



**HAL**  
open science

# Comparative Transcriptomics Analyses across Species, Organs, and Developmental Stages Reveal Functionally Constrained lncRNAs

Fabrice Darbellay, Anamaria Necsulea

► **To cite this version:**

Fabrice Darbellay, Anamaria Necsulea. Comparative Transcriptomics Analyses across Species, Organs, and Developmental Stages Reveal Functionally Constrained lncRNAs. *Molecular Biology and Evolution*, 2019, 10.1093/molbev/msz212 . hal-02400709

**HAL Id: hal-02400709**

**<https://cnrs.hal.science/hal-02400709v1>**

Submitted on 9 Dec 2019

**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.

**Comparative transcriptomics analyses across species, organs and developmental stages  
reveal functionally constrained lncRNAs**

Fabrice Darbellay<sup>1,§,\*</sup> and Anamaria Necsulea<sup>1,2,\*</sup>

<sup>1</sup>School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne,  
Switzerland

<sup>2</sup>Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive  
UMR 5558, F-69622 Villeurbanne, France

<sup>§</sup>Present address: Environmental Genomics and Systems Biology Division, Lawrence Berkeley  
National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.

\*Corresponding authors:

Fabrice Darbellay (fdarbellay@lbl.gov)

Anamaria Necsulea (anamaria.necsulea@univ-lyon1.fr)

Running title: Functionally constrained lncRNAs in embryonic development

Keywords: long non-coding RNAs; evolution; development; comparative transcriptomics.

27           **Abstract**

28           The functionality of long non-coding RNAs (lncRNAs) is disputed. In general, lncRNAs are under  
29 weak selective pressures, suggesting that the majority of lncRNAs may be non-functional. However,  
30 although some surveys showed negligible phenotypic effects upon lncRNA perturbation, key biological  
31 roles were demonstrated for individual lncRNAs. Most lncRNAs with proven functions were implicated  
32 in gene expression regulation, in pathways related to cellular pluripotency, differentiation and organ  
33 morphogenesis, suggesting that functional lncRNAs may be more abundant in embryonic  
34 development, rather than in adult organs. To test this hypothesis, we perform a multi-dimensional  
35 comparative transcriptomics analysis, across five developmental time-points (two embryonic stages,  
36 newborn, adult and aged individuals), four organs (brain, kidney, liver and testes) and three species  
37 (mouse, rat and chicken). We find that, overwhelmingly, lncRNAs are preferentially expressed in adult  
38 and aged testes, consistent with the presence of permissive transcription during spermatogenesis.  
39 lncRNAs are often differentially expressed among developmental stages and are less abundant in  
40 embryos and newborns compared to adult individuals, in agreement with a requirement for tighter  
41 expression control and less tolerance for noisy transcription early in development. For differentially  
42 expressed lncRNAs, we find that the patterns of expression variation among developmental stages are  
43 generally conserved between mouse and rat. Moreover, lncRNAs expressed above noise levels in  
44 somatic organs and during development show higher evolutionary conservation, in particular at their  
45 promoter regions. Thus, we show that functionally constrained lncRNA loci are enriched in developing  
46 organs, and we suggest that many of these loci may function in an RNA-independent manner.

47           **Introduction**

48           Long non-coding RNAs (lncRNAs, loosely defined as transcripts without protein-coding  
49 potential, at least 200 nucleotides long) are an excellent illustration of the ongoing conceptual tug-of-  
50 war between biochemical activity and biological function (Graur et al. 2013; Doolittle 2018). Recent  
51 sequencing-based studies identified thousands of lncRNAs in vertebrates (Guttman et al. 2009; Khalil  
52 et al. 2009; Iyer et al. 2015; Pertea et al. 2018). While this class of transcripts includes lncRNAs with  
53 undisputed biological roles, such as *Xist* (Brown et al. 1991) or *H19* (Brannan et al. 1990), experimental  
54 validations are lacking for the great majority of lncRNAs and their functionality is controversial.

55           The first functional characterizations of individual lncRNAs forged the idea that they are  
56 important contributors to gene expression regulatory networks. This has been unequivocally proven  
57 for some lncRNAs, such as *Xist*, whose transcription and subsequent coating of the X chromosome  
58 triggers a complex chain of molecular events leading to X inactivation in placental mammals  
59 (Munschauer et al. 2018). Other proposed mechanisms for gene expression regulation by lncRNAs  
60 include directing chromatin-modifying complexes at specific genomic locations, to control gene

61 expression in *trans* (Rinn et al. 2007); providing decoy targets for microRNAs (Cesana et al. 2011);  
62 enhancing neighboring gene expression through an RNA-dependent mechanism (Ørom et al. 2010).  
63 Biological functions unrelated to gene expression regulation were also proposed for lncRNAs. For  
64 example, the NORAD lncRNA was shown to assemble a topoisomerase complex critical for genome  
65 stability (Munschauer et al. 2018), while the X-linked Firre lncRNA is involved in chromatin super-loop  
66 formation on the inactive X chromosome (Hacisuleyman et al. 2014; Barutcu et al. 2018). Additional  
67 evidence that individual lncRNAs are undoubtedly biologically relevant comes from associations with  
68 human diseases, including cancer, as for the SAMMSON lncRNA (Vendramin et al. 2018).

69 Initial studies of lncRNA functionality generally asserted that biological functions are directly  
70 carried out by the transcribed RNA molecules. For some lncRNAs, this hypothesis was supported by  
71 thorough functional tests, including rescue experiments showing that phenotypic effects of lncRNA  
72 locus deletion can be reversed by expressing the lncRNAs *in trans* (Munschauer et al. 2018). Thus, for  
73 a subset of lncRNA loci, their biological function is undoubtedly achieved by the non-coding RNA  
74 molecule. However, in some cases, lncRNA function resides in the act of transcription at a given  
75 genomic location, rather than in the product of transcription (Latos et al. 2012). In other cases,  
76 biological functions are carried out by other elements embedded in the lncRNA genomic loci (Bassett  
77 et al. 2014). For example, transcription of *Linc-p21*, originally described as a *cis*-acting enhancer  
78 lncRNA, is not needed to regulate neighboring gene expression, which is instead controlled by multiple  
79 enhancer elements within the locus (Groff et al. 2016). Genetic engineering of multiple lncRNA loci in  
80 mouse likewise indicated that lncRNA transcripts are dispensable, and that gene expression regulation  
81 by lncRNA loci is instead achieved by the process of transcription and splicing, or by additional  
82 regulatory elements found in lncRNA promoters (Engreitz et al. 2016; Anderson et al. 2016).  
83 Furthermore, some attempts to look for lncRNA function through genetic engineering approaches  
84 showed that the tested lncRNA loci are altogether dispensable (Amândio et al. 2016; Zakany et al.  
85 2017; Goudarzi et al. 2019). These recent observations signal a paradigm shift in lncRNA biology, as it  
86 is increasingly acknowledged that, even when phenotypic effects can be unambiguously mapped to  
87 lncRNA loci, they may not be driven by the lncRNA transcripts themselves.

88 Importantly, this new perspective on lncRNA biology had been predicted by evolutionary  
89 analyses, traditionally used to evaluate the functionality of diverse genomic elements (Haerty and  
90 Ponting 2014; Ulitsky 2016). Evolutionary studies of lncRNAs in vertebrates agree that selective  
91 constraint on lncRNA primary sequences is weak, though significantly above the genomic background  
92 (Ponjavic et al. 2007; Kutter et al. 2012; Necsulea et al. 2014; Washietl et al. 2014; Hezroni et al. 2015).  
93 These observations are compatible with the hypothesis that many of the lncRNAs detected with  
94 sensitive transcriptomics techniques may be non-functional noise (Ponjavic et al. 2007), or that their  
95 function is carried out by small conserved elements, such that the selective constraint signal on the

106 entire lncRNA locus is overall weak (Ulitsky 2016). They also indicate that lncRNA functionality may not  
107 reside in the primary transcribed sequence. Indeed, mammalian lncRNA promoters show higher levels  
108 of sequence conservation, similar to protein-coding gene promoters (Necsulea et al. 2014), as  
109 expected if they carry out additional regulatory functions independently of the transcribed RNA  
110 molecule. Moreover, it was previously reported that, in multi-exonic lncRNAs, splicing regulatory  
111 elements are more conserved than exonic sequences (Schüler et al. 2014; Haerty and Ponting 2015),  
112 in agreement with the recent finding that lncRNA splicing can contribute to neighboring gene  
113 regulation (Engreitz et al. 2016). Thus, detailed evolutionary analyses of lncRNA loci can bring insights  
114 into their functionality, and can help prioritize candidates for experimental validation.

115 Most comparative lncRNA studies were so far restricted to adult organ transcriptomes. These  
116 comparisons showed that lncRNAs are preferentially expressed in adult testes, during  
117 spermatogenesis (Soumillon et al. 2013). This process is characterized by a permissive chromatin  
118 environment, which can promote non-functional transcription (Soumillon et al. 2013). The resulting  
119 lncRNA datasets may thus be enriched in non-functional transcripts. Additional lines of evidence  
120 suggest that the search for functional lncRNAs should be extended beyond adult organ transcriptomes.  
121 For example, involvement in developmental phenotypes was proposed for many experimentally-  
122 tested lncRNAs (Sauvageau et al. 2013; Ulitsky et al. 2011; Grote et al. 2013), and an enrichment for  
123 developmental transcription factor binding was reported for the promoters of conserved lncRNAs  
124 (Necsulea et al. 2014). These observations motivated us to add a temporal dimension to comparative  
125 lncRNA transcriptomics studies. Therefore, here we characterize lncRNAs across species, organs and  
126 developmental stages. We analyze the spatial and temporal expression patterns of protein-coding  
127 genes and lncRNAs, in conjunction with their evolutionary conservation. We find that, while lncRNAs  
128 are overall poorly conserved during evolution in terms of primary sequence or expression patterns,  
129 higher frequencies of constrained lncRNAs are observed in embryonic transcriptomes. For many of  
130 these loci, biological function may be RNA-independent, as the highest levels of sequence conservation  
131 are observed on promoter regions and on splice signals, rather than on lncRNA exonic sequence. Our  
132 results are thus compatible with unconventional, RNA-independent functions for evolutionarily  
133 conserved lncRNA loci, in particular for those that are expressed during embryonic development.

## 134 **Results**

### 135 Comparative transcriptomics across species, organs and developmental stages

136 To study protein-coding and lncRNA expression patterns across both developmental and  
137 evolutionary time, we generated strand-specific RNA-seq data for mouse and rat, for four major organs  
138 (brain, kidney, liver and testes) and five developmental time points, including two embryonic stages,  
139 newborn, young and aged adult individuals (Figure 1A, Supplementary Table 1, Materials and

130 methods). The selected time points allow us to obtain a broad view of major organ ontogenesis and to  
131 capture drastic physiological changes during development (Theiler 1989). We chose to include in our  
132 study both young adult (8-10 weeks old) and aged adult individuals (12 to 24 months old), thus  
133 completing our overview of temporal patterns of gene expression variation. At the earliest embryonic  
134 stage (day 13.5 post-conception for mouse, day 15 for rat), we dissected only the three somatic organs.  
135 Our experimental design for mouse and rat thus comprises 19 organ / developmental stage  
136 combinations. To obtain a broader evolutionary perspective we generated comparable RNA-seq data  
137 for the chicken, for the two earliest developmental stages (Figure 1A, Supplementary Table 1). We  
138 generated between 2 and 4 biological replicates for each species/organ/developmental stage  
139 combination (Supplementary Table 1). Additional RNA-seq samples from previous publications were  
140 included in the lncRNA annotation process, to increase detection sensitivity (Supplementary Table 2).

141 The organs and developmental stages included in our study differ greatly in terms of their  
142 cellular composition diversity. To verify that our whole-organ RNA-seq data reflects cellular  
143 composition heterogeneity, we assessed the expression patterns of cell population markers derived  
144 from single-cell transcriptomics studies (Tabula Muris Consortium 2018; Green et al. 2018) in our  
145 samples (Supplementary Figure 1, Supplementary Table 3, Supplementary Methods). This analysis  
146 confirms that our transcriptome collection reflects expected developmental patterns. For example,  
147 mature oligodendrocyte cell markers are systematically highly expressed in adult brain, while  
148 oligodendrocyte precursor markers are more highly expressed in the earliest developmental stages  
149 (Supplementary Figure 1). To further characterize our transcriptome collection, we sought to identify  
150 genes that could serve as markers for organ/developmental stage combinations. To do this, we  
151 selected genes that have narrow expression distributions, and for which maximum expression is  
152 observed in the same organ/developmental stage combination in mouse and rat (Supplementary  
153 Methods, Supplementary Table 4). Gene ontology enrichment analyses for these lists of genes are  
154 coherent with the cellular composition and biological processes at work. Thus, genes involved in  
155 forebrain neuron differentiation are over-represented in the mid-stage embryonic brain, while  
156 processes related to synaptic transmission are enriched among genes specifically expressed in adult  
157 brain (Supplementary Figure 2). In the kidney, the early developmental stages are enriched in genes  
158 involved in metanephric development (Supplementary Figure 3). The newborn liver stands out due to  
159 its strong enrichment in genes involved in immune response, while metabolic processes are over-  
160 represented in the adult liver (Supplementary Figure 4). Embryonic testes samples express genes  
161 implicated in gamete generation and gene silencing by miRNAs, including the *Piwi*-like genes, while  
162 adult testes transcriptomes are dominated by genes involved in spermatogenesis (Supplementary  
163 Figure 5). These patterns confirm that our whole-organ transcriptome collection captures the cell  
164 composition changes and physiological transitions that occur during organ development.

165 Developmental expression patterns are well conserved among species for protein-coding genes

166 To gain a first glimpse into the evolution of developmental gene expression patterns, we  
167 performed a principal component analysis (PCA) for 10,363 protein-coding genes shared among  
168 mouse, rat and chicken (Figure 1B, Supplementary Methods). This analysis revealed that the main  
169 source of gene expression variability among species, organs and developmental stages is the  
170 distinction between adult and aged testes and the other samples, which are separated on the first PCA  
171 axis (Figure 1B). In contrast, embryonic and newborn testes are grouped with kidney samples from  
172 similar developmental stages, in agreement with the common developmental origin of the kidney and  
173 the gonads (Nel-Themaat et al. 2010). The first axis of the PCA, which explains 67% of the total  
174 expression variance, also correlates with the developmental stage for the brain: samples derived from  
175 adult and aged individuals have higher coordinates on this axis than embryonic and newborn samples  
176 (Figure 1B, Supplementary Figure 6, Kruskal-Wallis test p-value 0.003). The second PCA axis (10%  
177 explained variability) mainly reflects the difference between brain and the other organs, but is also  
178 associated with the developmental stage for kidney and liver (Figure 1B, Supplementary Figure 6,  
179 Kruskal-Wallis test p-value  $4e^{-4}$  for kidney,  $4e^{-5}$  for liver). However, we note that the association  
180 between PC2 and developmental stages for kidney and liver may be confounded by differences in RNA  
181 among developmental stages for these organs (Supplementary Methods, Supplementary Figure 6).

182 While mouse and rat samples are almost undistinguishable on the PCA factorial map, there is  
183 considerably higher expression divergence between chicken and the two rodent species (Figure 1B).  
184 However, differences among major organs are stronger than differences among species, even at these  
185 broad evolutionary distances: brain samples all cluster together, irrespective of the species of origin,  
186 and are clearly separated from kidney and liver samples on the second PCA axis (Figure 1B). These  
187 patterns of gene expression variations are confirmed by a hierarchical clustering analysis based on  
188 Spearman's correlation coefficients between pairs of samples (Figure 1C). The strongest clustering is  
189 observed for adult and aged testes samples, followed by brain samples (Figure 1C).

190 The grouping among samples derived from similar organs and developmental stages,  
191 irrespective of the species of origin, is stronger for genes that are associated with embryonic  
192 development and with gene expression regulation (Supplementary Methods, Supplementary Figure  
193 6C, D). For this set of genes, both the principal component analysis and the hierarchical clustering  
194 analysis show a near-perfect separation of organs and developmental stages, for all three species  
195 (Supplementary Figure 6C, D). Chicken samples, which cluster apart from rodent samples in whole  
196 transcriptome analyses, are now grouped with the corresponding organs and developmental stages  
197 from mouse and rat. Our transcriptome collection can thus reveal highly conserved expression  
198 patterns for regulators of embryonic development, across amniotes.

199 Variations in transcriptome complexity among organs and developmental stages

200 We next sought to assess the transcriptome complexity in different organs across  
201 developmental stages. To predict lncRNAs, we used the RNA-seq data to reconstruct gene models with  
202 StringTie (Pertea et al. 2015), building on existing genomic annotations (Cunningham et al. 2019). We  
203 verified the protein-coding potential of newly annotated transcripts, based on the codon substitution  
204 frequency score (Lin et al. 2007, 2011) and on sequence similarity with known proteins, and we applied  
205 a stringent series of filters to reduce contaminations from un-annotated protein-coding UTRs and  
206 other artefacts (Materials and methods). We thus obtain a total of 18858 candidate lncRNAs in the  
207 mouse, 20159 in the rat and 5496 in the chicken, including both newly-annotated and previously  
208 known lncRNAs transcribed in our samples (Supplementary Dataset 1). The relative sizes of each  
209 species' lncRNA repertoires are consistent with previous studies (Necsulea et al. 2014; Sarropoulos et  
210 al. 2019). We note however that our power to detect lncRNAs in chicken is limited, due to the narrower  
211 organ and developmental stage sampling in this species (Supplementary Tables 1,2). Most candidate  
212 lncRNAs are expressed at very low levels. When imposing a minimum normalized expression level  
213 (transcript per million, or TPM) of 1, in at least one sample, the numbers of candidate lncRNAs falls to  
214 12199, 15319 and 2892 in the mouse, rat and chicken, respectively (Supplementary Datasets 2, 3).

215 The differences in lncRNA content among species may be affected by RNA-seq read coverage  
216 and sample distribution, as well as genome sequence and annotation quality. To correct for the effect  
217 of RNA-seq read coverage, we down-sampled the RNA-seq data to obtain the same number of uniquely  
218 mapped reads for each organ/developmental stage combination within each species (Supplementary  
219 Methods). After this procedure, the number of detectable protein-coding genes (supported by at least  
220 10 uniquely mapped reads) still shows broad variations among organs and developmental stages, with  
221 the highest numbers of genes detected in the testes, for all time points (Figure 2A). Large numbers of  
222 protein-coding genes (between approximately 12800 and 16700) are detected in all samples. In  
223 contrast, for lncRNAs, the pattern is much more striking: the young and aged adult testes express  
224 between 11000 and 12000 lncRNAs, in both mouse and rat, while in somatic organs and earlier  
225 developmental stages we can detect only between 1800 and 4800 lncRNAs (Figure 2B). This  
226 observation is in agreement with previous findings indicating that during spermatogenesis the  
227 chromatin environment is highly permissive to transcription (Soumillon et al. 2013).

228 Spatial and temporal expression patterns for protein-coding genes and lncRNAs

229 We next compared spatial and temporal expression patterns between protein-coding genes  
230 and lncRNAs. In agreement with previous findings (Soumillon et al. 2013), we show that lncRNAs are  
231 overwhelmingly preferentially expressed in the testes (Figure 3A). Indeed, more than 60% of lncRNAs  
232 reach their maximum expression level in this organ, compared to less than 35% of protein-coding



233 genes, for both mouse and rat (Figure 3A; Chi-square test  $p$ -value  $<1e^{-10}$ ). Almost 80% of lncRNAs are  
234 preferentially expressed in young or aged adult samples, which is significantly higher than the fraction  
235 observed for protein-coding genes (less than 65%, Chi-square test  $p$ -value  $<1e^{-10}$ , Figure 3B).

236 We found that between 57% and 80% of protein-coding genes are significantly differentially  
237 expressed (DE) among developmental stages, at a false discovery rate (FDR) below 1%, in each organ  
238 and species (Figure 3C, Supplementary Dataset 4). The proportions of DE lncRNAs are significantly  
239 lower than the proportions of DE protein-coding genes in somatic organs, between 17% and 41% (Chi-  
240 square test,  $p$ -value  $<1e^{-10}$ ). In the testes, we observed higher proportions of DE lncRNAs (63% in  
241 mouse and 67% in rat), but these values were still significantly lower than those observed for protein-  
242 coding genes (77% in mouse and 79% in rat; Chi-square test,  $p$ -value  $<1e^{-10}$ ; Figure 3C). We suspected  
243 that the lower proportion of DE lncRNAs could be due to their low expression levels, as total read  
244 counts affect the sensitivity of DE tests (Anders and Huber 2010). Indeed, lncRNAs are expressed at  
245 much lower levels than protein-coding genes (Supplementary Figure 7). To control for this, we down-  
246 sampled the read counts observed for protein-coding genes, bringing them to the same average counts  
247 as lncRNAs but preserving relative gene abundance (Materials and methods). Strikingly, after down-  
248 sampling, we observe higher proportions of DE loci for lncRNAs compared to protein-coding genes  
249 (Figure 3C). The differences are statistically significant (Chi-square test,  $p$ -value  $<1e^{-10}$ ) in all but one  
250 species / organ combination (mouse kidney, Chi-square test,  $p$ -value 0.15). We also observed that the  
251 expression amplitude among developmental stages are more important for lncRNAs than for protein-  
252 coding genes (Wilcoxon test,  $p$ -value  $<1e^{-10}$ , Supplementary Figure 8A), as expected given the lower  
253 lncRNA expression levels, which preclude detecting subtle expression shifts among time points. Finally,  
254 we observe that the developmental stage with maximum expression is generally different between  
255 protein-coding genes and lncRNAs, even when considering genes that are significantly DE among  
256 stages. For all organs, DE lncRNAs tend to show highest expression levels in the young and aged adults,  
257 while DE protein-coding genes are more homogeneously distributed among developmental stages  
258 (Chi-square test,  $p$ -value  $<1e^{-10}$ , Figure 3D, Supplementary Figure 8B).

259 Similar conclusions are reached when performing DE analyses between consecutive time  
260 points (Supplementary Figure 9, Supplementary Dataset 4). For both protein-coding genes and  
261 lncRNAs, the strongest expression changes are observed between newborn and young adult  
262 individuals. Almost 10000 lncRNAs are significantly up-regulated between newborn and young adult  
263 testes, confirming the strong enrichment for lncRNAs during spermatogenesis (Supplementary Figure  
264 9). As expected, the lowest numbers of DE genes are observed at the transition between young and  
265 aged adult organs. At this time point, we observe more changes for the rat than for the mouse,  
266 potentially due to a higher proportion of immune cell infiltrates in the rat aged organ samples. Genes

267 associated with antigen processing and presentation tend to be expressed at higher levels in aged  
268 adults than in young adults, for mouse kidney, rat brain and liver (Supplementary Dataset 4).

269 *Stronger selective constraint on lncRNAs expressed earlier in development*

270 We next analyzed the long-term evolutionary sequence conservation for lncRNAs, in  
271 conjunction with their spatio-temporal expression patterns (Supplementary Table 5). We used the  
272 PhastCons score (Siepel et al. 2005) for placental mammals (Casper et al. 2018), to assess sequence  
273 conservation for various aspects of mouse lncRNAs: exons, promoters (defined as 400 bp regions  
274 upstream of the transcription start site), splice sites (first and last two bases of the introns). As  
275 approximately 20% of lncRNAs overlap with exonic regions on the opposite strand (Supplementary  
276 Dataset 1), we masked exonic regions from other genes before evaluating sequence conservation.

277 As previously observed (Ponjavic et al. 2007; Haerty and Ponting 2013), exonic and splice site  
278 sequence conservation is much lower for lncRNAs (median exonic score 0.094, median splice site score  
279 0.075) than for protein-coding genes (median exonic score 0.42, median splice site score 0.85,  
280 Wilcoxon test p-value  $< 1e^{-10}$ , Supplementary Figure 10). Exonic lncRNA conservation scores are  
281 significantly above the conservation observed for intergenic regions genome-wide (median score  
282 0.076, Wilcoxon test, p-value  $< 1e^{-10}$ , Supplementary Figure 10). Interestingly, intergenic regions found  
283 in the vicinity of lncRNA loci (Supplementary Methods) had slightly lower conservation scores than all  
284 intergenic regions, on average (median 0.072, Wilcoxon test, p-value  $< 1e^{-6}$ , Supplementary Figure 10).  
285 Promoter conservation levels are more comparable between protein-coding genes (median score  
286 0.17) and lncRNAs (median score 0.08), though still significantly higher for the former (Wilcoxon test,  
287 p-value  $< 1e^{-10}$ , Supplementary Figure 10). Among lncRNA classes, the highest levels of promoter  
288 sequence conservation (median 0.14) are observed for bidirectional promoters shared with protein-  
289 coding genes (Supplementary Figure 10).

290 We next analyzed sets of protein-coding genes and lncRNAs that are expressed above noise  
291 levels (TPM $\geq$ 1, averaged across all biological replicates) in each organ / developmental stage  
292 combination (Supplementary Table 6). For all examined regions and for both categories of genes, the  
293 spatio-temporal expression pattern is associated with the level of sequence conservation. Globally,  
294 sequence conservation is higher for genes that are expressed earlier in development than for genes  
295 expressed later in development, and reaches its lowest values for genes expressed in adult and aged  
296 testes (Figure 4). For exonic sequences and splice sites, the amount of sequence conservation is  
297 significantly lower for lncRNAs than for protein-coding genes, irrespective of the organ and  
298 developmental stage in which they are expressed (Wilcoxon test, p-value  $< 1e^{-10}$ , Figure 4A, C).  
299 However, for promoter regions, the differences between the two gene categories are weaker, and are  
300 not statistically significant for the mid-stage embryonic brain (median 0.21 for protein-coding genes,

301 0.20 for lncRNAs, Wilcoxon test, p-value 0.08) and kidney (median 0.20 for protein-coding genes, 0.21  
302 for lncRNAs, Wilcoxon test, p-value 0.76), and for the late embryonic kidney (median 0.20 for protein-  
303 coding genes, 0.19 for lncRNAs, Wilcoxon test, p-value 0.15). As noted before, the highest levels of  
304 lncRNA promoter conservation are observed for lncRNAs that have bidirectional promoters shared  
305 with protein-coding genes or other non-coding loci (Supplementary Figure 11A-C).

306 Finally, we asked whether the highest level of evolutionary sequence conservation is seen at  
307 exons, promoter or splice site regions, for lncRNA loci taken individually. Here again, the answer  
308 depends on the expression pattern: for lncRNAs detected in somatic organs and in the developing  
309 testes, there is significantly higher conservation for promoters than for exons (Wilcoxon test, p-value  
310  $< 1e^{-3}$  for all organ / developmental stage combinations, Supplementary Figure 11D, E). We also  
311 observe significantly higher conservation for splice sites than for exons (Wilcoxon test, p-value  $<$   
312 0.005), in all samples except aged liver (Wilcoxon test, p-value 0.052). However, when we consider  
313 lncRNAs that are expressed above noise levels in the young and aged adult testes (which constitute  
314 the great majority of loci), the conservation scores are slightly but significantly higher for exons than  
315 for promoters or splice sites (Wilcoxon test, p-value  $< 1e^{-9}$ , Supplementary Figure 11D, E).

#### 316 Detection of homologous lncRNAs across species

317 We next sought to assess the conservation of lncRNA repertoires in mouse, rat and chicken.  
318 We detected lncRNA separately in each species, using only RNA-seq data and existing genome  
319 annotations, as previously suggested (Hezroni et al. 2015). We then searched for putative 1-to-1  
320 orthologous lncRNAs between species using pre-computed whole-genome alignments as a guide  
321 (Materials and methods), to increase the sensitivity of orthologous gene detection in the presence of  
322 rapid sequence evolution (Washietl et al. 2014). The orthologous lncRNA detection procedure involves  
323 several steps, including the identification of putative homologous (projected) loci across species,  
324 filtering to remove large-scale structural changes in the loci, and intersection with predicted loci in the  
325 target species (Materials and methods). As illustrated in Figure 5, for comparisons between rodents  
326 the extent of sequence divergence is low enough that more than 90% of 18858 lncRNA loci are  
327 successfully projected from mouse to rat (Figure 5A, Supplementary Dataset 5). However, only 54% of  
328 projected loci have detectable transcription in the target species (at least 10 uniquely mapped reads).  
329 Only 23% of mouse lncRNA loci have predicted 1-to-1 orthologues in the rat, and only 15% are  
330 orthologous to confirmed lncRNA loci in the rat (Figure 5A, Supplementary Dataset 5). The 1493 mouse  
331 lncRNAs that have non-lncRNAs orthologues in the rat are generally matched with loci discarded  
332 because of low read coverage, minimum exonic length or distance to protein-coding genes  
333 (Supplementary Dataset 5). Cases of lncRNA-protein-coding orthologues are rare at this evolutionary  
334 distance (Supplementary Dataset 5), and they may stem from gene classification errors. We note that

335 orthologous lncRNA gene structures are highly divergent across species, in terms of exonic length or  
336 number of exons (Supplementary Figure 12). At larger evolutionary distances, the rate of sequence  
337 evolution is the main factor hampering detection of orthologous lncRNAs. Only 2613 (14%) of mouse  
338 lncRNAs could be projected on the chicken genome, and after subsequent filtering we detect only 66  
339 mouse – chicken lncRNA orthologues, and 30 lncRNAs with orthologues in all three species  
340 (Supplementary Dataset 5, Supplementary Table 7).

341 Conserved lncRNAs differ from species-specific lncRNAs in terms of expression patterns. While  
342 only subtle differences can be observed when comparing mouse-rat orthologous lncRNAs to the  
343 mouse-specific lncRNA set, lncRNAs that are conserved between mouse and chicken are enriched in  
344 somatic organs and early developmental stages (Figure 5B, C). For example, only 15% of mouse-  
345 specific lncRNAs reach their maximum expression in the brain, which is significantly lower than the  
346 observed proportion for mouse lncRNAs with orthologues in rat (18%, Chi-square test, p-value  $3e^{-4}$ )  
347 and for mouse lncRNAs with orthologues in the chicken (39%, Chi-square test, p-value  $1.5e^{-7}$ ). Likewise,  
348 while only 9.9% of mouse-specific lncRNAs have their highest level of expression in one of the two  
349 embryonic stages, this proportion is significantly higher for lncRNAs with orthologues in the chicken  
350 (27%, Chi-square test, p-value 0.002). We note however that these results may be affected by our  
351 narrower sampling for the chicken, which is biased towards embryonic organs, although we did include  
352 data from adult organs for this species (Supplementary Table 2).

#### 353 *Patterns of lncRNA expression variation across species, organs and developmental stages*

354 We next assessed the global patterns of expression variation across species, organs and  
355 developmental stages, for predicted mouse – rat lncRNA orthologues (Supplementary Dataset 6). As  
356 for protein-coding genes, the main source of variability in a PCA performed on lncRNA expression levels  
357 is the difference between adult and aged testes and the other samples (Figure 6A). However, for  
358 lncRNAs, samples cluster according to the species of origin already on the second factorial axis (11.6%  
359 explained variance), confirming that lncRNA expression patterns evolve rapidly. Overall, differences  
360 between organs and developmental stages are less striking for lncRNAs, compared to differences  
361 between species (Figure 6A). This pattern is also visible on a hierarchical clustering analysis (performed  
362 on distances derived from Spearman's correlation coefficient): in contrast with what is observed for  
363 protein-coding genes, for lncRNAs samples generally cluster by species, with the exception of young  
364 and aged adult testes, which are robustly grouped (Figure 6B).

365 The higher rates of lncRNA expression evolution are also visible when analyzing within-species  
366 variations, through comparisons across biological replicates (Figure 7A). We sought to measure the  
367 selective pressures acting on expression patterns by contrasting between-species and within-species  
368 variations, in the spirit of a classical approach for coding sequences (McDonald and Kreitman 1991).

369 We constructed an expression conservation index by dividing the between-species and the within-  
370 species Spearman's correlation coefficient, computed on all genes from a category, for a given  
371 organ/developmental stage combination (Supplementary Methods). The resulting values are very high  
372 for protein-coding genes, in particular for the brain and the mid-stage embryonic kidney, where the  
373 expression conservation scores are above 0.95. However, there is significant less conservation  
374 between species for the adult and aged testes (expression conservation score  $\sim 0.88$ , bootstrap p-value  
375  $< 0.01$ , Figure 7B). For lncRNAs, expression conservation values vary between 0.5 and 0.7, significantly  
376 lower than for protein-coding genes (bootstrap p-value  $< 0.01$ ). The lowest conservation scores are  
377 observed for young and aged adult testes (Figure 7C).

### 378 Parallel patterns of temporal expression variation for mouse and rat lncRNAs

379 We delved deeper into the evolutionary comparison of protein-coding genes and lncRNA  
380 expression patterns, by asking whether temporal expression variations are shared between species.  
381 Several hundred orthologous lncRNAs are DE (FDR $< 0.01$ ) in both mouse and rat, in each organ  
382 (minimum 150 in liver, maximum 1583 in testes, Figure 8A). Likewise, between 6775 (in liver) and  
383 10608 (in testes) protein-coding genes are DE in both species (Supplementary Figure 13). Overall,  
384 shared DE lncRNAs show similar patterns of variation among developmental stages in mouse and rat,  
385 reaching their maximum expression in the same (or close) developmental stages (Figure 8A). For  
386 example, out of 42 lncRNAs that are DE in mouse brain and reach their maximum expression in the  
387 mid-stage embryo, 31 (74%) reach their maximum expression in the corresponding stage in the rat  
388 (Figure 8A). We clustered the relative expression profiles using the K-means algorithm (Supplementary  
389 Methods). Although individual gene trajectories show variations between species, the average  
390 expression profiles within each K-means cluster are generally similar between mouse and rat (Figure  
391 8B-E, Supplementary Figure 13). This is particularly striking for the brain, where all 5 lncRNA clusters  
392 show similar average expression profiles for the two species (Figure 8B). Greater differences between  
393 species are observed in other organs, such as the kidney, where 2 out of 5 clusters (120 genes in total,  
394 that is 31% of shared DE lncRNAs in kidney) have average expression profiles that reach their maximum  
395 in different stages in mouse and rat (Figure 8C). The promoters of shared DE lncRNAs in each cluster  
396 contain transcription factor binding sites that are coherent with the expression profile of the cluster,  
397 such as brain homeobox POU3F2/BRN2 binding sites for the first K-means cluster in the brain, which  
398 has maximum expression in the mid-stage embryo (Supplementary Table 8). We note that  
399 transcription factor binding site enrichments are generally not statistically significant for lncRNAs,  
400 partly due to the low gene counts in each cluster.

401 The testis is the only organ where opposite K-mean cluster expression profiles are observed in  
402 the two species (increasing with time in mouse and decreasing in rat, or vice-versa). For lncRNAs, this

403 occurs for 1 of the 4 detected clusters, containing 56 lncRNAs (3.5% of all shared DE lncRNAs in this  
404 organ, Figure 8E). For protein-coding genes, opposite average profiles are observed for 2 out of 4  
405 clusters, comprising 1182 and 1509 genes, i.e. 25% of all shared testes-DE protein-coding genes  
406 (Supplementary Figure 13). These clusters do not stand out in terms of transcription factor binding site  
407 (Supplementary Table 8) or gene ontology enrichment (Supplementary Dataset 4). This pattern  
408 confirms previous reports of rapid expression evolution in the adult testes (Brawand et al. 2011), and  
409 extends them by showing that patterns of variations among developmental stages are often species-  
410 specific in the testes, for protein-coding genes.

#### 411 Evolutionary divergence of individual lncRNA expression profiles

412 To further quantify lncRNA expression differences between species, we measured the  
413 Euclidean distance between relative expression profiles (average TPM values across biological  
414 replicates, normalized by dividing by the sum of all values for a gene, for each species), for mouse and  
415 rat orthologues (Supplementary Methods, Supplementary Dataset 7, Supplementary Table 9). The  
416 resulting expression divergence values correlate negatively with the average expression level (R-  
417 squared 0.13, t-test p-value  $< 1e^{-10}$ , Figure 9A), as expected given that abundance estimation is less  
418 reliable for weakly expressed genes. While the raw expression divergence values are significantly  
419 higher for lncRNAs (median 0.18) than for protein-coding genes (median 0.11, Wilcoxon test p-value  $<$   
420  $1e^{-10}$ , Figure 9B), this is largely due to the low lncRNA expression levels. Indeed, the effect disappears  
421 when analyzing the residual expression divergence after regressing the expression level (median value  
422  $-0.03$  for protein-coding genes,  $-0.06$  for lncRNAs, Wilcoxon test  $< 1e^{-10}$ , Figure 9C). These patterns  
423 remain true when analyzing separately protein-coding and lncRNAs with different types of promoters,  
424 bidirectional or unidirectional (Supplementary Figure 14A). For lncRNAs, we also observe a weak  
425 negative correlation between expression divergence and the extent of gene structure conservation (R-  
426 squared 0.04, t-test p-value  $< 1e^{-10}$ , Figure 9D). We measured the relative contribution of each  
427 organ/developmental stage to the expression divergence estimate (Figure 9E). For both protein-coding  
428 genes and lncRNAs, by far the highest contributors are the young adult and aged testes samples, which  
429 are responsible for almost 30% of the lncRNA expression divergence (Figure 9E). This is visible in the  
430 expression patterns of the 2 protein-coding and lncRNA genes with the highest residual expression  
431 divergence: the lncRNA expression divergence is mostly due to changes in adult testes, while more  
432 complex expression pattern changes seem to have occurred for the protein-coding genes  
433 (Supplementary Figure 14). The most divergent protein-coding genes are enriched in functions related  
434 to immunity (Supplementary Dataset 7), suggesting that differences in immune cell infiltrates among  
435 species could be responsible for these extreme cases of expression pattern divergence.

436 Candidate species-specific lncRNAs

437 We next investigated the most extreme cases of expression divergence: situations where  
438 expression can be robustly detected in one species, but not in the other one, despite almost perfect  
439 sequence alignment (Supplementary Methods). We selected lncRNA loci that were supported by at  
440 least 100 uniquely mapped reads in one species, with no reads detected in the predicted homologous  
441 region in the other species. With this convention, we obtain 1041 candidate mouse-specific and 1646  
442 candidate rat-specific loci (Supplementary Dataset 8). These lists include striking examples, such as the  
443 region downstream of the *Fzd4* protein-coding gene, which contains a mouse-specific and a rat-specific  
444 lncRNA candidate, each perfectly aligned in the other species (Supplementary Figure 15). Candidate  
445 species-specific lncRNAs are more frequently associated with predicted enhancers than orthologous  
446 lncRNAs (52% vs. 33%, Chi-square test, p-value  $<1e^{-10}$ ), are less often spliced (56% vs. 61%, Chi-square  
447 test p-value  $1.6e^{-3}$ ) and associated with bidirectional promoters (24% vs. 61%, Chi-square test, p-value  
448  $<1e^{-10}$ , Supplementary Figure 16). Moreover, we could confirm that their presence is associated with  
449 increased expression divergence in the neighboring genes. To test this, we selected species-specific  
450 and orthologous lncRNAs that are transcribed from bidirectional promoters shared with protein-  
451 coding genes, and evaluated the expression divergence of their protein-coding neighbors  
452 (Supplementary Figure 16D, E). Though the difference is subtle, genes that are close to species-specific  
453 lncRNAs have significantly higher expression divergence than the ones that have conserved lncRNA  
454 neighbors, even after correcting for expression levels (Wilcoxon test, p-value  $<1e^{-3}$ ). It thus seems that  
455 expression changes that led to the species-specific lncRNA transcription extend beyond the lncRNA  
456 locus and affect neighboring genes, as previously proposed (Kutter et al. 2012).

457 **Discussion**

458 Comparative transcriptomics across species, organs and developmental stages

459 More than a decade after the publication of the first genome-wide lncRNA datasets (Guttman  
460 et al. 2009; Khalil et al. 2009), the debate regarding their functionality is still not settled. Evolutionary  
461 approaches provide important tools to assess biological functionality (Haerty and Ponting 2014), and  
462 they have been already successfully applied to lncRNAs. However, most large-scale comparative  
463 transcriptomics studies available so far (Kutter et al. 2012; Washietl et al. 2014; Hezroni et al. 2015;  
464 Necsulea et al. 2014), with one recent exception (Sarropoulos et al. 2019), have focused on lncRNAs  
465 detected in adult organs. We hypothesized that lncRNAs expressed during development may be  
466 enriched in functional loci, as suggested by the increasing number of lncRNAs with proposed  
467 developmental roles (Rinn et al. 2007; Sauvageau et al. 2013; Grote et al. 2013; Grote and Herrmann  
468 2015). To test this hypothesis, we performed a multi-dimensional comparative transcriptomics  
469 analysis, following lncRNA and protein-coding genes across species, organs and developmental stages.

470 We ensured that our transcriptome collection reflects the changes in cellular composition and  
471 physiological functions that occur during major organ development, by analyzing cell-type specific  
472 gene markers derived from single-cell analyses (Tabula Muris Consortium 2018; Green et al. 2018). We  
473 showed that protein-coding gene expression profiles across major organs and developmental stages  
474 are well conserved among species, even at large evolutionary distances. Although differences among  
475 rodents and chicken are considerable when analyzing the full set of orthologous protein-coding genes  
476 (Figure 1), the expression profiles of genes that are known to be implicated in embryonic development  
477 and in gene expression regulation processes are highly conserved among species (Supplementary  
478 Figure 6). Our transcriptome collection thus enables detecting temporal expression patterns shared  
479 across divergent species, for key players in developmental regulatory networks. These observations  
480 are consistent with findings from a recent publication, which studied protein-coding gene expression  
481 patterns during major organ development in amniote species (Cardoso-Moreira et al. 2019). Our  
482 transcriptome dataset covers a narrower range of species and developmental stages than this  
483 comprehensive resource (Cardoso-Moreira et al. 2019), but uniquely includes aged individuals, thus  
484 completing the overview of temporal expression patterns. Thus, our work represents an additional  
485 resource for evolutionary studies of gene expression.

486 To our knowledge, together with a recent publication (Sarropoulos et al. 2019), our work is  
487 one of the first large-scale lncRNA evolutionary studies that include a temporal dimension, by sampling  
488 different developmental stages. Our manuscript and this recent work concur to reveal an enrichment  
489 for functional lncRNAs early in development (Sarropoulos et al. 2019). Here, we perform in-depth  
490 analyses of expression pattern evolution, short-term and long-term sequence evolution for different  
491 regions of lncRNAs loci, in conjunction with their expression patterns. We thus bring new insights into  
492 the evolution and functionality of lncRNAs.

#### 493 *Spatio-temporal lncRNA expression patterns*

494 Our first major observation is that lncRNAs are overwhelmingly expressed in the young and  
495 aged adult testes (Figure 3), in agreement with previous data (Soumillon et al. 2013). Their relative  
496 depletion in embryonic and newborn testes reinforces the association between lncRNA transcription  
497 and spermatogenesis, in accord with the hypothesis that the particular chromatin environment during  
498 spermatogenesis is a driver for promiscuous, non-functional transcription (Kaessmann 2010;  
499 Soumillon et al. 2013). Interestingly, we show that lncRNAs are significantly differentially expressed  
500 among developmental stages, at least as frequently as protein-coding genes, after correcting for their  
501 lower expression levels. However, in contrast with protein-coding genes, the majority of lncRNAs reach  
502 their highest expression levels in adult rather than in developing organs (Figure 3). As requirements  
503 for tight gene expression control are higher during embryonic development (Ben-Tabou de-Leon and



504 Davidson 2007), an explanation for the relative lncRNA depletion in embryonic and newborn  
505 transcriptomes is that transcriptional noise is deleterious and thus more efficiently blocked during the  
506 early stages of development. Differences in cellular composition heterogeneity may also be part of the  
507 explanation. Expression analyses of cell-type specific markers suggest that adult organ transcriptomes  
508 may be a mix of more diverse cell types, including substantial immune cell infiltrates (Supplementary  
509 Figure 1). A higher cell type diversity may explain the increased abundance of lncRNAs in young and  
510 aged adult organs, especially given that lncRNAs are thought to be cell-type specific (Liu et al. 2016).

511 We found that lncRNA expression patterns are generally similar between young and aged adult  
512 individuals: less than 50 lncRNAs are significantly DE between these two stages, for most organs  
513 (Supplementary Figure 9). Moreover, the levels of sequence and expression conservation are globally  
514 similar between young and aged adults, for both protein-coding and lncRNA genes (Figures 4,7).  
515 Overall, our analyses indicate that, with our sampling (Supplementary Table 1), the physiological  
516 processes at work in aged organs are highly similar to those acting in juvenile organs, suggesting that  
517 developmental stage sampling should be further extended for in-depth analyses of the aging process.

#### 518 *Functionally constrained lncRNAs are enriched in developmental transcriptomes*

519 Our long-term sequence conservation analyses confirm that lncRNAs are overall under weak,  
520 but significant selective constraint (Ponjavic et al. 2007): lncRNA sequence conservation scores are  
521 much lower than those of protein-coding genes, but above those of intergenic regions (Figure 4,  
522 Supplementary Figures 10-11). Interestingly, intergenic regions flanking lncRNAs are on average less  
523 conserved than the genomic intergenic average (Figure 4), suggesting that the rapid lncRNA evolution  
524 may be a general feature of the genomic regions in which they reside. The underlying mechanisms are  
525 unclear, but may reflect a lower density of constrained expression regulatory elements in the vicinity  
526 of lncRNAs, or a higher accumulation of lineage-specific transposable elements (Kapusta et al. 2013).

527 We show that, for those lncRNAs that are expressed above noise levels ( $TPM \geq 1$ ) in somatic  
528 organs and in the embryonic and newborn developmental stages, there is a higher proportion of  
529 evolutionarily constrained loci than in testes-expressed lncRNAs (Figure 4). Strikingly, we find that the  
530 level of long-term sequence conservation for lncRNA promoter regions is similar to the one observed  
531 for protein-coding promoters, when we analyze genes that are robustly expressed in embryonic brain  
532 and kidney. Furthermore, we show that lncRNAs expressed in somatic organs and in the developing  
533 testes differ from those expressed in the adult testes not only in terms of overall levels of sequence  
534 conservation, but also with respect to the regions of the lncRNA loci that are under selective constraint.  
535 Thus, for lncRNAs expressed in somatic organs and in the developing testes, there is significantly more  
536 evolutionary constraint on promoters and splice sites than on exons, while these patterns are not seen  
537 for the bulk of lncRNAs, expressed in adult and aged testes (Supplementary Figure 11). We are thus

538 able to modulate previous reports of increased constraint on splicing regulatory regions in mammalian  
539 lncRNAs (Schüler et al. 2014; Haerty and Ponting 2015), by showing that this pattern is specific to  
540 lncRNAs that are expressed in somatic organs and in the developing testes.

541 These results are also in agreement with recent findings suggesting that biological function  
542 may reside in the presence of additional non-coding regulatory elements at the lncRNA promoter  
543 rather than in the production of a specific transcript (Engreitz et al. 2016; Groff et al. 2016). While the  
544 elevated sequence conservation at splicing regulatory signals could indicate that the production of a  
545 specific mature lncRNA is required, splicing of lncRNA transcripts was recently proposed to affect the  
546 expression of neighboring protein-coding genes (Engreitz et al. 2016). Thus, while there is evidence for  
547 increased functionality for lncRNA loci that are detected in developmental transcriptomes or in adult  
548 somatic organs, in agreement with a recent report (Sarropoulos et al. 2019), our sequence  
549 conservation analyses are compatible with the hypothesis that their biological functions may be  
550 carried out in an RNA-independent manner, as exons are under less constraint than promoters or  
551 splice sites. Alternatively, their function may be carried out by small conserved elements, such that  
552 the sequence conservation on the entire lncRNA exonic sequence is weak (Ulitsky 2016).

#### 553 *Evolutionary divergence of spatio-temporal expression profiles for lncRNAs*

554 We previously showed that lncRNA expression patterns evolve rapidly across species in adult  
555 organs (Necsulea et al. 2014). Here, we show that this rapid evolution of lncRNA expression is also true  
556 for embryonic and newborn developmental stages. Expression comparisons across species, organs and  
557 developmental stages are dominated by differences between species for lncRNAs (Figure 6), while  
558 similarities between organs and developmental stages are predominant for protein-coding genes,  
559 even across distantly-related species (Figure 1). We assessed the extent of expression conservation by  
560 contrasting between-species and within-species expression variations and we showed that lncRNAs  
561 have significantly lower levels of conservation than protein-coding genes, for all organs and  
562 developmental stages (Figure 7). However, lncRNA expression is more conserved in somatic organs  
563 and in early embryonic stages than in the adult testes. Moreover, when orthologous lncRNAs are  
564 differentially expressed among developmental stages in both mouse and rat, they generally show  
565 parallel profiles of expression variation in both species (Figure 8). This result is in agreement with a  
566 recent publication, which showed that temporal patterns of expression variation tend to be  
567 evolutionarily conserved for developmentally dynamic lncRNAs (Sarropoulos et al. 2019). We note that  
568 these temporal patterns of variation may in fact be caused by spatially-restricted lncRNA expression.  
569 Previous reports indicated that lncRNA expression may be cell type-specific (Liu et al. 2016). The  
570 differentially expressed lncRNAs, shared across mouse and rat, could be specific of cell types that  
571 change their relative abundance in whole-organ transcriptomes with developmental time.

572 Interestingly, when we evaluate expression divergence individually for each orthologous gene  
573 pair, correcting for the lower lncRNA expression levels, we find that lncRNAs are not more divergent  
574 than protein-coding genes (Figure 8). This observation indicates that much of the between-species  
575 differences in lncRNA expression patterns is tightly linked with the low expression levels of lncRNAs. It  
576 is not clear however whether this is purely an indication of technical biases, that hamper expression  
577 estimation for lowly expressed lncRNAs, or whether the low lncRNA expression levels are a sign that  
578 these transcripts are non-functional. For cell type-specific lncRNAs, low expression in whole-organ  
579 transcriptomes are expected. This question may soon be directly addressed, as single-cell assays  
580 become more sensitive and allow investigation of lncRNAs (Liu et al. 2016).

#### 581 Candidate species-specific lncRNAs

582 Finally, we analyzed extreme cases of expression divergence between species, where  
583 transcription can be robustly detected in one species but not in the other, despite the presence of  
584 good sequence conservation. We identify more than a thousand candidate species-specific lncRNAs,  
585 in both mouse and rat. Interestingly, we observe that candidate mouse-specific lncRNAs are more  
586 frequently transcribed from enhancers than lncRNAs conserved between mouse and rat  
587 (Supplementary Figure 11). This observation is consistent with previous reports that enhancers and  
588 enhancer-associated lncRNAs evolve rapidly (Villar et al. 2015; Marques et al. 2013). Moreover, we  
589 show that these lncRNA expression changes do not occur in an isolated manner. When species-specific  
590 lncRNA transcription was inferred at protein-coding genes bidirectional promoters, the neighboring  
591 protein-coding genes also showed increased expression divergence, compared to genes that are  
592 transcribed from conserved lncRNA promoters. We thus confirm that lncRNA turnover is associated  
593 with changes in neighboring gene expression (Kutter et al. 2012). While lncRNAs changes may be  
594 directly affecting gene expression, another probable hypothesis is that a common mechanism affects  
595 both lncRNAs and protein-coding genes transcribed from bidirectional promoters.

#### 596 Conclusions

597 Our comparative transcriptomics approach confirms that lncRNAs repertoires, sequences and  
598 expression patterns evolve rapidly across species, and shows that accelerated rates of lncRNA  
599 evolution are also seen in developmental transcriptomes, albeit less frequently. These observations  
600 are consistent with the hypothesis that the majority of lncRNAs (or at least of those detected with  
601 sensitive transcriptome sequencing approaches, in particular in the adult testes) may be non-  
602 functional. However, we are able to modulate this conclusion, by showing that there are increased  
603 levels of functional constraint on lncRNAs expressed during embryonic development, in particular in  
604 the developing brain and kidney. These increased levels of constraint apply to all analyzed aspects of  
605 lncRNAs, including sequence conservation for exons, promoter and splice sites, but also expression

606 pattern conservation. For many of these loci, biological function may be RNA-independent, as the  
607 highest levels of selective constraint are observed on promoter regions and on splice signals, rather  
608 than on lncRNA exonic sequences. Our results are thus compatible with unconventional, RNA-  
609 independent functions for lncRNAs expressed during embryonic development.

## 610 **Materials and methods**

### 611 Biological sample collection

612 We collected samples from three species (mouse C57BL/6J strain, rat Wistar strain and chicken  
613 White Leghorn strain), four organs (brain, kidney, liver and testes) and five developmental stages  
614 (including two embryonic stages, newborn, young and aged adult individuals). We sampled the  
615 following stages in the mouse: embryonic day post-conception (dpc) 13.5 (E13.5 dpc, hereafter mid-  
616 stage embryo); E17 to E17.5 dpc (late embryo); post-natal day 1 to 2 (newborn); young adult (8-10  
617 weeks old); aged adult (24 months old). For the rat, we sampled the following stages: E15 dpc (mid-  
618 stage embryo); E18.5 to E19 dpc (late embryo); post-natal day 1 to 2 (newborn); young adult (8-10  
619 weeks old); aged adult (24 months, with the exception of kidney samples and two of four liver samples,  
620 derived from 12 months old individuals). The embryonic and neonatal developmental stages were  
621 selected for maximum comparability based on Carnegie stage criteria ([Theiler 1989](#)). For chicken, we  
622 collected samples from Hamburger-Hamilton stages 31 and 36 (hereafter termed mid-stage and late  
623 embryo), selected for comparability with the two embryonic stages in mouse and rat (Hamburger and  
624 Hamilton 1951). Each sample corresponds to one individual, except for mouse and rat mid-stage  
625 embryonic kidney, for which tissue from several embryos was pooled prior to RNA extraction. For adult  
626 and aged organs, multiple tissue pieces from the same individual were pooled and homogenized prior  
627 to RNA extraction. For brain dissection, we sampled the cerebral cortex. For mouse and rat, with the  
628 exception of the mid-stage embryonic kidney, individuals were genotyped and males were selected  
629 for RNA extraction. Between two and four biological replicates were obtained for each  
630 species/organ/stage combination, amounting to 97 samples in total (Supplementary Table 1).

### 631 RNA-seq library preparation and sequencing

632 We performed RNA extractions using RNeasy Plus Mini kit from Qiagen. We assessed RNA  
633 quality with the Agilent 2100 Bioanalyzer. RNA integrity numbers (RIN) are available in Supplementary  
634 Table 1; see Supplementary Methods for additional RNA integrity analyses. Sequencing libraries were  
635 produced with the Illumina TruSeq stranded mRNA protocol with polyA selection, and sequenced as  
636 101 base pairs (bp) single-end reads, at the Genomics Platform of iGE3 and the University of Geneva.  
637 Libraries are strand-specific and the sequenced strand is complementary to the RNA molecule.

638 Additional RNA-seq data

639 To improve detection power for lowly expressed lncRNAs, we complemented our RNA-seq  
640 collection with samples generated with the same technology for Brown Norway rat adult organs  
641 (Cortez et al. 2014). We added published data for adult chicken (red jungle fowl strain UCD001) organs  
642 (McCarthy et al. 2019), as well as for embryonic chicken (White Leghorn) organs (Uebbing et al. 2015;  
643 Ayers et al. 2013). As the data were not comparable with our own in terms of library preparation and  
644 animal strains, these samples were only used to increase lncRNA detection sensitivity.

645 RNA-seq data processing

646 We used HISAT2 (Kim et al. 2015) release 2.0.5 to align the RNA-seq data on reference  
647 genomes. The genome sequences (assembly versions mm10/GRCm38, rn6/Rnor\_6.0 and  
648 galGal5/Gallus\_gallus-5.0) were downloaded from the Ensembl database (Cunningham et al. 2019).  
649 Genome indexes were built using only genome sequence information. To improve detection  
650 sensitivity, at the alignment step we provided known splice junction coordinates extracted from  
651 Ensembl. We set the maximum intron length for splice junction detection at 1 million base pairs (Mb).  
652 The following command-line arguments were used: --rna-strandness R --known-splicesite-  
653 infile=SpliceSites\_Ensembl.txt --max-intronlen 1000000 --dta-cufflinks, where SpliceSites\_Ensembl.txt  
654 corresponds to the exon junction coordinates extracted with hisat2\_extract\_splice\_sites.py. See also  
655 Supplementary Methods for additional RNA-seq data quality analyses.

656 Transcript assembly and filtering

657 We assembled transcripts for each sample using StringTie (Pertea et al. 2015), release 1.3.5,  
658 based on read alignments obtained with HISAT2. We provided genome annotations from Ensembl  
659 release 94 as a guide for transcript assembly. We filtered Ensembl annotations to remove transcripts  
660 that spanned a genomic length above 2.5 Mb. For protein-coding genes, we kept only protein-coding  
661 transcripts, discarding isoforms annotated as “retained\_intron”, “processed\_transcript” etc. We set  
662 the minimum exonic length at 150 bp, the minimum anchor length for splice junctions at 8bp and the  
663 minimum isoform fraction at 0.05. The following StringTie command-line arguments were used: -G  
664 Ensembl\_annotations.gtf -m 150 -a 8 -f 0.05 -p 8 -rf, where Ensembl\_annotations.gtf correspond to  
665 the Ensembl annotations filtered as described above. We compared the resulting assembled  
666 transcripts with Ensembl annotations and we discarded read-through transcripts, overlapping with  
667 multiple multi-exonic Ensembl-annotated genes. For strand-specific samples, we discarded transcripts  
668 for which the ratio of sense to antisense unique read coverage was below 0.01. We discarded multi-  
669 exonic transcripts that were not supported by splice junctions with correctly assigned strands. The  
670 filtered transcripts obtained for each sample were assembled into a single dataset *per* species using  
671 the merge option in StringTie. For increased sensitivity, we removed the minimum FPKM and TPM

672 thresholds, but required a minimum isoform fraction of 0.05 for transcript inclusion. The following  
673 StringTie command-line arguments were used: stringtie -v --merge -G Ensembl\_annotatons.gtf -m 150  
674 -a 8 -p 8 -F 0 -T 0 -f 0.05. We constructed a combined annotation dataset, starting with Ensembl  
675 annotations, to which we added newly-assembled transcripts that had no exonic overlap with Ensembl  
676 genes. We also included newly-annotated isoforms for known genes if they had exonic overlap with  
677 exactly one Ensembl gene, thus discarding potential read-through transcripts or gene fusions.

#### 678 Protein-coding potential of assembled transcripts

679 To determine whether the newly assembled transcripts were protein-coding or non-coding,  
680 we mainly relied on the codon substitution frequency (CSF) score (Lin et al. 2007). As in a previous  
681 publication (Necsulea et al. 2014) we scanned whole genome alignments and computed CSF scores in  
682 75 bp sliding windows moving with a 3 bp step. We used pre-computed alignments downloaded from  
683 the UCSC Genome Browser (Casper et al. 2018), including the alignment between the mouse genome  
684 and 59 other vertebrates (for mouse classification), between the human genome and 99 other  
685 vertebrates (for rat and chicken classification) and between the rat genome and 19 other vertebrates  
686 (for rat classification). For each window, we computed the score in each of the 6 possible reading  
687 frames and extracted the maximum score for each strand. We considered that transcripts are protein-  
688 coding if they overlapped with positive CSF scores on at least 150 bp. As positive CSF scores may also  
689 appear on the antisense strand of protein-coding regions due to the partial strand-symmetry of the  
690 genetic code, in this analysis we considered only exonic regions that did not overlap with other genes.  
691 In addition, we searched for sequence similarity between assembled transcripts and known protein  
692 sequences from the SwissProt 2017\_04 (The UniProt Consortium 2017) and Pfam 31.0 (El-Gebali et al.  
693 2019) databases. We kept only SwissProt entries with confidence scores 1, 2 or 3 and we used the  
694 Pfam-A curated section of Pfam. We searched for sequence similarity using the blastx utility in the  
695 BLAST+ 2.8.1 package (Camacho et al. 2009; Altschul et al. 1990), keeping hits with maximum e-value  
696  $1e^{-3}$  and minimum protein sequence identity 40%, on repeat-masked cDNA sequences. We considered  
697 that transcripts were protein-coding if they overlapped with blastx hits over at least 150 bp. Genes  
698 were said to be protein-coding if at least one of their isoforms was classified as protein-coding, based  
699 on either the CSF score or on sequence similarity with known proteins.

#### 700 Long non-coding RNA selection

701 To construct a reliable lncRNA dataset, we selected newly-annotated genes classified as non-  
702 coding based on both the CSF score and on sequence similarity with known proteins and protein  
703 domains, as well as Ensembl-annotated genes with non-coding biotypes ("lincRNA",  
704 "processed\_transcript", "antisense", "TEC", "macro\_lncRNA", "bidirectional\_promoter\_lncRNA",  
705 "sense\_intronic"). For newly detected genes, we applied several additional filters: we required a

706 minimum exonic length (corresponding to the union of all annotated isoforms) of at least 200 bp for  
707 multi-exonic loci and of at least 500 bp for mono-exonic loci; we eliminated genes that overlapped for  
708 more than 5% of their exonic length with unmappable regions; we kept only loci that were classified  
709 as intergenic and at least 5 kb away from Ensembl-annotated protein-coding genes on the same strand;  
710 for multi-exonic loci, we required that all splice junctions be supported by reads with correct strand  
711 assignment (cf. above). For both *de novo* and Ensembl annotations, we removed transcribed loci that  
712 overlapped on at least 50% of their length with retrotransposed gene copies, annotated by the UCSC  
713 Genome Browser and from a previous publication (Carelli et al. 2016); we discarded loci that  
714 overlapped with UCSC-annotated tRNA genes and with RNA-type elements from RepeatMasker (Smit  
715 et al. 2003) on at least 25% of their length. We kept loci supported by at least 10 uniquely mapped  
716 RNA-seq reads and for which a ratio of sense to antisense transcription of at least 1% was observed in  
717 at least one sample. Although the fraction of reads stemming from the wrong strand due to errors in  
718 library preparations is very low in our samples (Supplementary Table 1), loci situated on the antisense  
719 strand of highly expressed genes can have unreliable expression estimates. Thus, for loci that had  
720 sense/antisense exonic overlap with other genes, we computed expression levels either on complete  
721 gene annotations, or only on exonic regions that had no overlap with other genes, and computed  
722 Spearman's correlation coefficient between the two expression estimates, across all samples. We  
723 discarded loci for which the correlation coefficient was below 0.9. Full gene annotations and lncRNA  
724 selection criteria are provided in Supplementary Dataset 1 online.

#### 725 Gene expression estimation

726 We computed the number of uniquely mapping reads unambiguously attributed to each gene  
727 using the Rsubread package in R (Liao et al. 2019), discarding reads that overlapped with multiple  
728 genes. We also estimated read counts and TPM (transcript *per* million) values *per* gene using Kallisto  
729 (Bray et al. 2016). To approach absolute expression levels estimates, for better comparisons across  
730 samples, we further normalized TPM values using a scaling approach (Brawand et al. 2011). Briefly, we  
731 ranked the genes in each sample according to their TPM values, we computed the variance of the ranks  
732 across all samples for each gene, and we identified the 100 least-varying genes, found within the inter-  
733 quartile range (25%-75%) in terms of average expression levels across samples. We derived  
734 normalization coefficients for each sample such that the median of the 100 least-varying genes be  
735 identical across samples. We then used these coefficients to normalize TPM values for each sample.  
736 We excluded mitochondrial genes from expression estimations and analyses, as these genes are highly  
737 expressed and can be variable across samples. For differential expression analyses, we used *per*-gene  
738 unique read counts computed with Rsubread. For all downstream analyses we used normalized TPM

739 values. When indicated, we transformed TPM values with the following formula:  $x \rightarrow \log_2(x+1)$ . Gene  
740 expression data is available in Supplementary Dataset 2 online.

#### 741 Differential expression analyses

742 We used the DESeq2 (Love et al. 2014) package release 1.22.2 in R release 3.5.0 (R Core Team  
743 2018) to test for differential expression across developmental stages, separately for each organ and  
744 species. We analyzed both protein-coding genes and lncRNAs, selected according to the criteria  
745 described above. We first performed a global differential expression analysis, using the likelihood ratio  
746 test to contrast a model including an effect of the developmental stage against the null hypothesis of  
747 homogeneous expression across all developmental stages. This analysis was performed on all protein-  
748 coding and lncRNA genes for each species, as well as on 1-to-1 orthologues for mouse and rat. In  
749 addition, we down-sampled the numbers of reads assigned to protein-coding genes to obtain identical  
750 average numbers of reads for protein-coding genes and lncRNAs. The resampled read counts were  
751 directly proportional to the original counts for each protein-coding gene. We also contrasted  
752 consecutive developmental stages, for each species and organ, using the Wald test implemented in  
753 DESeq2. Differential expression results are available in Supplementary Dataset 4 online.

#### 754 Homologous lncRNA family prediction

755 We used existing whole-genome alignments as a guide to predict homologous lncRNAs across  
756 species, as previously proposed (Washietl et al. 2014). We first constructed for each gene the union of  
757 its exon coordinates across all isoforms, hereafter termed “exon blocks”. We projected exon block  
758 coordinates between pairs of species using the liftOver utility and whole-genome alignments  
759 generated with blastz ([http://www.bx.psu.edu/miller\\_lab/](http://www.bx.psu.edu/miller_lab/)), available through the UCSC Genome  
760 Browser (Casper et al. 2018). To increase detection sensitivity, for the initial liftOver projection we  
761 required only that 10% of the reference bases remap on the target genome. Projections were then  
762 filtered, retaining only cases where the size ratio between the projected and the reference region was  
763 between 0.33 and 3 for mouse and rat (0.2 and 5 for comparisons involving chicken). To exclude recent  
764 lineage-specific duplications, regions with ambiguous or split liftOver projections were discarded. For  
765 genes where multiple exon blocks could be projected across species, we defined the consensus  
766 chromosome and strand in the target genome and discarded projected exon blocks that did not match  
767 this consensus. We then evaluated the order of the projected exon blocks on the target genes, to  
768 identify potential internal rearrangements. If internal rearrangements were due to the position of a  
769 single projected exon block, the conflicting exon block was discarded; otherwise, the entire projected  
770 gene was eliminated. As the projected reference gene coordinates could overlap with multiple genes  
771 in the target genome, we constructed gene clusters based on the overlap between projected exon  
772 block coordinates and target annotations, using a single-link clustering approach. We then realigned



773 entire genomic loci for each pair of reference-target genes found within a cluster, using lastz  
774 ([http://www.bx.psu.edu/miller\\_lab/](http://www.bx.psu.edu/miller_lab/)) and the threaded blockset aligner TBA (Blanchette et al. 2004).  
775 Using this alignment, we computed the percentage of exonic sequences aligned without gaps and the  
776 percentage of identical exonic sequence, for each pair of reference-target genes. We then extracted  
777 the best hit in the target genome for each gene in the reference genome based on the percentage of  
778 identical exonic sequence, requiring that the ratio between the maximum percent identity and the  
779 percent identity of the second-best hit be above 1.1. Reciprocal best hits were considered to be 1-to-  
780 1 orthologous loci between pairs of species. For analyses across all three species, we constructed  
781 clusters of reciprocal best hits from pairwise species comparisons, using a single-link clustering  
782 approach. Resulting clusters with more than 1 representative *per* species were discarded. The results  
783 of the homology prediction pipeline, sequence alignment statistics and Ensembl orthology  
784 relationships for protein-coding genes are available in Supplementary Dataset 5 online.

#### 785 **Availability of data and materials**

786 The RNA-seq data were submitted to the NCBI Gene Expression Omnibus (GEO), under  
787 accession number GSE108348. Supplementary Datasets containing additional processed data are  
788 available at the address: [ftp://pbil.univ-lyon1.fr/pub/datasets/Darbellay\\_LncEvoDevo](ftp://pbil.univ-lyon1.fr/pub/datasets/Darbellay_LncEvoDevo)

789 The scripts used to analyse the data are available through a GitHub repository:  
790 <https://github.com/anecsulea/LncEvoDevo> . See also Supplementary Methods.

#### 791 **Author contributions**

792 FD performed organ dissections, RNA extractions, quality control, prepared samples for  
793 sequencing and contributed to study design and manuscript preparation. AN designed the study,  
794 performed computational analyses and wrote the manuscript.

#### 795 **Acknowledgements**

796 We thank Denis Duboule and all members of the laboratory for advice and support, Mylène  
797 Docquier, Brice Petit and the Genomics Platform of the University of Geneva for RNA-seq data  
798 production, Amanda Cooksey and the Chickspress Team at the University of Arizona  
799 (<http://geneatlas.arl.arizona.edu>) for granting us access to chicken RNA-seq data, Ioannis Xenarios and  
800 the Vital-IT team for computational support. We thank Jean-Marc Matter (University of Geneva) for  
801 providing chicken eggs. This work was performed using the computing facilities of the CC LBBE/PRABI,  
802 the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics  
803 (<http://www.vital-it.ch>) and the computing center of the French National Institute of Nuclear and  
804 Particle Physics (CC-IN2P3). This project was funded by the Swiss National Science Foundation (SNSF  
805 Ambizione grant PZ00P3\_142636), the Agence Nationale pour la Recherche (ANR JJC 2017  
806 LncEvoSys). FD was supported by a FP7 IDEAL grant (259679).

807

808 **Figure legends**

809 **Figure 1. Comparative transcriptomics across species, organs and developmental stages.**

810 **A.** Experimental design. The developmental stages selected for mouse, rat and chicken are  
811 marked on a horizontal axis. Organs sampled for each species and developmental stage are shown  
812 below. Abbreviations: br, brain; kd, kidney; lv, liver; ts, testes.

813 **B.** First factorial map of a principal component analysis, performed on log<sub>2</sub>-transformed TPM  
814 values, for 10,363 protein-coding genes with orthologues in mouse, rat and chicken. Colors represent  
815 different organs and developmental stages, point shapes represent different species.

816 **C.** Hierarchical clustering, performed on a distance matrix derived from Spearman correlations  
817 between pairs of samples, for 10,363 protein-coding genes with orthologues in mouse, rat and chicken.  
818 Organ and developmental stages are color-coded, shown below the heatmap. Species of origin is color-  
819 coded, shown on the right. Sample clustering is shown on the left.

820

821 **Figure 2. Transcriptome complexity in different species, organs and developmental stages.**

822 **A.** Number of protein-coding genes supported by at least 10 uniquely mapped reads in each  
823 sample, after read resampling to homogenize coverage (Supplementary Methods). Colors represent  
824 different organs, point shapes represent different species. Developmental stages are indicated by  
825 numeric labels, 1 to 5, on the X-axis. We analyzed a total of 19,356 protein-coding genes in the mouse,  
826 19,274 in the rat and 15,509 in the chicken.

827 **D.** Number of lncRNAs supported by at least 10 uniquely mapped reads in each sample, after  
828 read resampling to homogenize coverage. We analyzed a total of 18,858 candidate lncRNAs in the  
829 mouse, 20,159 in the rat and 5,496 in the chicken.

830

831 **Figure 3. Different expression patterns for protein-coding genes and lncRNAs.**

832 **A.** Distribution of the organ in which maximum expression is observed, for protein-coding  
833 genes (pc) and lncRNAs (lnc), for mouse, rat and chicken. Organs are color-coded, shown above the  
834 plot. We defined the sample in which maximum expression is observed based on average expression  
835 values across replicates, for each organ / developmental stage combination (Supplementary Methods).

836 **B.** Distribution of the developmental stage in which maximum expression is observed, for  
837 protein-coding genes and lncRNAs, for mouse, rat and chicken. Developmental stages are color-coded,  
838 shown above the plot.

839 **C.** Percentage of protein-coding and lncRNA genes that are significantly (FDR<0.01)  
840 differentially expressed (DE) among developmental stages, with respect to the total number of genes  
841 tested for each organ. Left panel: differential expression analysis performed with all RNA-seq reads.

842 Right panel: differential expression analysis performed after down-sampling read counts for protein-  
843 coding genes, to match those of lncRNAs (Materials and methods).

844 **D.** Distribution of the developmental stage in which maximum expression is observed, for  
845 protein-coding genes and lncRNAs that are significantly DE (FDR<0.01) in each organ, for the mouse.  
846 Percentages are computed with respect to the total number of DE genes in each organ and gene class.

847

848 **Figure 4. Increased levels of long-term sequence conservation for lncRNAs expressed early**  
849 **in development.**

850 **A.** Sequence conservation scores (PhastCons scores, placental mammals) for protein-coding  
851 and lncRNAs exonic regions, for subsets of genes expressed above noise levels (TPM>=1) in each organ  
852 and developmental stage. Dots represent medians, vertical bars represent 95% confidence intervals.  
853 Numbers of analyzed genes are provided in Supplementary Table 5. Organs are color-coded;  
854 developmental stages are indicated (numbers 1 to 5) on the X-axis. The gray dots and vertical bars  
855 represent the median value and 95% confidence interval for all intergenic regions, genome-wide, or  
856 for intergenic regions flanking lncRNA loci (Supplementary Methods).

857 **B.** Same as A, for promoter regions (400 bp upstream of transcription start sites). Exonic  
858 sequences were masked before assessing conservation.

859 **C.** Same as B, for splice sites (first and last two bases of each intron).

860

861 **Figure 5. Orthologous lncRNA families for mouse, rat and chicken.**

862 **A.** Number of mouse protein-coding genes and lncRNAs in different classes of evolutionary  
863 conservation. From left to right: all loci, loci with conserved sequence in the rat, loci for which  
864 transcription could be detected (at least 10 unique reads) in predicted orthologous locus in the rat,  
865 loci with predicted 1-to-1 orthologues, loci for which the predicted orthologue belonged to the same  
866 class (protein-coding or lncRNA) in the rat, loci with conserved sequence in the chicken, loci for which  
867 transcription could be detected (at least 10 unique reads) in predicted orthologous locus in the  
868 chicken, loci with predicted 1-to-1 orthologues, loci for which the predicted orthologue belonged to  
869 the same class (protein-coding or lncRNA) in the chicken. We analyze 19,356 protein-coding genes and  
870 18,858 candidate lncRNAs in the mouse.

871 **B.** Distribution of the organ in which maximum expression is observed, for mouse protein-  
872 coding and lncRNA genes that have no orthologues in the rat or chicken, for genes with orthologues in  
873 the rat and for genes with orthologues in chicken. The sample in which maximum expression is  
874 observed is computed based on average expression values across biological replicates, for each organ  
875 / developmental stage combination (Supplementary Methods).

876 C. Same as B, for the distribution of the developmental stage in which maximum expression is  
877 observed.

878

879 **Figure 6. Global comparison of lncRNA expression patterns across species.**

880 A. First factorial map of a principal component analysis, performed on log<sub>2</sub>-transformed TPM  
881 values, for 2,893 orthologous lncRNAs between mouse and rat. Colors represent different organs and  
882 developmental stages, point types represent species.

883 B. Hierarchical clustering, performed on a distance matrix derived from Spearman correlations  
884 between pairs of samples, for 2,893 orthologous lncRNAs between mouse and rat. Organ and  
885 developmental stages are shown below the heatmap. Species of origin is shown on the right. Sample  
886 clustering is shown on the left.

887

888 **Figure 7. Global estimates of expression conservation across organs and developmental**  
889 **stages.**

890 A. Example of between-species and within-species variation of expression levels, for protein-  
891 coding genes (left) and lncRNAs (right), for orthologous genes between mouse and rat, for the mid-  
892 stage embryonic brain. Spearman's correlation coefficients ( $\rho$ ) are shown above each plot. We show  
893 a smoothed color density representation of the scatterplots, obtained through a 2D kernel density  
894 estimate (smoothScatter function in R).

895 B. Expression conservation index, defined as the ratio of the between-species and the within-  
896 species expression level correlation coefficients, for protein-coding genes, for each organ and  
897 developmental stage. The vertical segments represent minimum and maximum values obtained from  
898 100 bootstrap replicates. We analyzed 15,931 pairs of orthologous protein-coding genes.

899 C. Same as B, for lncRNAs. We analyzed 2,893 orthologous mouse and rat lncRNAs.

900

901 **Figure 8. Conservation of developmental expression patterns of differentially expressed**  
902 **lncRNAs.**

903 A. Comparison of the developmental stage in which maximum expression is observed, for  
904 orthologous lncRNAs that are significantly differentially expressed (FDR<0.01) among developmental  
905 stages, for both mouse and rat. The sample in which maximum expression is observed is computed  
906 based on average expression values across biological replicates, for each organ / developmental stage  
907 combination (Supplementary Methods). Genes are divided into classes based on the developmental  
908 stage where maximum expression is observed in mouse organs (X-axis). The Y axis represents the  
909 percentage of orthologous genes that reach maximum expression in each developmental stage, in the  
910 rat. Numbers of analyzed genes are shown below the plot.

911           **B.** Expression profiles of orthologous lncRNAs that are significantly differentially expressed  
912 (FDR<0.01) among developmental stages, for both mouse and rat, in the brain. TPM values were  
913 averaged across replicates and normalized by dividing by the maximum, for each species. The resulting  
914 relative expression profiles were combined across species and clustered with the K-means algorithm.  
915 Dots represent the average profiles of the genes belonging to each cluster. Gray lines represent profiles  
916 of individual genes from a cluster. Numbers of genes in each cluster are shown in the plot.

917           **C.** Same as B, for the kidney.

918           **D.** Same as B, for the liver.

919           **E.** Same as B, for the testes. For this organ, we searched for only 4 clusters with the K-means  
920 algorithm.

921

922           **Figure 9. Per-gene estimates of expression pattern divergence between species.**

923           **A.** Relationship between the per-gene expression divergence measure (Euclidean distance of  
924 relative expression profiles among organs/stages, between mouse and rat), and the average  
925 expression values (log<sub>2</sub>-transformed TPM) across all mouse and rat samples. We show a smoothed  
926 color density representation of the scatterplots, obtained through a 2D kernel density estimate  
927 (smoothScatter function in R). Red line: linear regression.

928           **B.** Distribution of the expression divergence value for all protein-coding and lncRNA genes with  
929 predicted 1-to-1 orthologues in mouse and rat.

930           **C.** Distribution of the residual expression divergence values, after regressing the average  
931 expression level, for protein-coding genes and lncRNAs.

932           **D.** Relationship between expression divergence and exonic sequence conservation (% exonic  
933 sequence aligned without gaps between mouse and rat), for protein-coding genes and lncRNAs.

934           **E.** Average contribution of each organ/developmental stage combination to expression  
935 divergence, for protein-coding genes and lncRNAs.

936

937

938           **References**

- 939   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*  
940       215: 403–410.
- 941   Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D. 2016. Hotair is dispensible for mouse  
942       development. *PLoS Genet* 12: e1006232.
- 943   Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* 11:  
944       R106.
- 945   Anderson KM, Anderson DM, McAnally JR, Shelton JM, Bassel-Duby R, Olson EN. 2016. Transcription  
946       of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*  
947       539: 433–436.
- 948   Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, Oshlack A, Smith CA. 2013.  
949       RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in  
950       chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol* 14.
- 951   Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. 2018. A TAD boundary is preserved upon  
952       deletion of the CTCF-rich Firre locus. *Nat Commun* 9: 1444.
- 953   Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, Ephrussi A, Ferguson-Smith AC,  
954       Gingeras TR, Haerty W, et al. 2014. Considerations when investigating lncRNA function in vivo.  
955       *eLife* 3: e03058.
- 956   Ben-Tabou de-Leon S, Davidson EH. 2007. Gene regulation: gene control network in development.  
957       *Annu Rev Biophys Biomol Struct* 36: 191.
- 958   Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson  
959       H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset  
960       aligner. *Genome Res* 14: 708–715.
- 961   Brannan CI, Dees EC, Ingram RS, Tilghman SM. 1990. The product of the H19 gene may function as an  
962       RNA. *Mol Cell Biol* 10: 28–36.
- 963   Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A,  
964       Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature*  
965       478: 343–348.
- 966   Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification.  
967       *Nat Biotechnol* 34: 525–527.
- 968   Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991. A gene from  
969       the region of the human X inactivation centre is expressed exclusively from the inactive X  
970       chromosome. *Nature* 349: 38–44.
- 971   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:  
972       architecture and applications. *BMC Bioinformatics* 10: 421.
- 973   Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascensão K, Rummel C,  
974       Ovchinnikova S, et al. 2019. Gene expression across mammalian organ development. *Nature*  
975       571: 505–509.

- 976 Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. 2016. The life history of retrocopies  
977 illuminates the evolution of new mammalian genes. *Genome Res* 26: 301–314.
- 978 Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Karolchik  
979 D, et al. 2018. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* 46: D762–  
980 D769.
- 981 Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. 2011.  
982 A long noncoding RNA controls muscle differentiation by functioning as a competing  
983 endogenous RNA. *Cell* 147: 358–369.
- 984 Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. 2014.  
985 Origins and functional evolution of Y chromosomes across mammals. *Nature* 508: 488–93.
- 986 Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu  
987 S, et al. 2019. Ensembl 2019. *Nucleic Acids Res* 47: D745–D751.
- 988 Doolittle WF. 2018. We simply cannot go on being so vague about “function.” *Genome Biol* 19: 223.
- 989 El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA,  
990 Smart A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47: D427–  
991 D432.
- 992 Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES.  
993 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing.  
994 *Nature* 539: 452–455.
- 995 Goudarzi M, Berg K, Pieper LM, Schier AF. 2019. Individual long non-coding RNAs have no overt  
996 functions in zebrafish embryogenesis, viability and fertility. *eLife* 8.
- 997 Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. 2013. On the immortality of television sets:  
998 “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome*  
999 *Biol Evol* 5: 578–590.
- 1000 Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB,  
1001 et al. 2018. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-  
1002 Seq. *Dev Cell* 46: 651-667.e10.
- 1003 Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E, Elcavage L, Plana O,  
1004 Sanchez LV, Lee JC, et al. 2016. In vivo characterization of Linc-p21 reveals functional cis-  
1005 regulatory DNA elements. *Cell Rep* 16: 2178–2186.
- 1006 Grote P, Herrmann BG. 2015. Long noncoding RNAs in organogenesis: making the difference. *Trends*  
1007 *Genet TIG* 31: 329–335.
- 1008 Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, et  
1009 al. 2013. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall  
1010 development in the mouse. *Dev Cell* 24: 206–214.
- 1011 Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et  
1012 al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs  
1013 in mammals. *Nature* 458: 223–227.

- 1014 Hacısuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG,  
1015 Sauvageau M, Kelley DR, et al. 2014. Topological organization of multichromosomal regions by  
1016 the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* 21: 198–206.
- 1017 Haerty W, Ponting CP. 2013. Mutations within lncRNAs are effectively selected against in fruitfly but  
1018 not in human. *Genome Biol* 14: R49.
- 1019 Haerty W, Ponting CP. 2014. No gene in the genome makes sense except in the light of evolution. *Annu*  
1020 *Rev Genomics Hum Genet* 15: 71–92.
- 1021 Haerty W, Ponting CP. 2015. Unexpected selection to retain high GC content and splicing enhancers  
1022 within exons of multiexonic lncRNA loci. *RNA N Y N* 21: 333–346.
- 1023 Hamburger V, Hamilton HL. 1951. A series of normal stages in the development of the chick embryo. *J*  
1024 *Morphol* 88: 49–92.
- 1025 Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding  
1026 RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* 11:  
1027 1110–1122.
- 1028 Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S,  
1029 et al. 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47:  
1030 199–208.
- 1031 Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res* 20: 1313–  
1032 1326.
- 1033 Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013.  
1034 Transposable elements are major contributors to the origin, diversification, and regulation of  
1035 vertebrate long noncoding RNAs. *PLoS Genet* 9: e1003470.
- 1036 Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein  
1037 BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate  
1038 with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 106:  
1039 11667–11672.
- 1040 Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements.  
1041 *Nat Methods* 12: 357–60.
- 1042 Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, Odom DT, Marques AC. 2012. Rapid  
1043 turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet* 8:  
1044 e1002841.
- 1045 Latos PA, Pauler FM, Koerner MV, Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement  
1046 RM, Warczok KE, et al. 2012. Airn transcriptional overlap, but not its lncRNA products, induces  
1047 imprinted Igf2r silencing. *Science* 338: 1469–1472.
- 1048 Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for  
1049 alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.*
- 1050 Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, Park S, Wan KH, Schroeder AJ, Gramates LS, St  
1051 Pierre SE, et al. 2007. Revisiting the protein-coding gene catalog of *Drosophila melanogaster*  
1052 using 12 fly genomes. *Genome Res* 17: 1823–1836.

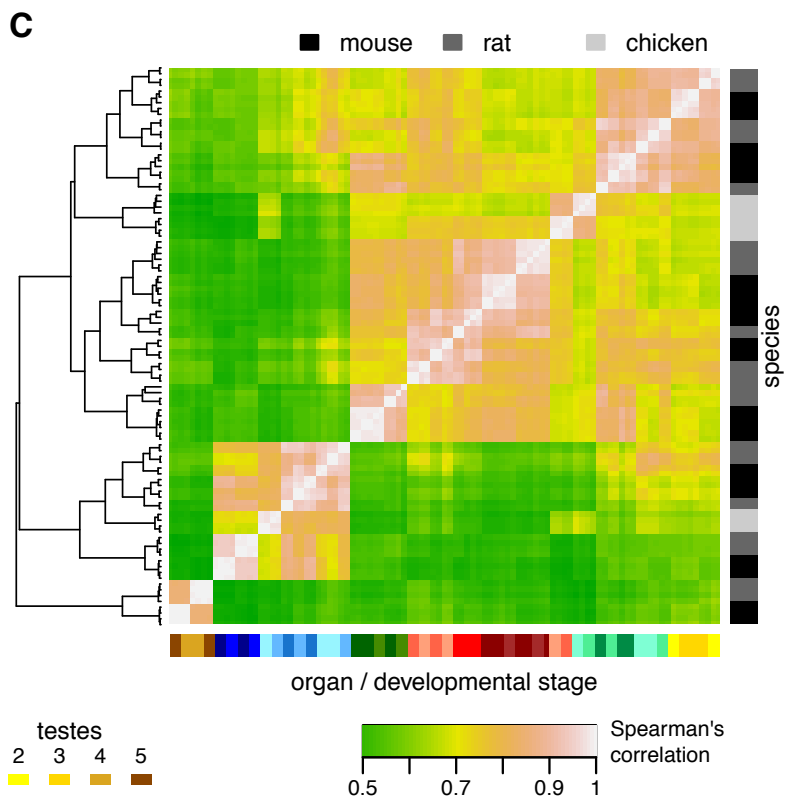
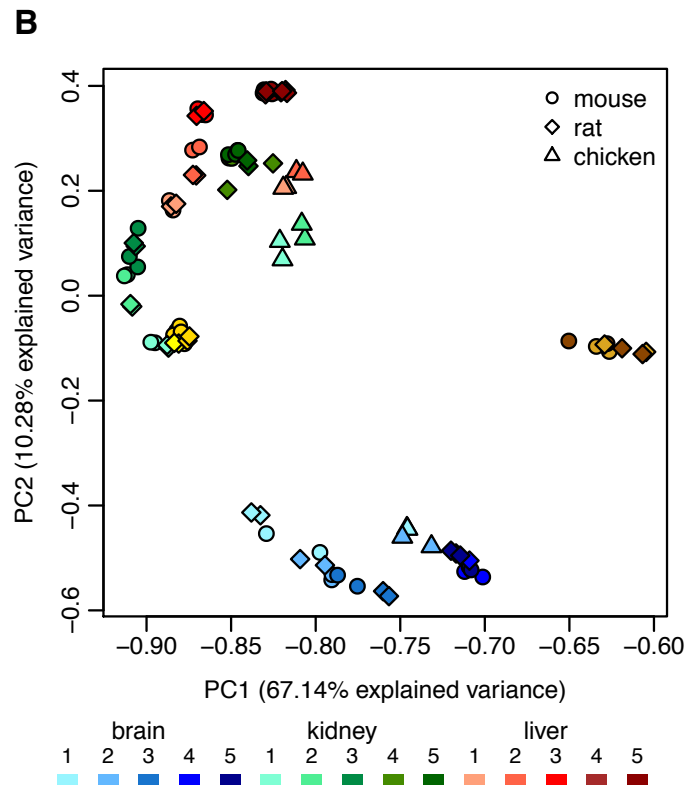
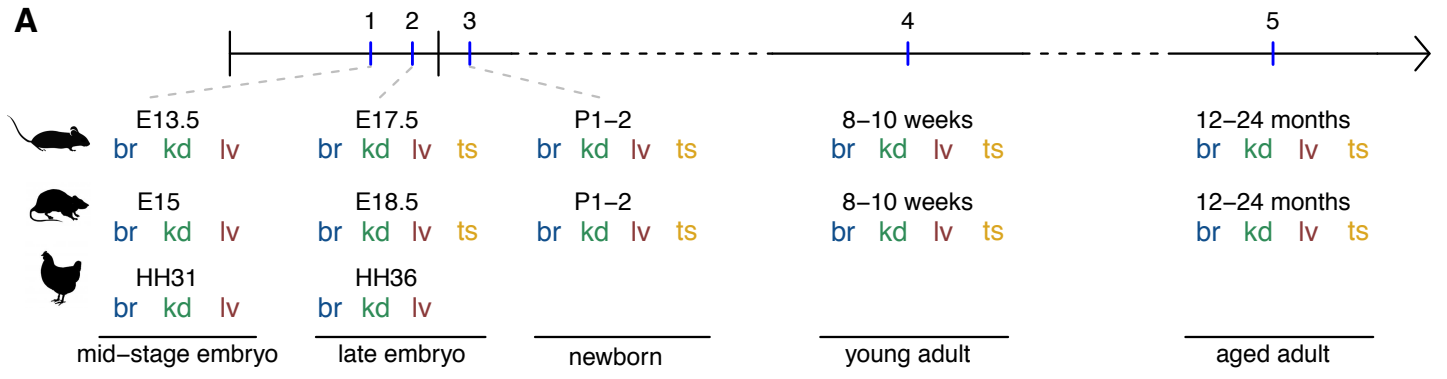


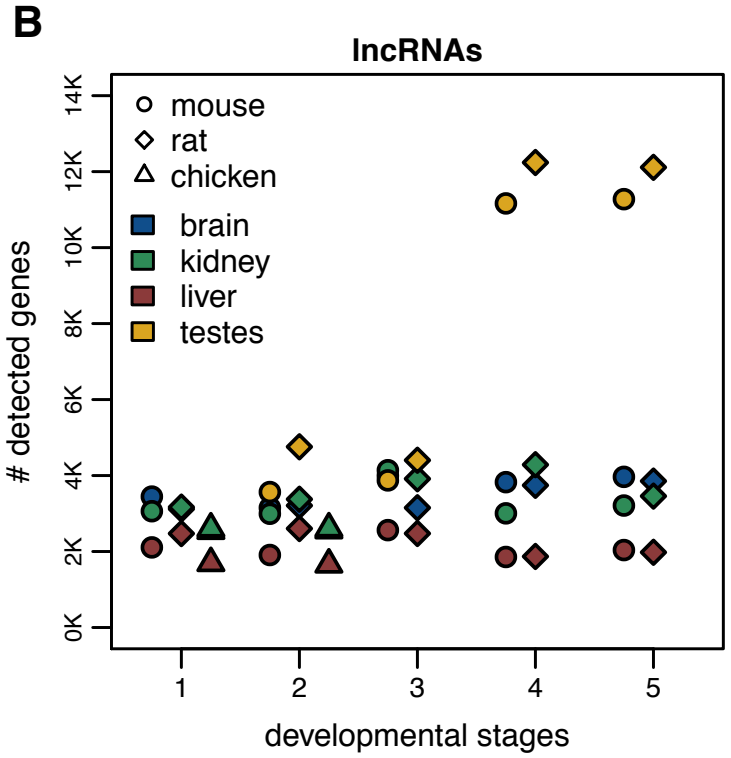
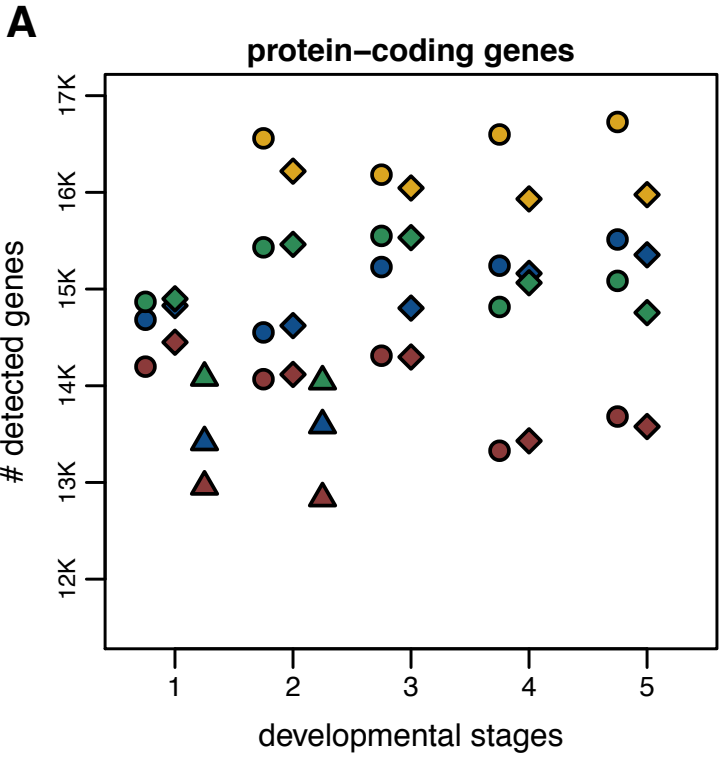
- 1053 Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein  
1054 coding and non-coding regions. *Bioinforma Oxf Engl* 27: i275-282.
- 1055 Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, He D, Weissman JS, Kriegstein AR,  
1056 Diaz AA, et al. 2016. Single-cell analysis of long non-coding RNAs in the developing human  
1057 neocortex. *Genome Biol* 17: 67.
- 1058 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq  
1059 data with DESeq2. *Genome Biol* 15: 550.
- 1060 Marques AC, Hughes J, Graham B, Kowalczyk MS, Higgs DR, Ponting CP. 2013. Chromatin signatures at  
1061 transcriptional start sites separate two equally populated yet distinct classes of intergenic long  
1062 noncoding RNAs. *Genome Biol* 14: R131.
- 1063 McCarthy FM, Pendarvis K, Cooksey AM, Gresham CR, Bomhoff M, Davey S, Lyons E, Sonstegard TS,  
1064 Bridges SM, Burgess SC. 2019. Chickspress: a resource for chicken gene expression. *Database*  
1065 2019. <https://academic.oup.com/database/article/doi/10.1093/database/baz058/5512474>  
1066 (Accessed August 11, 2019).
- 1067 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*  
1068 351: 652–654.
- 1069 Munschauer M, Nguyen CT, Sirokman K, Hartigan CR, Hogstrom L, Engreitz JM, Ulirsch JC, Fulco CP,  
1070 Subramanian V, Chen J, et al. 2018. The NORAD lncRNA assembles a topoisomerase complex  
1071 critical for genome stability. *Nature* 561: 132–136.
- 1072 Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Grutzner F, Kaessmann H. 2014. The  
1073 evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505: 635–640.
- 1074 Nel-Themaat L, Gonzalez G, Akiyama H, Behringer RR. 2010. Illuminating testis morphogenesis in the  
1075 mouse. *J Androl* 31: 5–10.
- 1076 Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C,  
1077 Huang Q, et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell*  
1078 143: 46–58.
- 1079 Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables  
1080 improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33: 290–5.
- 1081 Perteau M, Shumate A, Perteau G, Varabyou A, Breitwieser FP, Chang Y-C, Madugundu AK, Pandey A,  
1082 Salzberg SL. 2018. CHESS: a new human gene catalog curated from thousands of large-scale  
1083 RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol* 19: 208.
- 1084 Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection  
1085 within long noncoding RNAs. *Genome Res* 17: 556–65.
- 1086 R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. [https://www.R-](https://www.R-project.org/)  
1087 [project.org/](https://www.R-project.org/).
- 1088 Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ,  
1089 Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human  
1090 HOX loci by noncoding RNAs. *Cell* 129: 1311–1323.

- 1091 Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. 2019. Developmental dynamics of lincRNAs  
1092 across mammalian organs and species. *Nature* 571: 510–514.
- 1093 Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacısuleyman  
1094 E, Li E, Spence M, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required  
1095 for life and brain development. *eLife* 2: e01749.
- 1096 Schüler A, Ghanbarian AT, Hurst LD. 2014. Purifying selection on splice-related motifs, not expression  
1097 level nor RNA folding, explains nearly all constraint on human lincRNAs. *Mol Biol Evol* 31: 3164–  
1098 3183.
- 1099 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW,  
1100 Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and  
1101 yeast genomes. *Genome Res* 15: 1034–50.
- 1102 Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A. 2009. BioMart--  
1103 biological queries made easy. *BMC Genomics* 10: 22.
- 1104 Smit AFA, Hubley R, Green P. 2003. *RepeatMasker Open-4.0*. <http://www.repeatmasker.org>.
- 1105 Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke  
1106 A, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the  
1107 mammalian testis. *Cell Rep* 3: 2179–2190.
- 1108 Tabula Muris Consortium. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.  
1109 *Nature* 562: 367–372.
- 1110 Theiler K. 1989. *The house mouse: atlas of embryonic development*. Springer-Verlag, Berlin Heidelberg  
1111 <https://www.springer.com/la/book/9783642884207>.
- 1112 The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45:  
1113 D158–D169.
- 1114 Uebbing S, Konzer A, Xu L, Backström N, Brunström B, Bergquist J, Ellegren H. 2015. Quantitative mass  
1115 spectrometry reveals partial translational regulation for dosage compensation in chicken. *Mol*  
1116 *Biol Evol* 32: 2716–2725.
- 1117 Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding  
1118 RNAs. *Nat Rev Genet* 17: 601–614.
- 1119 Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate  
1120 embryonic development despite rapid sequence evolution. *Cell* 147: 1537–1550.
- 1121 Vendramin R, Verheyden Y, Ishikawa H, Goedert L, Nicolas E, Saraf K, Armaos A, Delli Ponti R,  
1122 Izumikawa K, Mestdagh P, et al. 2018. SAMMSON fosters cancer cell fitness by concertedly  
1123 enhancing mitochondrial and cytosolic translation. *Nat Struct Mol Biol* 25: 1035–1046.
- 1124 Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT,  
1125 Jasinska AJ, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* 160: 554–566.
- 1126 Washietl S, Kellis M, Garber M. 2014. Evolutionary dynamics and tissue specificity of human long  
1127 noncoding RNAs in six mammals. *Genome Res* 24: 616–28.

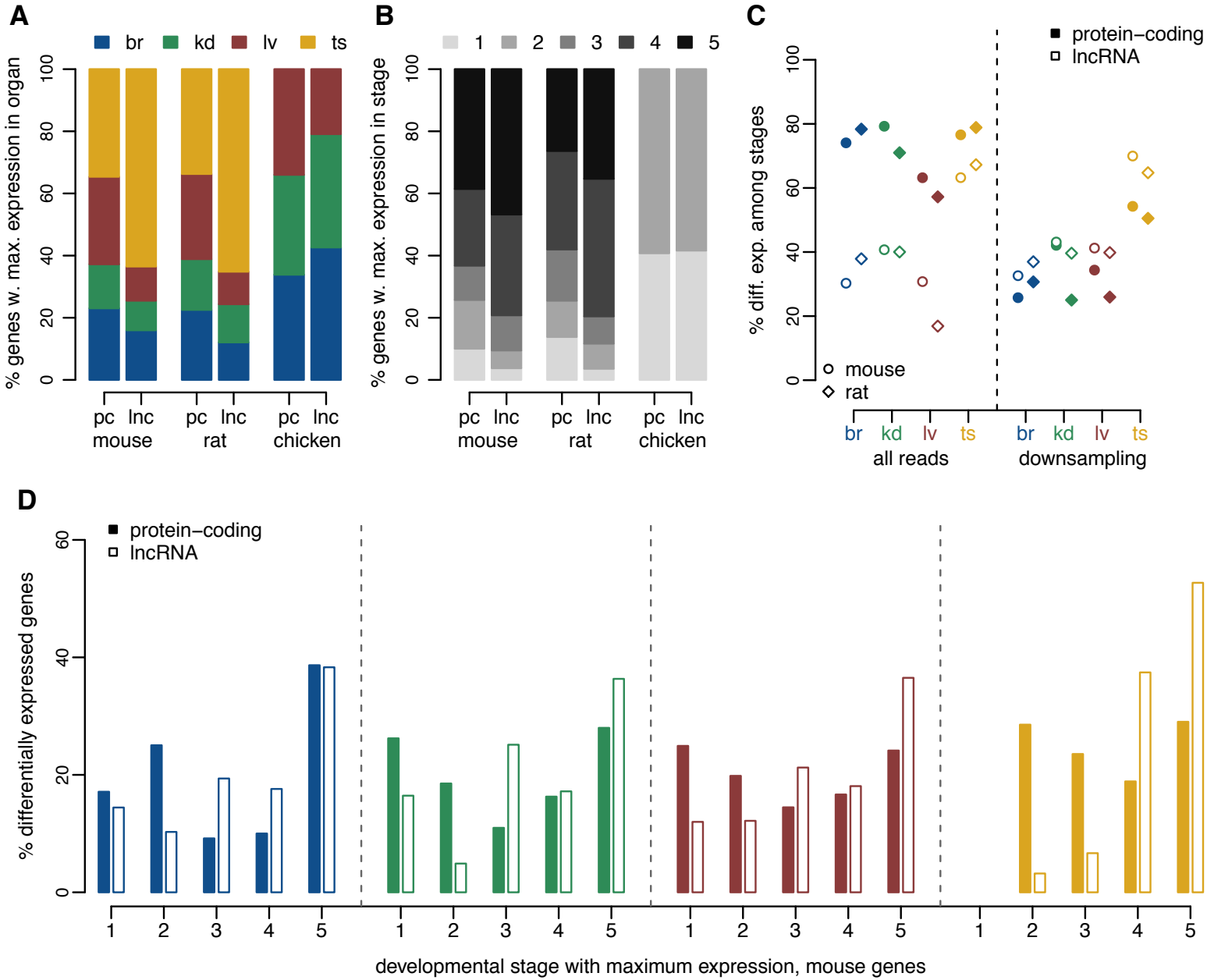
1128 Zakany J, Darbellay F, Mascrez B, Necsulea A, Duboule D. 2017. Control of growth and gut maturation  
1129 by HoxD genes and the associated lncRNA Haglr. *Proc Natl Acad Sci U S A* 114: E9290–E9299.  
1130

**Darbellay and Necselea, Figure 1**

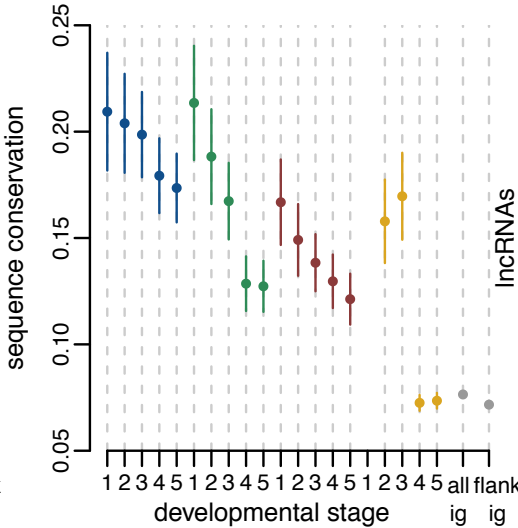
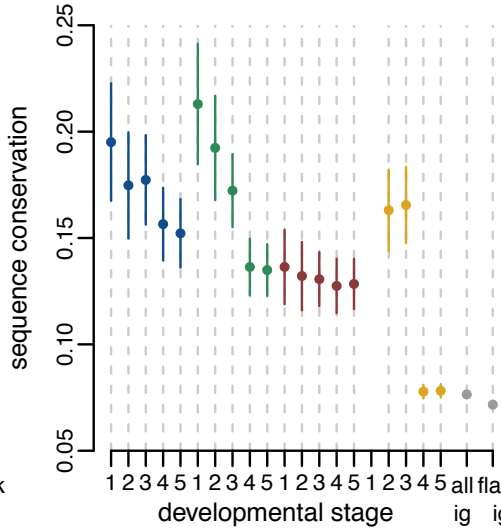
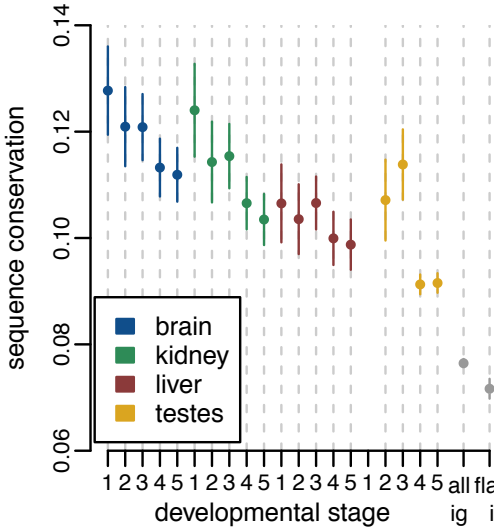
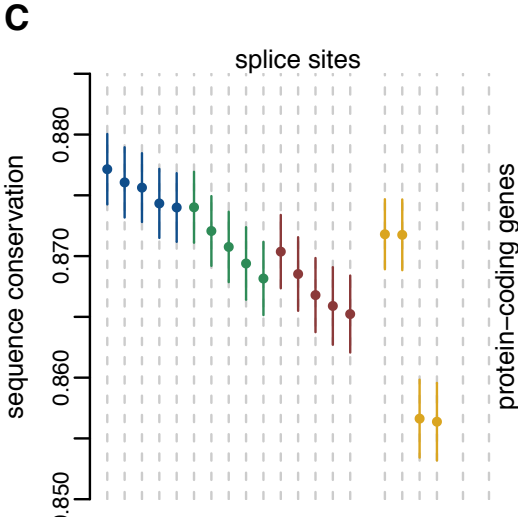
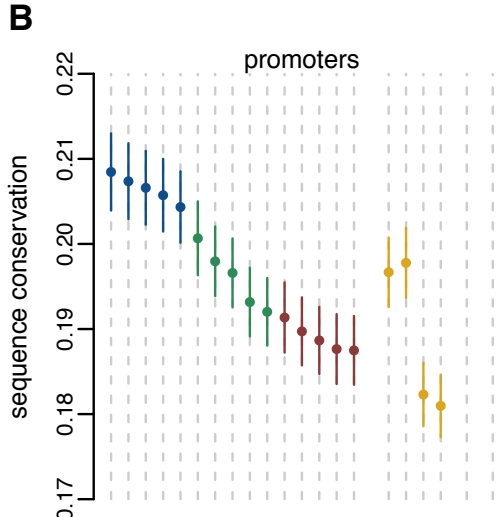
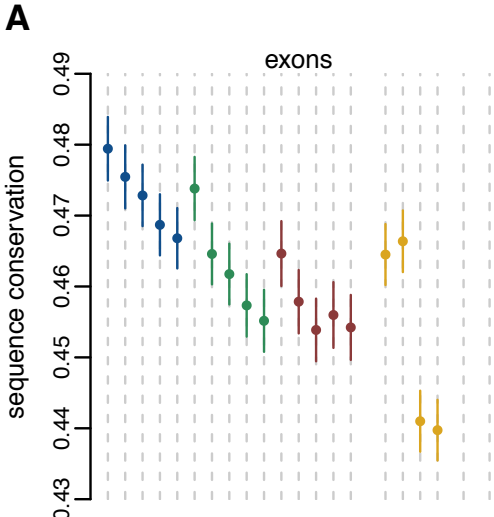




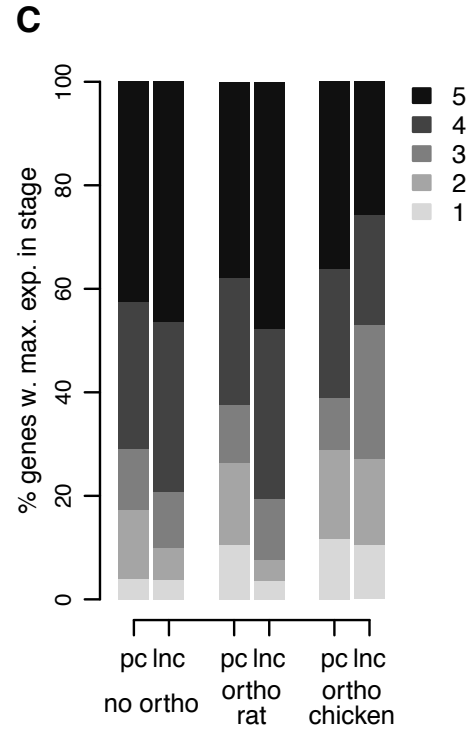
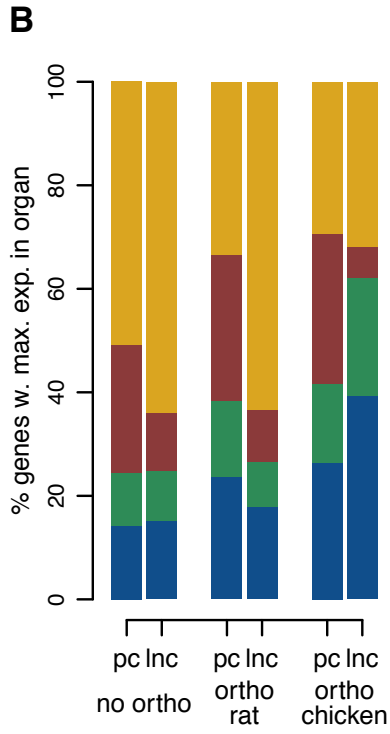
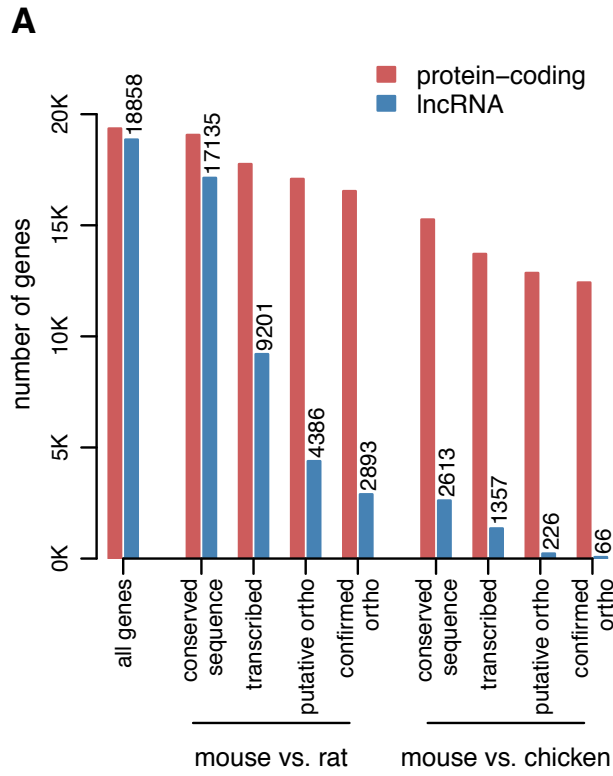
Darbellay and Neacsulea, Figure 3



Darbelay and Necsulea, Figure 4

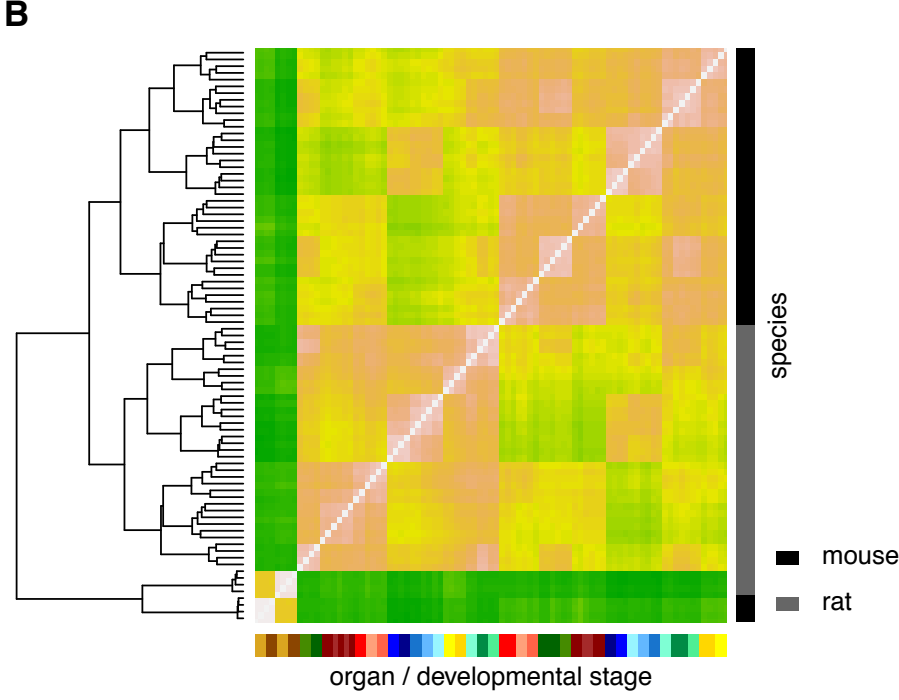
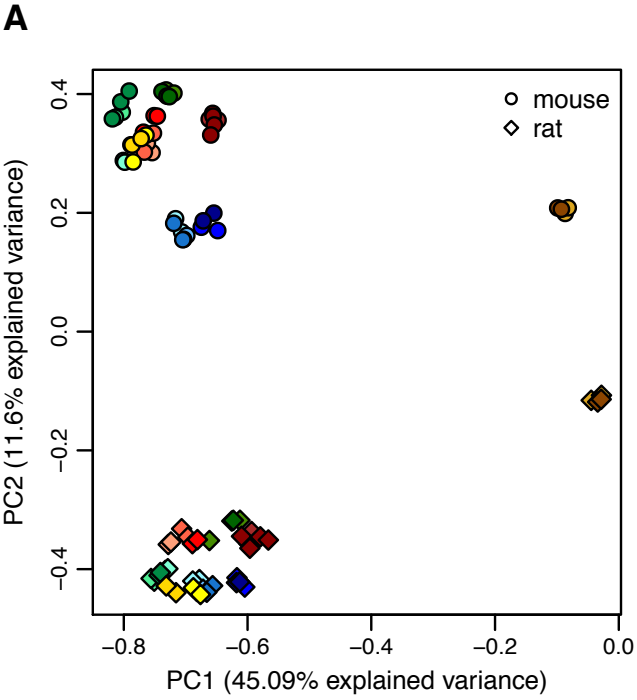


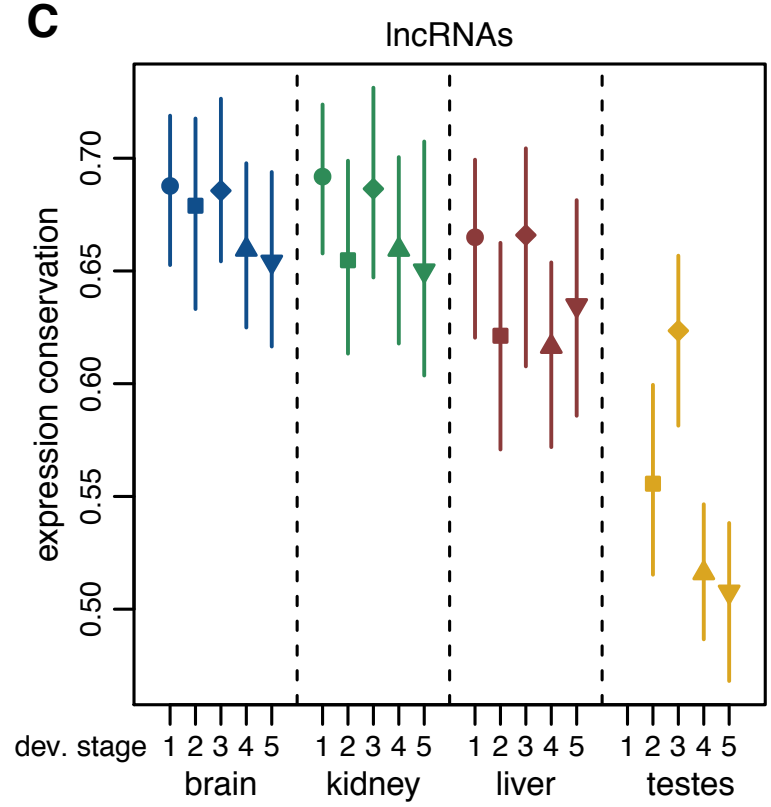
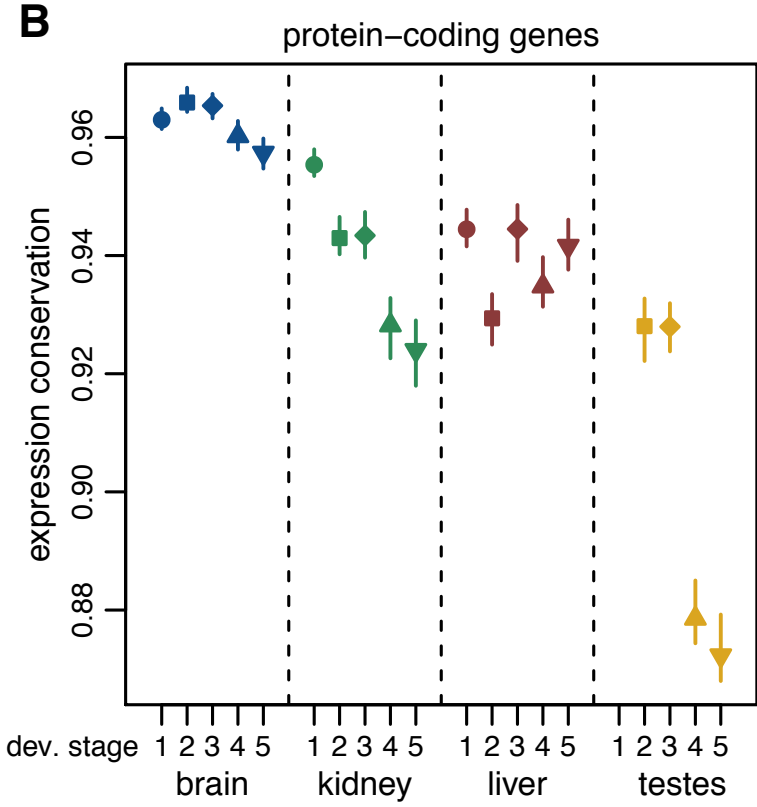
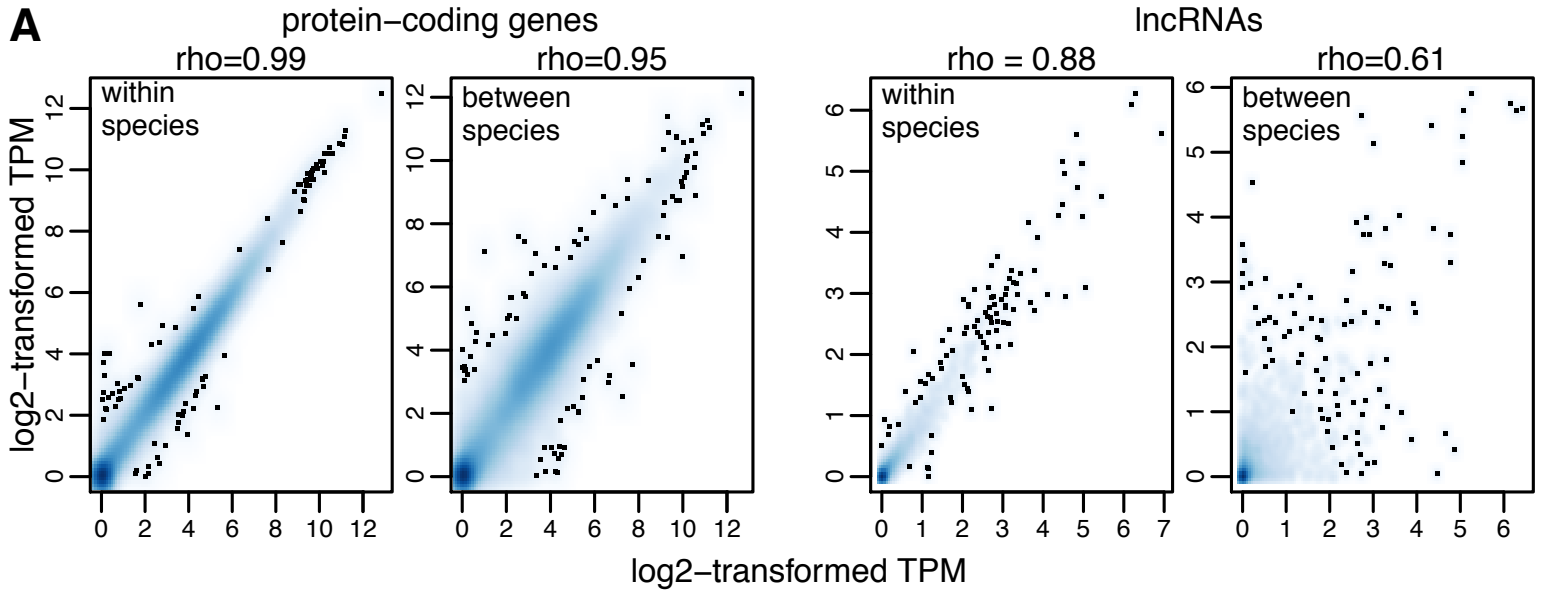
Darbella and Neacsulea, Figure 5





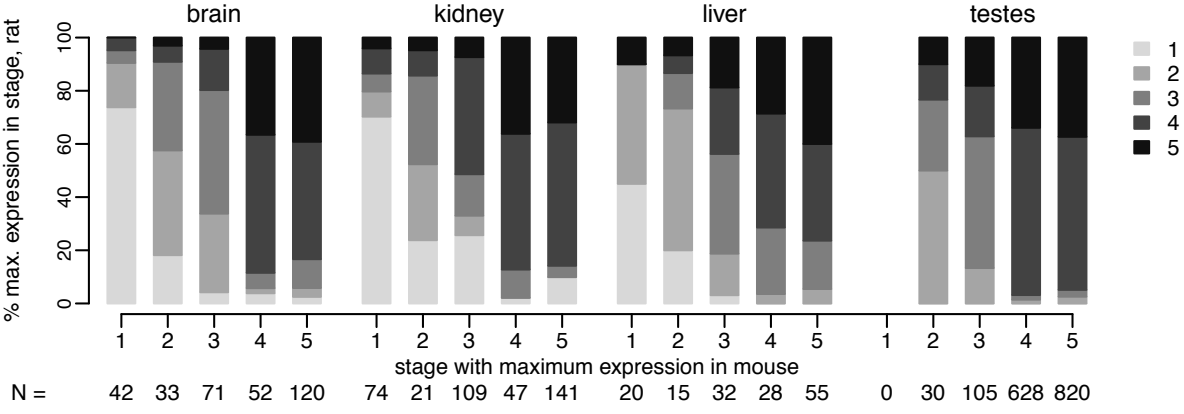
Darbellay and Neacsulea, Figure 6



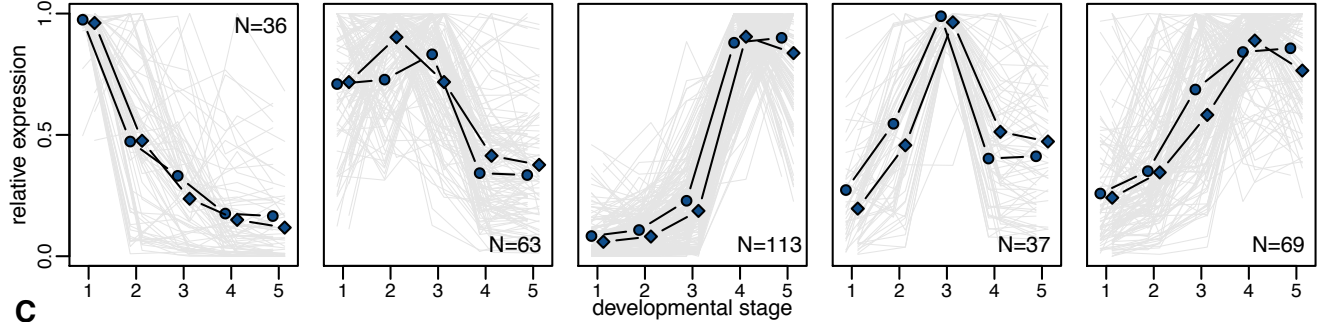


**Darbelay and Necsulea, Figure 8**

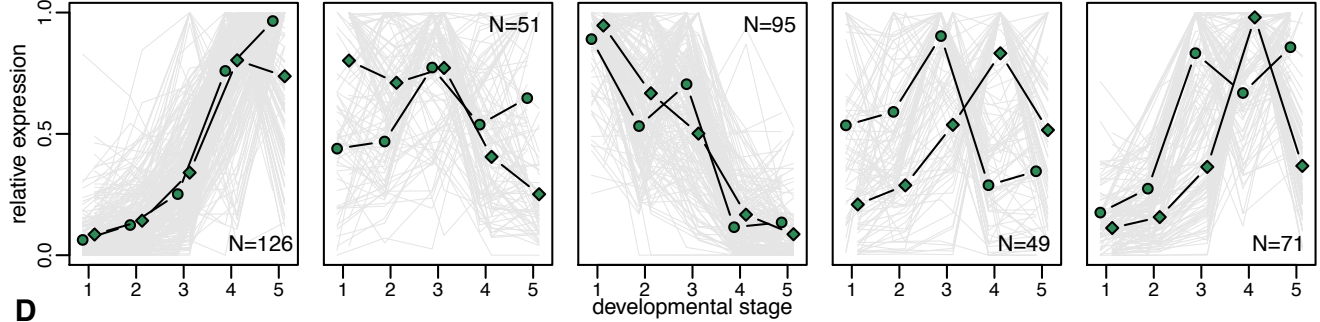
**A**



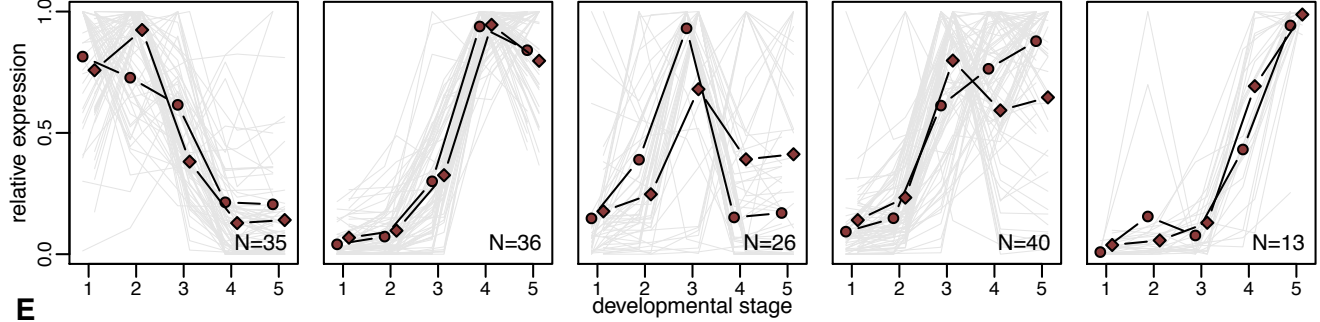
**B**



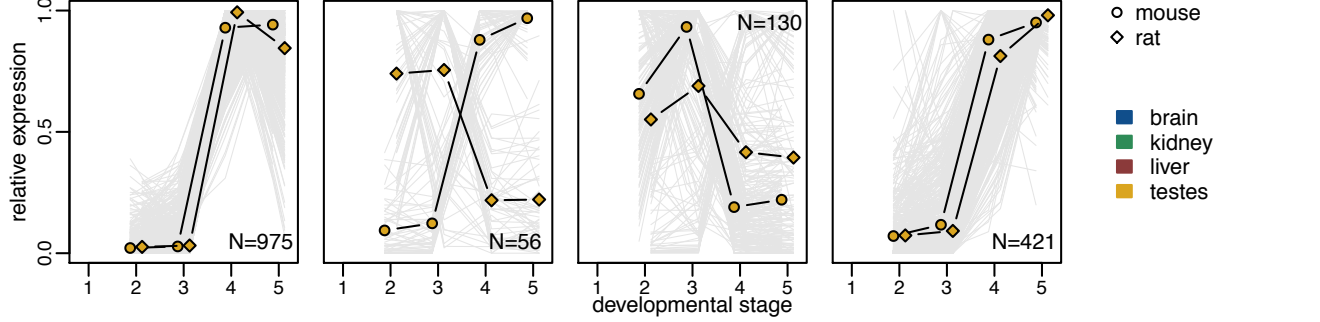
**C**



**D**



**E**



○ mouse  
◇ rat

■ brain  
■ kidney  
■ liver  
■ testes

Darbellay and Necselea, Figure 9

