien Bobe <sup>(1)</sup> *
6	
7	<sup>(1)</sup> INRA LPGP UR1037, Campus de Beaulieu, 35042 Rennes, FRANCE.
8	<sup>(2)</sup> CNRS/ UMR6290, Université de Rennes 1, 35000 Rennes, FRANCE.
9	
10	* Corresponding author E-mail: julien.bobe@inra.fr
11	Authors contributed equally to this work: #
12	
13	Short Title: <i>npm2</i> genes have maternal effect in vertebrates

1

#### 14 Abstract

15 *Nucleoplasmin 2 (npm2)* is an essential maternal-effect gene that mediates early embryonic 16 events through its function as a histone chaperone that remodels chromatin. Here we report 17 the existence of two *npm2* (*npm2a* and *npm2b*) genes in zebrafish. We examined the 18 evolution of *npm2a* and *npm2b* in a variety of vertebrates, their potential phylogenetic 19 relationships, and their biological functions using knockout models via the CRISPR/cas9 20 system. We demonstrated that the two npm2 duplicates exist in a wide range of vertebrates, 21 including sharks, ray-finned fish, amphibians, and sauropsids, while *npm2a* was lost in 22 Coelacanth and mammals, as well as some specific teleost lineages. Using phylogeny and 23 synteny analyses, we traced their origins to the early stages of vertebrate evolution. Our 24 findings suggested that *npm2a* and *npm2b* resulted from an ancient local gene duplication, 25 and their functions diverged although key protein domains were conserved. We then 26 investigated their functions by examining their tissue distribution in a wide variety of species 27 and found that they shared ovarian-specific expression, a key feature of maternal-effect 28 genes. We also showed that both *npm2a* and *npm2b* are maternally-inherited transcripts in 29 vertebrates. Moreover, we used zebrafish knockouts to demonstrate that *npm2a* and *npm2b* 30 play essential, but distinct, roles in early embryogenesis. npm2a functions very early during 31 embryogenesis, at or immediately after fertilization, while *npm2b* is involved in processes 32 leading up to or during zygotic genome activation. These novel findings will broaden our 33 knowledge on the evolutionary diversity of maternal-effect genes and underlying mechanisms 34 that contribute to vertebrate reproductive success.

Version preprint

35

36

#### 37 Author Summary

38 The protein and transcript of the *npm2* gene have been previously demonstrated as maternal 39 contributions to embryos of several vertebrates. Recently, two npm2 genes, denoted here as 40 *npm2a* and *npm2b*, were discovered in zebrafish. This study was conducted to explore the 41 evolutionary origin and changes that occurred that culminated in their current functions. We 42 found that an ancient local duplication of the ancestral npm2 gene created the current two 43 forms, and while most vertebrates retained both genes, notably, mammals and certain species 44 of fish lost *npm2a* and, albeit rarely, both *npm2a* and *npm2b*. Our functional analyses showed 45 that *npm2a* and *npm2b* have diverse but essential functions during embryogenesis, as *npm2a* 46 mutants failed to undergo development at the earliest stage while *npm2b* mutants developed, 47 although abnormally, until the zygotic genome activation stage after which their development 48 was arrested followed subsequently by death. Our study is the first to clearly demonstrate the 49 evolution, diversification, and functional analyses of the *npm2* genes, which are essential 50 maternal factors that are required for proper embryonic development and survival. 51

52

#### 53 Introduction

54 In animals and plants, early embryonic development relies strictly on maternal 55 products until maternal-to-zygotic transition (MZT) during which zygotic genome activation 56 (ZGA) occurs [1]. Maternal-effect genes are those that are transcribed from the maternal 57 genome and whose products, which include transcripts, proteins, and other biomolecules, are 58 deposited into the oocytes during their production in order to coordinate embryonic 59 development before ZGA [2]. MZT is a key step that is needed firstly for clearance of 60 maternal components, and secondly to activate zygotic gene expression and to allow 61 subsequent embryonic development. Among the maternally-inherited transcripts that play

62 important roles during early development, some genes were demonstrated to regulate zygotic 63 program activation such as *nanog*, *pou5f1*, and *soxb1* in zebrafish (*Danio rerio*) [3]. 64 Henceforth, all gene and protein nomenclature will be based on that of zebrafish regardless of 65 species for simplification purposes. Those factors were shown to participate in the regulation 66 of "first wave" zygotic genes and among them, mir-430, a conserved microRNA that has 67 been shown to be involved in clearance of maternal mRNAs in zebrafish [4][5], as well as 68 mir-427, its orthologue in Xenopus [6]. microRNAs may also have maternal effects by 69 functioning in a similar manner to protein-coding genes, and recent revelations showed that 70 there is a subset of microRNAs that are predominantly expressed in the fish ovary and may 71 function in oogenesis and early embryogenesis [7]. Another gene, *nucleoplasmin 2 (npm2)*, 72 belongs to the family of nucleoplasmins/nucleophosmins that is maternally-inherited at both 73 protein and mRNA levels, whereby both play important roles in early development [8]. 74 Historically, this protein was identified and defined as a nuclear chaperone in Xenopus 75 [9,10]. While the protein has been shown to be the most abundant nuclear protein in the 76 Xenopus oocyte [11] and to play a crucial role at fertilization due to its role in sperm 77 chromatin decondensation [12,13], its maternally-inherited mRNA has been recently 78 demonstrated to be translated as a *de novo* synthesized protein that could play a crucial role 79 during ZGA in zebrafish [8]. Further, npm2 is one of the first identified maternal-effect genes 80 in mouse whereby its deficiency results in developmental defects and eventual embryonic 81 mortality [14]. 82 The *npm2* gene belongs to the *npm* gene family that encompasses four members, 83 npm1, npm2, npm3, and npm4. The diversity of the Npm family has been shown to result 84 from the two rounds of whole genome duplication (WGD) that occurred in early vertebrates 85 (vertebrate genome duplication 1 and 2, or VGD1 & VGD2, respectively) [15,16]. Former

86 evolutionary studies clearly provided a phylogenic model of this family; VGD1 produced two

87	genes, <i>npm3/2</i> and <i>npm4/1</i> , from an ancestral <i>npm</i> gene and the following WGD, VGD2,
88	further created the current four npm types with subsequent loss of npm4 from mammals, but
89	retained in most fish species [17-20]. Recently, two npm2 genes were automatically
90	annotated in the zebrafish genome, i.e. npm2a (ENSDARG00000076391) and npm2b
91	(previously known as npm2, ENSDARG00000053963). As the teleost ancestor experienced
92	an extra WGD event (TGD, or teleost-specific genome duplication) [21], doubling of genes
93	and other types of genomic rearrangement may be present in teleost species compared with
94	other vertebrates. Moreover, a fourth round of duplication occurred more recently in
95	salmonids (salmonid-specific genome duplication or SaGD) [22,23], leading to further
96	possible doubling of genes and other genomic rearrangements. As multiple evolutionary
97	events (impact of TGD, local duplication, teleost-specific duplication, etc.) could have led to
98	the npm2a/npm2b diversity in zebrafish, which in turn may have significant impact on
99	vertebrate reproduction, we investigated the evolution of <i>npm2a</i> and <i>npm2b</i> in a wide range
100	of vertebrate species, their potential phylogenetic relationship, as well as their biological
101	functions using transgenic zebrafish models created by the CRISPR/cas9 system in order to
102	broaden our knowledge on the evolutionary diversity of maternal-effect genes and the
103	underlying mechanisms that contribute to reproductive success in vertebrates.
104	
105	
106	Results and Discussion
107	As previously shown, <i>npm1</i> , <i>npm2</i> , <i>npm3</i> and <i>npm4</i> genes are thought to have

108 originated from the first two rounds of WGD, VGD1 and VGD2, which occurred early on in

109 vertebrate evolution [17–20]. However, two *npm2* genes have been automatically annotated

110 in the zebrafish genome, *i.e. npm2a* and *npm2b*. In order to verify if these two *npm2* genes

111 are paralogous to each other and to determine their origination, we used a Blast search

- approach in various public databases to retrieve a multitude of sequences that could be related
- to *npm2* genes. All retrieved sequences are compiled in Supplemental Table S1.
- 114

124

#### 115 Diversity of npm2a and npm2b in vertebrates

#### 116 <u>Phylogenetic analysis</u>

In order to verify that the retrieved protein sequences (Supplemental Table S1) were homologous to zebrafish Npm2a and Npm2b, a phylogenetic analysis on Npm2 was performed. Based on the alignment of 76 vertebrate Npm2-related sequences, and using vertebrate Npm1 and Npm3 amino acid sequences as out-groups, a phylogenetic tree was generated (Fig. 1). As shown in Fig. 1, vertebrate Npm2-related sequences clustered into two clades, Npm2a and Npm2b, which were supported by significant bootstrap values, 84.2% and 78.2%, respectively.

There were 28 sequences that clustered in the Npm2a clade, which encompassed

sequences belonging to species from various vertebrate groups, including chondrichthyans
(such as dogfish Npm2a), ray-finned fish, and sarcopterygians (lobe-finned fish and
tetrapods). The sequences belonging to ray-finned fish included gar and bowfin Npm2a, as
well as teleost sequences such as zebrafish, northern pike, and rainbow trout Npm2a. In
contrast, no Npm2a sequence was identified in neoteleostei (medaka, European perch, and
Atlantic cod). The sequences belonging to sarcopterygians included amphibian (Xenopus)
Npm2a, as well as sauropsid sequences such as Chinese alligator, chicken, penguin, and ibis

132 Npm2a, although no Npm2a sequence could be identified in mammals. These results

133 provided the first evidence for the presence of orthologs to zebrafish Npm2a across

134 evolutionary divergent vertebrate groups (i.e. sauropsids, amphibians, and fish).

135 The Npm2b clade included 48 of the retrieved sequences. Like the Npm2a clade, the

136 Npm2b clade also encompassed sequences belonging to species from all of the main

Comment citer ce document : Cheung, C., Pasquier, J., Bouleau, A., Nguyen, T. T. V., Chesnel, F., Guiguen, Y., Bobe, J. (2017). Double maternal effect: duplicated nucleoplasmin 2 genes, npm2a and npm2b, are shared by fish and tetrapods, and have distinct and essential roles in early embryogenesis. BioRxiv, 1-39.

6

137 vertebrate groups, including chondrichthyans (such as dogfish Npm2b), ray-finned fish, and 138 sarcopterygians. Ray-finned fish sequences included gar and bowfin Npm2b in addition to 139 teleost sequences such as zebrafish and northern pike Npm2b. Our analysis also demonstrated 140 the existence of two paralogous Npm2b proteins (*i.e.* Npm2b1 and Npm2b2) in all of the 141 investigated salmonid species (rainbow trout, brown trout, and brook trout). In contrast, no 142 Npm2b sequence could be identified in any neoteleostean species, including medaka, cod, 143 perch, fugu, tetraodon, and stickleback. The sequences belonging to sarcopterygians included 144 amphibian sequences such as Xenopus Npm2b, sauropsid sequences such as Chinese 145 alligator, ostrich, penguin and chicken Npm2b, as well as mammalian sequences such as 146 human and mouse Npm2. These results demonstrated, for the first time, that proteins 147 previously reported as Npm2 (Xenopus, cattle, mouse, human, zebrafish) are orthologous to 148 Npm2b in all investigated vertebrate species, and should therefore now be referred to as 149 Npm2b. In addition, with the presence of Npm2b1 and Npm2b2 in all investigated salmonid 150 species, we provided for the first time evidence of the existence of two Npm2b protein forms 151 in vertebrate species. 152 The existence of the two Npm2 clades indicated that Npm2a and Npm2b are likely 153 paralogous to each other. In addition, comparison of Npm2a and Npm2b amino acid 154 sequences in the species that harbor both revealed that they share between 30.2% and 46.4% 155 homology, depending on the species (Supplemental Table S2). The low sequence identity is 156 consistent with an ancient duplication event that gave rise to *npm2a* and *npm2b* genes. 157 However, neither the topology of the Npm2 phylogenetic tree nor the comparison between

158 Npm2a and Npm2b could indicate the kind of duplication event that had occurred. In

159 contrast, the high sequence identity shared by Npm2b1 and Npm2b2 in salmonids (between

160 75.1% and 86%) suggested a more recent duplication event. This observation in all of the

5.1% and 86%) suggested a more recent duplication event. This observation in all o

Comment citer ce document : Cheung, C., Pasquier, J., Bouleau, A., Nguyen, T. T. V., Chesnel, F., Guiguen, Y., Bobe, J. (2017). Double maternal effect: duplicated nucleoplasmin 2 genes, npm2a and npm2b, are shared by fish and tetrapods, and have distinct and essential roles in early embryogenesis. BioRxiv, 1-39. DOI : 10.1101/104760

7

161 investigated salmonid species is consistent with the hypothesis that *npm2b1* and *npm2b2* 

- 162 genes likely resulted from the SaGD.
- 163 <u>Synteny analysis</u>

164 In order to further understand the origin of the *npm2a* and *npm2b* genes in vertebrates, 165 we performed a synteny analysis of their neighboring genes in representative vertebrate 166 genomes. We focused our study on two mammals (human and mouse), two sauropsids 167 (Chinese alligator and chicken), two amphibians (Xenopus tropicalis and laevis L), one basal 168 sarcopterygian (coelacanth), one basal actinopterygian (spotted gar), and two teleosts 169 (zebrafish and tetraodon) (Fig. 2). Analysis was also performed on the *Xenopus laevis* S 170 subgenome, but since the results were very similar to that of Xenopus laevis L subgenome, 171 we only showed the latter's data [24]. 172 The human, mouse, Chinese alligator, *Xenopus tropicalis* and *laevis*, coelacanth, 173 spotted gar, and zebrafish *npm2b* genes are located in genomic regions containing common 174 loci, including dok2, xpo7, fgf17, dmtn, lgi3, bmp1, sorbs3, pdlim2, gpr124, prlhr, got111, 175 adrb3, and nkx6-3. Together with the phylogenetic analysis, this indicates that npm2b genes 176 investigated here are orthologous. Synteny analysis revealed the absence of the *npm2b* gene 177 in tetraodon although the above-mentioned neighbouring genes are present in its genome 178 (Fig. 2 and Supplemental Table S3). 179 The Chinese alligator, chicken, Xenopus tropicalis and laevis, spotted gar, and 180 zebrafish *npm2a* genes are located in genomic regions containing the same loci as *npm2b* 181 conserved regions (Fig. 2). Indeed, Chinese alligator and spotted gar *npm2a* and *npm2b* genes 182 are located in the vicinity of each other on scaffold 98.1 and the linkage group LG1, 183 respectively. The presence of *npm2a* and *npm2b* genes in the same genomic region in 184 representative species of sarcopterygians (Chinese alligator and Xenopus tropicalis and

185	<i>laevis</i> ) and actinopterygians	(spotted gar)	strongly suggested	that <i>npm2a</i> and	npm2b genes
	· · · · · · · · · · · · · · · · · · ·	(T	0,00	· · · · · · · · · ·	1 0

186 could have resulted from a unique local duplication of an ancestral *npm2* gene.

187

#### 188 Evolutionary history of npm2 genes in vertebrates

189 The presence/absence of *npm2a* and *npm2b* in the current vertebrate phyla and species 190 is summarized in Fig.3, and we also propose an evolutionary scenario for the diversification 191 of the *npm2* genes across vertebrate evolution.

192 In this study, we demonstrated that *npm2a* and *npm2b* may be paralogous genes

193 present in the different vertebrate groups, chondrichthyans (Fig.1), sarcopterygians, and

actinopterygians (Figs. 1 and 2), which strongly suggested that the *npm2* genes originated

195 from a duplication event prior to the divergence of chondrichthyans and osteichthyans. Since

the four *npm* family members (*npm1*, *npm2*, *npm3*, and *npm4*) are thought to be produced

197 from the first two rounds of WGD (VGD1 & VGD2) that occurred early on in vertebrate

198 evolution, we can thus hypothesize that the duplication event that generated *npm2a* and

*npm2b* took place after VGD2, but before emergence of chondrichthyans and osteichthyans,

between 450 and 500 million years ago (Mya) [25]. The findings from our synteny analysis

201 demonstrated that in representative species of actinopterygians (spotted gar) and

202 sarcopterygians (*Xenopus tropicalis* and *laevis* and *Chinese alligator*), *npm2a* and *npm2b* 

203 genes are at two distinct loci located on the same chromosomic region (Fig. 2). These results

strongly suggested that *npm2a* and *npm2b* genes arose from local gene duplication rather than

a whole genome (or chromosome) duplication event (Fig. 3).

206 The teleost ancestor experienced an extra WGD event (TGD) [21], but in all

207 investigated teleosts, we observed a maximum of one *npm2a* ortholog and one *npm2b* 

208 ortholog (except in salmonids) (Fig. 1). In addition, *npm2a* and *npm2b* are located on two

209 TGD ohnologous regions on chromosomes 8 and 10, respectively (Fig. 2), in zebrafish,

210 which is consistent with the loss of one of the *npm2a* duplicates from one region and the loss 211 of one *npm2b* duplicate from the corresponding ohnologous region. Thus, *npm2* diversity is 212 most likely due to the early loss of one of the two *npm2a* and *npm2b* TGD ohnologs (Fig. 3). 213 This observation in zebrafish strengthened the hypothesis that TGD did not impact the *npm2a* 214 and *npm2b* diversity in teleosts. In addition, the lack of *npm2a* and *npm2b* in neoteleostei 215 species, such as tetraodon, suggested that additional gene losses occurred early in the 216 evolutionary history of this group (Fig. 2). 217 In salmonids, we identified two *npm2b* paralogs, i.e. *npm2b1* and *npm2b2*, in all 218 investigated salmonid species (Fig. 1). Considering the high sequence identity shared by 219 *npm2b1* and *npm2b2*, it is strongly hypothesized that these duplicates originated from SaGD. 220 In contrast, we identified only one *npm2a* gene in all investigated salmonids suggesting that 221 SaGD did not have any impact on the current salmonid *npm2a* diversity mostly due to an 222 early loss of the SaGD ohnolog of this gene (Fig. 3). 223 In addition to the early gene losses after WGD, various other independent and 224 phylum-specific gene losses may have contributed to shape the current diversity of *npm2a* 225 and *npm2b* in vertebrates. In fact, although both *npm2a* and *npm2b* have been globally 226 conserved in sarcopterygians and actinopterygians, some phyla in each group lack at least one 227 of the genes. In sarcopterygians, *npm2a* is conserved in amphibians such as *Xenopus* 228 tropicalis and laevis, and in sauropsids such as Chinese alligator and chicken (Figs. 1-3). In 229 contrast, we did not find any *npm2a* genes in the Coelacanth genome (Figs. 1-3), which could 230 be due to the lower assembly quality of the concerned genomic region (see Supplemental 231 Table S3), and in the mammalian genomes including human and mouse (Figs. 1-3), whose 232 absence may have been due to the loss of this gene in the common ancestor of mammals. On 233 the other hand, *npm2b* is conserved in most investigated species. In actinopterygians, *npm2a* 234 and *npm2b* are both conserved in all investigated species except neoteleostei in which neither

235 *npm2a* nor *npm2b* is found. In contrast, one *npm2a* and two *npm2b*, i.e. *npm2b1* and *npm2b2*, 236 are found in salmonids, representing to date the most important diversity of *npm2* genes in 237 vertebrates (Figs. 1-3). Considering the essentialness of the *npm2* gene in embryonic 238 development, its lack in some neoteleosteon species, such as Atlantic cod and medaka, raises 239 the question on how evolution can cope with its loss. Further analyses of data from the 240 Phylofish database [26] (data not shown) showed that *npm3* had the strongest homology to 241 npm2 and was predominantly expressed in the ovaries in medaka and cod which suggest that 242 it could potentially compensate for npm2 deficiency. In contrast, npm3 does not show an 243 ovarian-predominant expression in spotted gar and zebrafish, thus suggesting that its strong 244 ovarian expression could be restricted to neoteleostei.

245

#### 246 Evolution of npm2a and npm2b coding sequences

247 To examine the evolutionary adaptation of the *npm2* proteins, estimation of the ratio 248 of substitution rates (dN/dS) between the paralogous npm2a and npm2b CDS for all of the 249 species harbouring multiple *npm2* paralogs was performed (Supplemental Table S4). For all 250 investigated species, the dN/dS values were well below 1, indicating that whilst *npm2a* and 251 *npm2b* genes diverged approximately 500 Mya, they have remained under strong purifying 252 selection and evolutionary pressure, which may have tended to conserve their distinct protein 253 structures and functions. Thus, in this study, we found that Npm2a and Npm2b may have 254 diverged from an ancient duplication and currently share low sequence identity, suggesting 255 that they could have thus evolved with different roles, which were likely conserved through 256 strong negative selection.

257

Comment citer ce document : Cheung, C., Pasquier, J., Bouleau, A., Nguyen, T. T. V., Chesnel, F., Guiguen, Y., Bobe, J. (2017). Double maternal effect: duplicated nucleoplasmin 2 genes, npm2a and npm2b, are shared by fish and tetrapods, and have distinct and essential roles in early embryogenesis. BioRxiv, 1-39.

11

#### 258 Expression domains of Npm2a and Npm2b in vertebrates

259	In order to investigate further the potential functions of Npm2a and Npm2b, we first
260	explored the tissue distributions of both transcripts using two different approaches, qPCR and
261	RNA-Seq, the latter of which was obtained from the Phylofish online database [26]. In
262	zebrafish and Xenopus tropicalis, we observed that npm2a and npm2b were both
263	predominantly expressed in the ovary, and to a lesser extent, the muscle, as well as in the
264	zebrafish gills (Fig. 4A and 4B). We also demonstrated that <i>npm2a</i> and <i>npm2b</i> were
265	predominantly expressed in the ovary of bowfin, elephantnose fish, panga, European eel,
266	sweetfish, and northern pike (Fig. 4C to 4H). In the investigated salmonid species, brook
267	trout and brown trout, npm2a, npm2b1, and npm2b2 were also predominantly expressed in
268	the ovary (Fig. 4I and 4J). In addition, npm2 transcripts were expressed at a very low level in
269	the testis of European eel, northern pike, and salmonid species (Fig. 4F, 4H, 4I, and 4J). In
270	teleosts, npm2a mRNA levels globally tended to be lower than npm2b (or npm2b1 and
271	npm2b2 for salmonids) (Fig. 4A and 4D-J). The clear ovarian-specific expression profiles of
272	npm2a and npm2b transcripts were conserved in all investigated species, from teleosts to
273	tetrapods. This is consistent with the npm2b mRNA profiles reported in the literature for
274	mouse [14], cattle [27], Xenopus tropicalis [28], and zebrafish [8]. Despite the long passage
275	of time since their divergence (approximately 500 Mya) [25] as demonstrated above, npm2a
276	and <i>npm2b</i> appeared to have conserved their ovarian-specific expression profiles, which
277	suggested that they both have roles in female reproduction and/or embryonic development.
278	
279	Embryonic expression of npm2a and npm2b
280	Thus, in order to delve deeper into their functions in reproduction and embryogenesis,

281 we investigated *npm2a* and *npm2b* mRNA expression during oogenesis and early

development in zebrafish. During zebrafish oogenesis, both npm2a and npm2b transcripts

293	Peptidic domains of npm2 paralogs
292	
291	<i>npm2a</i> is also a maternal-effect gene.
290	are typical features of maternally-inherited mRNAs, which highly suggested that the novel
289	previous studies on zebrafish npm2a and npm2b transcripts [8,29]. Their expression profiles
288	mRNAs are strictly maternal ( <i>i.e.</i> not re-expressed by the zygote), which is consistent with
287	reaching very low levels at 24 hours post-fertilization (hpf) (Fig. 5B). This suggests that both
286	levels progressively decreased during embryonic development after fertilization before
285	unfertilized egg (i.e. metaphase 2 oocyte) (Fig. 5B). Thereafter, npm2a and npm2b transcript
284	both npm2 during oogenesis, they still can be detected at reasonable amounts in the
283	were found at high levels in oocytes (Fig. 5A), and despite gradual decreases in the levels of

294 We further examined the functions of the Npm2 proteins by analysis of their protein 295 domains. To date, only the role of *npm2b* has been investigated in *Xenopus tropicalis* 296 [12,30,31], mouse [14,32], humans [33–35], cattle [27], and zebrafish [8]. As previously 297 demonstrated, *npm2b* is a maternal-effect gene whose transcripts are accumulated in the 298 growing oocyte and maternally-inherited by the zygote, where it functions as a histone 299 chaperone to decondense sperm DNA as well as reorganize chromatin, thus, it has also been 300 suggested to contribute to ZGA as well [8,14,27,28]. Npm2b is thought to be activated by 301 various post-translational modifications and homo-pentamerisation [36], and subsequently 302 interacts with chromatin by exchanging sperm-specific basic proteins with histones via its 303 core domain and acidic tracts A1, A2 and A3 [34,37–42]. However, no data is available on 304 the structure and function of Npm2a. Using a predictive approach, we identified the presence 305 of the Npm core domain, acidic tract A1, and acidic tract A2 in the various investigated 306 Npm2 sequences. The nucleoplasmin family is defined by the presence of an Npm core 307 domain, which enables oligomerization of Npm proteins, in all of its members [38,43]. We

13

308	were able to predict the presence of this domain in all investigated Npm2a and Npm2b
309	sequences (Supplemental Fig. S1), suggesting that Npm2a proteins could also form homo- or
310	hetero-polymers. The acidic tract domains were demonstrated to facilitate histone binding by
311	increasing the recognition and affinity for different histones [40,44]. Acidic tract A1 was
312	demonstrated to be absent from most of the Npm2b investigated so far, except Xenopus
313	tropicalis Npm2b, and acidic tract A2 was predicted to be in all investigated Npm2b apart
314	from guinea pig Npm2b and American whitefish Npm2b1 (Supplemental Fig. S1). This
315	strongly suggested that the histone and basic protein binding activity of Npm2b is mediated
316	predominantly by acidic tract A2. On the other hand, in all investigated Npm2a proteins, we
317	predicted an acidic tract A1 except in that of dogfish, zebrafish, and allis shad (Supplemental
318	Fig. S1). In contrast, only half of the investigated Npm2a proteins harbored an acidic tract
319	A2, which is additionally shorter than the one present in Npm2b proteins (Supplemental Fig.
320	S1). Our results demonstrated that all Npm2a proteins, except those of Allis shad and
321	dogfish, possess acidic tracts, which could potentially mediate histone and basic protein
322	interactions.
323	

#### 324 Functional analysis of npm2a and npm2b in zebrafish

325 Lastly, we performed functional analysis of these two Npm2 proteins by genetic 326 knockout using the CRISPR/cas9 system. One-cell staged embryos were injected with the 327 CRISPR/cas9 guides that targeted either *npm2a* or *npm2b* and allowed to grow to adulthood. 328 Mosaic founder mutant females (F0) were identified by fin clip genotyping and subsequently 329 mated with wild-type (WT) males, and embryonic development of the F1 fertilized eggs was 330 recorded. Since the mutagenesis efficiency of the CRISPR/cas9 system was very high, as 331 previously described [45,46], the *npm2* genes were sufficiently knocked-out even in the 332 transgenic mosaic F0 females. This was evidenced by the substantially lower transcript levels

333 of *npm2a* and *npm2b* in the F1 embryos as compared to those from control WT pairings 334 (Fig.6A). Thus, the phenotypes of npm2a (n=3) and npm2b (n=4) mutants could be observed 335 even in the F0 generation. Since none of the mutated genes were transmissible to future 336 generations neither through the male nor the female, therefore, all of our observations were 337 obtained from the F0 generation. We observed that most of the embryos from the *npm2b* 338 mutant females underwent cellular division during the very early stages of development (1-3 339 hpf) despite a considerable number of embryos with abnormal morphology (24.00±7.84%) 340 versus 0% in controls) (Fig.6B), which included smaller size, enlarged yolk to membrane 341 ratio, and small volk to membrane ratio. The diameter of the embryo is demonstrated by the 342 red dotted lines at oblong and germ ring stages, which reveal the extremely reduced size of 343 *npm2a* embryos (Fig.6C). Notably, around two thirds of the embryos showed abnormal cell 344 division, even in those with normal morphology, that culminated in developmental arrest at 345 around 4 hpf following which the cells stopped dividing and appeared to regress. Thus, the 346 developmental success at 4 hpf (oblong/sphere stage) was 38.09±7.31% vs. 85.09±4.88% in 347 controls (Fig.6B and 6C [right column]). By the 24 somite stage (24 hpf), these embryos were 348 all dead while the remaining embryos that showed normal development and normal cell 349 division continued to progress normally. This finding corresponded to our previous results, 350 which showed that *npm2b*-deficient embryos targeted by morpholino arrested at 4 hpf and 351 eventually died [8]. 352 On the other hand, *npm2a* mutant-derived embryos had a very low developmental 353 success even at a very early stage of growth (1 hpf)  $(17.62\pm5.06\% \text{ vs. } 85.09\pm4.88\% \text{ in})$ 354 controls) as defined by a complete lack of cell division, and a large population of them 355 (36.36±6.98% vs. 0% in controls) had an abnormal morphology (Fig.6B and 6C [middle 356 columns]). The abnormal morphology of these embryos was similar to that found in *npm2b*-

358 to display a complete lack of development at 64-cell, oblong, dome, germ ring, and finally 359 somite stages (Fig.6B [middle columns]). Similar to the *npm2b* mutant-derived embryos, the 360 *npm2a*-deficient embryos were all dead by 24 hpf while the remaining embryos that showed 361 normal development and normal cell division continued to progress normally. This novel 362 finding showed for the first time that *npm2a* is essential for early development of embryos, 363 and is therefore a crucial maternal-effect gene. Further, we demonstrated that while *npm2a* 364 and *npm2b* share similar tissue distribution, as both are found specifically in the ovaries and 365 early stage embryos, their roles are distinct and essential to embryogenesis, and one could not 366 compensate for the other. Notably, *npm2a* probably plays a major role at or immediately after 367 fertilization since the embryos are arrested before the first mitotic division, which is in line 368 with a role in sperm DNA decondensation since lack of proper processing of the paternal-369 derived DNA would lead to developmental arrest at the earliest stage and embryonic death. In 370 contrast, *npm2b* appears to function at a later stage or other factors can replace its role in the 371 early stages as the *npm2b*-deficient embryos are capable of dividing until around 4 hpf, which 372 corresponds to previous reports [8,14]. These findings are in line with those demonstrated for 373 other maternal-effect genes in mammals, such as *mater*, *floped*, *filia*, *tle6* [47], and *bcas2* 374 [48] as well as in zebrafish, including *fue* [49] and *cellular atoll* [50], all of which function to 375 moderate the early events of fertilization and embryogenesis and their ablation leads to arrest 376 and mortality at the earliest stages of embryogenesis. 377 In consequence, the dominancy of the mutant allele combined with the strong 378 maternal effect of these genes, which led to massive early embryonic mortality such that none 379 of the mutant carriers survived, very likely contributed to the lack of germline transmission of 380 the mutations despite reported heritability rates of 10-30% in CRISPR/cas9-targeted zebrafish 381 from previous studies [45,46]. This was supported by detection of the mutant allele via PCR 382 genotyping in the genome of F1 npm2a- and npm2b-deficient embryos that were

developmentally-arrested at the germ ring stage, but not in the surviving animals at 24 hpf(Supplemental Fig.S2).

385

407

#### 386 Conclusions

387 Since two annotated genes corresponding to *npm2*, which is known as an essential 388 maternal-effect gene in mammals and amphibians, were recently found in zebrafish, we set 389 out to investigate their evolution and function. In this study, we demonstrated that the two 390 duplicates of *npm2*, *npm2a* and *npm2b*, exist in a wide range of vertebrate species, including 391 ray-finned fish, amphibians, and bird. We also found that the mammalian *npm*<sup>2</sup> gene is in 392 fact an ortholog of *npm2b*. Using phylogeny and synteny analyses, we traced the origins of 393 those two duplicates to the early stages of vertebrate evolution. Our findings indicate that 394 *npm2a* and *npm2b* genes resulted from a local gene duplication that may have occurred 395 between VGD2 and the divergence of ray-finned fish and tetrapods (~450-500 Mya). This 396 ancient origin is in line by the low sequence identity between the two *npm2* genes, although 397 the main protein domains remain conserved. Both genes exhibit a strict ovarian expression 398 and corresponding transcripts are maternally-inherited in fish and amphibians. Moreover, we 399 demonstrated and confirmed by CRISPR/cas9-directed genetic knockout of npm2a and 400 *npm2b* that they are both maternal-effect genes that play essential, but distinct, roles in early 401 embryogenesis. *npm2a* probably plays a major role at or immediately after fertilization, and 402 most likely in chromatin sperm decondensation. In contrast, *npm2b* appears to function at a 403 later stage and could participate in ZGA. Our findings will help us gain further insight into 404 the evolutionary diversity of maternal-effect genes and understand the underlying 405 mechanisms that contribute to reproductive success in vertebrates. 406

17

bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

#### 408 Material and Methods

#### 409 Genomic databases

- 410 The following genomic data were extracted and investigated from the ENSEMBL genomic
- 411 database (<u>http://www.ensembl.org/index.html</u>): human, *Homo sapiens*; mouse, *Mus*
- 412 musculus; chicken, Gallus gallus; Xenopus, Xenopus tropicalis; coelacanth, Latimeria
- 413 *chalumnae*; spotted gar, *Lepisosteus oculatus*; zebrafish, *Danio rerio*; and tetraodon,
- 414 *Tetraodon nigroviridis*. The Chinese alligator (*Alligator sinensis*) genome was extracted and
- 415 investigated from the NCBI genomic database (<u>http://www.ncbi.nlm.nih.gov/genome/22419</u>).
- 416 The Xenopus laevis L and S genomes were analyzed from the Xenbase database
- 417 (www.xenbase.org).

#### 418 Transcriptomic databases

- 419 The following actinopterygian transcriptomes were retrieved and investigated from the
- 420 Phylofish database (http://phylofish.sigenae.org/index.html): bowfin, *Amia calva*; spotted gar,
- 421 Lepisosteus oculatus; elephantnose fish, Gnathonemus petersi; arowana, Osteoglossum
- 422 *bicirrhosum*; butterflyfish, *Pantodon buchholzi*; European eel, *Anguilla anguilla*; rainbow
- 423 trout, Oncorhynchus mykiss; allis shad, Alosa alosa; zebrafish, Danio rerio; panga,
- 424 Pangasius hypophthalmus; northern pike, Esox lucius; grayling, Thymallus thymallus;
- 425 Atlantic cod, *Gadhus morua*; medaka, *Oryzias latipes*; European perch, *Perca fluviatilis*;
- 426 brown trout, *Salmo trutta*; European whitefish, *Coregonus lavaretus*; brook trout, *Salvelinus*
- 427 *fontinalis*; Astyanax, Astyanax mexicanus; lake whitefish, Coregonus clupeaformis; eastern
- 428 mudminnow, Umbra pygmae, and sweetfish, Plecoglossus altivelis.
- 429 *Gene predictions*
- 430 <u>TBLASTN search</u>
- 431 Genomic data were analyzed using the TBLASTN algorithm (search sensitivity: near exact
- 432 match short) on the ENSEMBL website or the NCBI browser for the Chinese alligator

- 433 genome. The TBLASTN algorithm on the SIGENAE platform was used on the
- 434 transcriptomic data.
- 435 Predictions of npm2 genes
- 436 The peptidic sequences of zebrafish Npm2a and Npm2b were used as query in TBLASTN
- 437 search to identify the open reading frame (ORF) encoding *npm2* genes in the various
- 438 investigated genomes and transcriptomes.
- 439 *Phylogenetic analysis*
- 440 Amino acid sequences of 75 Npm2, 3 Npm3, and 3 Npm1 proteins were first aligned using
- 441 ClustalW. The JTT (Jones, Taylor, and Thornton) protein substitution matrix of the resulting
- 442 alignments was determined using ProTest software. Phylogenetic analysis of Npm proteins
- 443 was performed using the Maximum Likelihood method (MEGA 5.1 software) with 1,000
- 444 bootstrap replicates.
- 445 Synteny analyses
- 446 Synteny maps of the conserved genomic regions in human, mouse, chicken, Xenopus,
- 447 coelacanth, spotted gar, zebrafish, and tetraodon were produced using PhyloView on the
- 448 Genomicus v75.01 website (http://www.genomicus.biologie.ens.fr/genomicus-75.01/cgi-
- 449 <u>bin/search.pl</u>). Synteny analysis of the Chinese alligator conserved genomic regions was
- 450 performed using TBLASTN searches in the corresponding genomic database. For each gene,
- 451 the peptidic sequences of human and chicken were used as query, as far as they were
- 452 referenced in the databases.
- 453 RNA-seq
- 454 RNA-seq data were deposited into Sequence Read Archive (SRA) of NCBI under accession
- 455 references SRP044781-84, SRP045138, SRP045098-103, and SRP045140-146. The
- 456 construction of sequencing libraries, data capture and processing, sequence assembly,
- 457 mapping, and interpretation of read counts were all performed as previously reported with

some modifications [26]. In order to study the transcript expression patterns and levels of *npm2* for each actinopterygian species presenting 2 or 3 *npm2* genes, we mapped the double
stranded RNA-seq reads to the corresponding *npm2* CDS using BWA-Bowtie with stringent
mapping parameters (maximum number of allowed mismatches –aln 2). Mapped reads were
counted using the idxstat command in SAMtools, with a minimum alignment quality value (–
q 30) to discard ambiguous mapping reads. For each species, the number of mapped reads
was then normalized for each *npm2* gene across the 11 tissues using RPKM normalization.

465 Quantitative real-time PCR (QPCR)

466 For each sample, total RNA was extracted using Tri-Reagent (Molecular Research Center,

467 Cincinnati, OH) according to the manufacturer's instructions. Reverse transcription (RT) was

468 performed using 1 μg of RNA from each sample with M-MLV reverse transcriptase and

469 random hexamers (Promega, Madison, WI). Briefly, RNA and dNTP were denatured for

470 6 min at 70°C and then chilled on ice for 5 min before addition of the RT reagents. RT was

471 performed at 37°C for 1 h and 15 min followed by a 15-min incubation step at 70°C. Control

472 reactions were run without reverse transcriptase and used as negative control in the QPCR

473 study. For each studied tissue, cDNA originating from three individual fish were pooled and

474 subsequently used for QPCR. Follicular oocytes at different stages of oogenesis were

475 obtained from at least 3 different wildtype animals, and pooled embryos were obtained from

476 at least 3 clutches from each individual mutant. QPCR experiments were performed with the

477 Fast-SYBR GREEN fluorophore kit (Applied Biosystems, Foster City, CA) as per the

478 manufacturer's instructions using 200 nM of each primer in order to keep PCR efficiency

479 between 90% and 100%, and an Applied Biosystems StepOne Plus instrument. RT products,

480 including control reactions, were diluted 1/25, and 4  $\mu$ l of each sample were used for each

481 PCR. All QPCR were performed in triplicate. The relative abundance of target cDNA was

482 calculated from a standard curve of serially diluted pooled cDNA and normalized to 18S,  $\beta$ -

483 actin,  $\beta$ -2 microglobulin, and EF1 $\alpha$  transcripts. The primer sequences can be found in

484 Supplemental Table S5.

#### 485 Comparison of npm2a and npm2b peptidic sequences

- 486 For the species harbouring Npm2a and Npm2b, or Npm2a, Npm2b1, and Npm2b2, we
- 487 compared the corresponding protein sequences using EMBOSS Matcher tool on the EBI
- 488 website (<u>http://www.ebi.ac.uk/Tools/psa/emboss\_matcher/</u>). Results of the pair-wise
- 489 comparisons are presented as identity percentages in Supplemental Table S2. We also
- 490 estimated the synonymous substitution rates (dS) and non-synonymous substitution rates
- 491 (dN) between the paralogous Npm2a and Npm2b coding sequences (CDS) using JCoDA
- 492 v1.14 software (<u>http://www.tcnj.edu/~nayaklab/jcoda</u>). The alignment options were set to
- 493 "clustal", and the Yang and Nielsen dN/dS substitution models were used.

#### 494 CrispR-cas9 genetic knockout

495 CRISPR/cas9 guide RNA (gRNA) were designed using the ZiFiT online software and were

496 made against 3 targets within each gene to generate large genomic deletions, ranging from

497 250-1600 base pairs, that span exons which allow the formation of non-functional proteins.

498 Nucleotide sequences containing the gRNA were ordered, annealed together, and cloned into

- the DR274 plasmid. *In vitro* transcription of the gRNA from the T7 initiation site was
- 500 performed using the Maxiscript T7 kit (Applied Biosystems), and their purity and integrity
- 501 were assessed using the Agilent RNA 6000 Nano Assay kit and 2100 Bioanalyzer (Agilent
- 502 Technologies, Santa Clara, CA). Zebrafish embryos at the one-cell stage were micro-injected

503 with approximately 30-40 pg of each CRISPR/cas9 guide along with 8-9 nM of purified cas9

- 504 protein (a generous gift from Dr. Anne de Cian from the National Museum of Natural History
- 505 in Paris, France). The embryos were allowed to grow to adulthood, and genotyped using fin
- 506 clip and PCR that detected the deleted regions. The PCR bands of the mutants were then sent
- 507 for sequencing to verify the deletion. Once confirmed, the mutant females were mated with

508	wildtype males to produce	F1 embryos, whose phenotypes	were subsequently recorded.
-----	---------------------------	------------------------------	-----------------------------

- 509 Images were captured with a Carl Zeiss microscope (Jena, Germany) and ToupCam camera
- 510 (ToupTek, Hangzhou, China).

#### 511 Genotyping by PCR

- 512 Fin clips were harvested from animals under anesthesia (0.1% phenoxyethanol) and lysed
- 513 with 5% chelex containing 100  $\mu$ g of proteinase K at 55°C for 2 hrs and then 99°C for 10
- 514 minutes. The extracted DNA was subjected to PCR using Advantage2 system (Clontech,
- 515 Mountain View, CA) for *npm2b* and Jumpstart Taq polymerase (Sigma-Aldrich, St. Louis,
- 516 MO) for *npm2a*. The primers are listed in Supplemental Table S5.

#### 517 Statistical Analysis

- 518 Comparison of two groups was performed using the GraphPad Prism statistical software (La
- 519 Jolla, CA), and either the Student's t-test or Mann-Whitney U-test was conducted depending
- 520 on the normality of the groups based on the Anderson-Darling test.
- 521

Version preprint

522

#### 523 Acknowledgements

- 524 This work was supported by ANR grants PHYLOFISH (ANR-10-GENM-017) and
- 525 Maternal Legacy (ANR-13-BSV7-0015) to JB. Authors would like to thank Dr. Yann Audic
- 526 for providing Xenopus tropicalis animals and Ms. Amélie Patinote for zebrafish rearing and
- 527 egg production.

528

529

#### 530 **References**

- 1. Baroux C, Autran D, Gillmor CS, Grimanelli D, Grossniklaus U. The maternal to
- 532 zygotic transition in animals and plants. Cold Spring Harb Symp Quant Biol. 2008;73:

Lindeman RE, Pelegri F. Vertebrate maternal-effect genes: Insights into fertilization,

533	89–100.

2.

534

535		early cleavage divisions, and germ cell determinant localization from studies in the
536		zebrafish. Mol Reprod Dev. 2010;77: 299–313.
537	3.	Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, et al.
538		Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-
539		zygotic transition. Nature. 2013;503: 360-364.
540	4.	Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, et al.
541		Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs.
542		Science. 2006;312: 75–79.
543	5.	Giraldez AJ. microRNAs, the cell's Nepenthe: clearing the past during the maternal-
544		to-zygotic transition and cellular reprogramming. Curr Opin Genet Dev. 2010; 20:369-
545		375.
546	6.	Lund E, Liu M, Hartley RS, Sheets MD, Dahlberg JE. Deadenylation of maternal
547		mRNAs mediated by miR-427 in Xenopus laevis embryos. Rna. 2009;15: 2351–2363.
548	7.	Bouchareb A, Le Cam A, Montfort J, Gay S, Nguyen T, Bobe J, et al. Genome-wide
549		identification of novel ovarian-predominant miRNAs: new insights from the medaka
550		(Oryzias latipes). Sci Rep. 2017;7: 40241.
551	8.	Bouleau A, Desvignes T, Traverso JM, Nguyen T, Chesnel F, Fauvel C, et al.
552		Maternally inherited npm2 mRNA is crucial for egg developmental competence in
553		zebrafish. Biol Reprod. 2014;91: 43.
554	9.	Laskey RA, Honda BM, Mills AD, Finch JT. Nucleosomes are assembled by an acidic
555		protein which binds histones and transfers them to DNA. Nature. 1978;275: 416-420.
556	10.	Laskey RA, Earnshaw WC. Nucleosome assembly. Nature. 1980;286: 763–767.

557 11. Mills AD, Laskey RA, Black P, De Robertis EM. An acidic protein which assembles

bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

558 nucleosomes in vitro is the most abundant protein in Xenopus oocyte nuclei. J M
---

- 559 Biol. 1980;139: 561–568.
- 560 12. Philpott A, Leno GH, Laskey RA. Sperm decondensation in Xenopus egg cytoplasm is
  561 mediated by nucleoplasmin. Cell. 1991;65: 569–578.
- 562 13. Philpott A, Leno GH. Nucleoplasmin remodels sperm chromatin in Xenopus egg
  563 extracts. Cell. 1992;69: 759–767.
- 14. Burns KH, Viveiros MM, Ren Y, Wang P, DeMayo FJ, Frail DE, et al. Roles of
- 565 NPM2 in chromatin and nucleolar organization in oocytes and embryos. Science.
- 566 2003;300: 633–636.
- 567 15. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral
  568 vertebrate. PLoS Biol. 2005;3: e314.
- 569 16. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome
  570 duplications. Nat Rev Genet. 2009:10: 725–732.
- 571 17. Eirin-Lopez JM, Frehlick LJ, Ausio J. Long-term evolution and functional
- 572 diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear
- 573 chaperones. Genetics. 2006;173: 1835–1850.
- 574 18. Frehlick LJ, Eirin-Lopez JM, Jeffery ED, Hunt DF, Ausio J. The characterization of
- amphibian nucleoplasmins yields new insight into their role in sperm chromatin
- 576 remodeling. BMC Genomics. 2006;7: 99.
- Wotton KR, Weierud FK, Dietrich S, Lewis KE. Comparative genomics of Lbx loci
  reveals conservation of identical Lbx ohnologs in bony vertebrates. BMC Evol Biol.
- 579 2008;8: 171.
- 580 20. Jovelin R, Yan YL, He X, Catchen J, Amores A, Canestro C, et al. Evolution of
- 581 developmental regulation in the vertebrate FgfD subfamily. J Exp Zool B Mol Dev
- 582 Evol. 2010;314: 33–56.

583	21.	Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its
584		evolutionary consequences. Mol Genet Genomics. 2014;289: 1045-1060.
585	22.	Volff JN. Genome evolution and biodiversity in teleost fish. Heredity (Edinb).
586		2005;94: 280–294.
587	23.	Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, et al. The
588		rainbow trout genome provides novel insights into evolution after whole-genome
589		duplication in vertebrates. Nat Commun. 2014;5: 3657.
590	24.	Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome
591		evolution in the allotetraploid frog Xenopus laevis. Nature. 2016;538: 336-343.
592	25.	Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, et al. Resolution of
593		ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci U S A.
594		2012;109: 13698–13703.
595	26.	Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al. Gene
596		evolution and gene expression after whole genome duplication in fish: the PhyloFish
597		database. BMC Genomics. 2016;17: 368.
598	27.	Lingenfelter BM, Tripurani SK, Tejomurtula J, Smith GW, Yao J. Molecular cloning
599		and expression of bovine nucleoplasmin 2 (NPM2): a maternal effect gene regulated
600		by miR-181a. Reprod Biol Endocrinol. 2011;9: 40.
601	28.	Burglin TR, Mattaj IW, Newmeyer DD, Zeller R, De Robertis EM. Cloning of
602		nucleoplasmin from Xenopus laevis oocytes and analysis of its developmental
603		expression. Genes Dev. 1987;1: 97–107.
604	29.	Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, Stemple D, et al.
605		Identification of the zebrafish maternal and paternal transcriptomes. Development.
606		2013;140: 2703–2710.
607	30.	Laskey RA, Mills AD, Philpott A, Leno GH, Dilworth SM, Dingwall C. The role of

bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

608		nucleoplasmin in chromatin assembly and disassembly. Philos Trans R Soc Lond B
609		Biol Sci. 1993;339: 263–269.
610	31.	Prado A, Ramos I, Frehlick LJ, Muga A, Ausio J. Nucleoplasmin: a nuclear chaperone.
611		Biochem Cell Biol. 2004;82: 437–445.
612	32.	De La Fuente R, Viveiros MM, Burns KH, Adashi EY, Matzuk MM, Eppig JJ. Major
613		chromatin remodeling in the germinal vesicle (GV) of mammalian oocytes is
614		dispensable for global transcriptional silencing but required for centromeric
615		heterochromatin function. Dev Biol. 2004;275: 447-458.
616	33.	Lee HH, Kim HS, Kang JY, Lee BI, Ha JY, Yoon HJ, et al. Crystal structure of human
617		nucleophosmin-core reveals plasticity of the pentamer-pentamer interface. Proteins.
618		2007;69: 672–678.
619	34.	Platonova O, Akey I V, Head JF, Akey CW. Crystal structure and function of human
620		nucleoplasmin (npm2): a histone chaperone in oocytes and embryos. Biochemistry.
621		2011;50: 8078–8089.
622	35.	Okuwaki M, Sumi A, Hisaoka M, Saotome-Nakamura A, Akashi S, Nishimura Y, et
623		al. Function of homo- and hetero-oligomers of human nucleoplasmin/nucleophosmin
624		family proteins NPM1, NPM2 and NPM3 during sperm chromatin remodeling.
625		Nucleic Acids Res. 2012;40: 4861–4878.
626	36.	Onikubo T, Nicklay JJ, Xing L, Warren C, Anson B, Wang WL, et al.
627		Developmentally Regulated Post-translational Modification of Nucleoplasmin
628		Controls Histone Sequestration and Deposition. Cell Rep. 2015. pii:S2211-
629		1247(15)00196-5. doi: 10.1016/j.celrep.2015.02.038.
630	37.	Dingwall C, Dilworth SM, Black SJ, Kearsey SE, Cox LS, Laskey RA. Nucleoplasmin
631		cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences
632		homologous to putative nuclear localization signals. Embo J. 1987;6: 69–74.

633	38.	Dutta S. Akev I	V. Dingwall C	. Hartman KL	. Laue T.	Nolte RT	. et al. The	crystal
			· • • • • • • • • • • • • • • •	,	,		,	

- structure of nucleoplasmin-core: implications for histone binding and nucleosome
  assembly. Mol Cell. 2001;8: 841–853.
- 636 39. Banuelos S, Hierro A, Arizmendi JM, Montoya G, Prado A, Muga A. Activation
- 637 mechanism of the nuclear chaperone nucleoplasmin: role of the core domain. J Mol
  638 Biol. 2003;334: 585–593.
- 40. Salvany L, Chiva M, Arnan C, Ausio J, Subirana JA, Saperas N. Mutation of the small
  acidic tract A1 drastically reduces nucleoplasmin activity. FEBS Lett. 2004;576: 353–
  357.
- 41. Taneva SG, Banuelos S, Falces J, Arregi I, Muga A, Konarev P V, et al. A mechanism
  for histone chaperoning activity of nucleoplasmin: thermodynamic and structural
  models. J Mol Biol. 2009;393: 448–463.
- 42. Fernandez-Rivero N, Franco A, Velazquez-Campoy A, Alonso E, Muga A, Prado A. A

646 Quantitative Characterization of Nucleoplasmin/Histone Complexes Reveals

- 647 Chaperone Versatility. Sci Rep. 2016;6: 32114.
- 43. Namboodiri VM, Akey IV, Schmidt-Zachmann MS, Head JF, Akey CW. The
- 649 structure and function of Xenopus NO38-core, a histone chaperone in the nucleolus.
- 650 Structure. 2004;12: 2149–2160.
- 44. Ramos I, Fernandez-Rivero N, Arranz R, Aloria K, Finn R, Arizmendi JM, et al. The
- 652 intrinsically disordered distal face of nucleoplasmin recognizes distinct
- oligomerization states of histones. Nucleic Acids Res. 2014;42: 1311–1325.
- 45. Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. Highly efficient
- 655 CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair.
- 656 Genome Res. 2014;24: 142–153.
- 46. Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, et al. Efficient

27

658	mutagenesis by (	Cas9 prot	ein-mediated	oligonucl	eotide i	insertion a	and large-s	cale
	0 1			0			0	

- assessment of single-guide RNAs. PLoS One. 2014;9: e98186.
- 660 47. Li L, Baibakov B, Dean J. A subcortical maternal complex essential for

preimplantation mouse embryogenesis. Dev Cell. 2008;15: 416–425.

- 48. Xu Q, Wang F, Xiang Y, Zhang X, Zhao ZA, Gao Z, et al. Maternal BCAS2 protects
- genomic integrity in mouse early embryonic development. Development. 2015;142:

664 <u>3943–3953</u>.

- 665 49. Dekens MP, Pelegri FJ, Maischein HM, Nusslein-Volhard C. The maternal-effect gene
- futile cycle is essential for pronuclear congression and mitotic spindle assembly in the
  zebrafish zygote. Development. 2003;130: 3907–3916.
- 50. Yabe T, Ge X, Pelegri F. The zebrafish maternal-effect gene cellular atoll encodes the
  centriolar component sas-6 and defects in its paternal function promote whole genome
  duplication. Dev Biol. 2007;312: 44–60.
- Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. Stages of embryonic
  development of the zebrafish. Dev Dyn. 1995;203: 253–310.
- 674

673

675 Figure legends

#### 676 Figure 1: Consensus phylogenetic tree of Npm2 proteins.

- 677 This phylogenetic tree was constructed based on the amino acid sequences of Npm2 proteins
- 678 (for the references of each sequence see Supplemental Table S1) using the Neighbour Joining
- 679 method with 1,000 bootstrap replicates. The number shown at each branch node indicates the
- bootstrap value (%). The tree was rooted using Npm3 and Npm1 sequences. The Npm2a
- sequences are in blue, the Npm2b sequences are in red, the salmonid Npm2b1 sequences are
- in purple, and the salmonid Npm2b2 sequences are in pink.

Version preprint

```
28
```

#### 683 Figure 2: Conserved genomic synteny of *npm2* genes.

684 Genomic synteny maps comparing the orthologs of *npm2a*, *npm2b*, and their neighbouring

685 genes. *npm2* genes are named as *npm2a* and *npm2b* (formerly known as *npm2*). The other

- 686 genes were named after their human orthologs according to the Human Genome Naming
- 687 Consortium (HGNC). Orthologs of each gene are shown in the same color. The direction of
- arrows indicates the gene orientation, with the ID of the genomic segment indicated above
- and the position of the gene (in  $10^{-6}$  base pairs) indicated below. The full gene names and
- 690 detailed genomic locations are given in Supplemental Table S3.

691 Figure 3: Current status and proposed evolutionary history of *npm2* genes among

- 692 gnathostomes.
- 693 The names of the current representative species of each phylum are given at the end of the
- final branches, together with a red and/or blue X to denote the *npm2* genes they possess

695 (*npm2a*=red, *npm2b*=blue). The black X upon an *npm2* gene symbol indicates a gene loss.

696 LGD: local gene duplication; TGD: teleost-specific whole genome duplication; SaGD:

697 salmonid-specific whole genome duplication.

#### **Figure 4: Tissue distribution of** *npm2a* **and** *npm2b* **in different species.**

- 699 A) Tissue expression analysis by QPCR of *npm2a* and *npm2b* mRNAs in zebrafish and
- 700 Xenopus. (B) Expression level is expressed as a percentage of the expression in the ovary for
- the most expressed gene. Data were normalized using 18S expression. (C-J) Tissue
- 702 expression level by RNA-Seq of npm2a and npm2b mRNAs in different fish species. mRNA
- reads (RPKM). In salmonids (I and J),
- the two 4R isoforms of *npm2b* are *npm2b1* and *npm2b2*. Br, brain ; Gi, gills ; Lu, lung ; In,
- intestine ; Li, liver ; Mu, muscle ; He, heart ; Bo, bone ; Ki, kidney ; Ov, ovary ; and Te,
- 706 testis.

#### 707 Figure 5: npm2a and npm2b expression during oogenesis and early development.

708	QPCR analysis of (A) npm2a and (B) npm2b expression during oogenesis and early
709	embryonic development in zebrafish. Data were normalized using luciferase, and relative
710	expression was based on npm2b expression at the indicated stage. UF, unfertilized egg; hpf,
711	hours post-fertilization.
712	Figure 6: CRISPR/cas9 knockout of <i>npm2a</i> and <i>npm2b</i> in zebrafish.
713	(A) Relative expression level of <i>npm2a</i> and <i>npm2b</i> transcripts by QPCR in the fertilized
714	zebrafish eggs from crosses between npm2a or npm2b mutant female and wildtype (WT)
715	male, respectively. (B) Developmental success as measured by the proportion of fertilized
716	eggs that underwent normal cell division and reached normal developmental milestones
717	based on Kimmel et al. [51] from crosses between WT animals (control), and <i>npm2a</i> or
718	npm2b mutant female and WT male at 1 and 4 hours post-fertilization (hpf). Shaded grey
719	columns denote normal morphology and black columns denote abnormal morphology of the
720	fertilized eggs at 1 hpf. (C) Representative images demonstrating development of fertilized
721	eggs from crosses between control animals (left panels), and npm2a (middle panels) or
722	npm2b (right panels) mutant female and WT male at 64-cell, oblong, dome, germ ring, and
723	24-somite stages according to Kimmel et al [51]. N=3 for npm2a mutant and N=4 for npm2b
724	mutant. All assessments were performed from at least 3 clutches from each mutant. Scale
725	bars denote 500 nm. Red dotted lines define the diameter of the embryo. ** $p<0.01$ ,
726	***p<0.001, ****p<<0.0001
727	Supplemental Figure S1: Npm2a and Npm2b protein domain conservation.
728	Presence of the Npm core (red color bar) and acidic tract (blue color bar) domains in <i>npm2a</i>
729	and <i>npm2b</i> types in various vertebrate species.
730	Supplemental Figure S2: Genotyping by PCR for <i>npm2a</i> and <i>npm2b</i> in F1 embryos.
731	npm2a and npm2b knockout animals were crossed with wildtype (WT) zebrafish, and

732 individual F1 embryos were harvested at 6 hours post-fertilization and subjected to

- genotyping by PCR with gene-specific primers. The wildtype PCR product for *npm2a* was
- 1634 base pairs (bp) and the CRISPR/cas9-targeted knockout mutant band was 240 bp while
- the WT *npm2b* band was 813 bp and the CRISPR/cas9-targeted knockout mutant band was
- 560 bp. The yellow asterisks denote the embryos carrying the mutant allele.

#### 737 Supplementary Data

- 738 Supplemental Table S1: list of Npm2 protein sequences used in the phylogenetic analysis.
- 739 **Supplemental Table S2**: Npm2a and Npm2b protein sequence comparisons.
- 740 **Supplemental Table S3**: list of the genes from the conserved syntenic region of *npm2* genes.
- 741 **Supplemental Table S4:** Estimation of the ratio of substitution rates (dN/dS) in *npm2a* and
- 742 npm2b genes.
- 743 **Supplemental Table S5**: QPCR and PCR primers.
- 744 Supplemental Fig.S1: Npm2a and Npm2b protein domain conservation.
- 745 **Supplemental Fig.S2:** Genotyping of F1 embryos from *npm2a* and *npm2b* mutants.

31

bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.



# Figure 2







## Figure 4



bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

### Figure 5



bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.



### Supplementing of the start of t



bioRxiv preprint first posted online May. 22, 2017; doi: http://dx.doi.org/10.1101/104760. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

## **Supplemental Figures**

