Supplemental Online Material

Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period

Tena JJ*, González-Aguilera C*, Fernández-Miñán A, Vázquez-Marín J, Parra-Acero H, Cross JW, Rigby PWJ, Carvajal JJ, Wittbrodt J, Gómez-Skarmeta JL#, Martínez-Morales JR#.

Supplemental Data includes: 7 Supplementary Figures, 1 Supplementary movie, 6 Supplementary tables, Supplementary methods and References.

Supplementary Figure Legends:

Figure S1. Comparative anatomy of medaka (A) and zebrafish (B) at 4 additional comparable stages, two before and two after our reference stage (stg 24 in medaka / 24 hpf in zebrafish; at the center). Time-lapse analysis of embryonic movements shows that whereas stage 24 (44 hpf) medaka embryos are immobile (C), spontaneous contractions occur in zebrafish at 24 hpf (D). See also Supplementary Movie 1.

Figure S2. Analysis of the expression levels of four vertebrates at the selected phylotypic stage. (A) Heat map of Pearson correlation coefficients of zebrafish (*D. rerio*), medaka (*O. latipes*), xenopus (*X. tropicalis*) and mouse (*M. musculus*). Correlations were calculated based on the expression levels ($\log_2$ RPKM) of the 7118 orthologs with expression higher than 1 count per million reads (c.p.m.). Numbers within the cells indicate correlation values. The average of two independent replicates was used for the analyses. (B) Comparison of the expression levels for additional tissues in medaka and zebrafish (see also Figure 2). Bottom and top of boxes indicate percentile 25th and 75th, respectively, and lines in the boxes indicate medians. Whiskers indicate the lowest or the highest data point within 1.5x interquartile range from the box. (C) Pairwise analysis of the mRNA levels in different tissues in the four species described above. *$p<0.05$, **$p<10^{-3}$, ***$p<10^{-4}$ (Wilcoxon rank sum test).

Figure S3. Characterization of SPARRs and OPARRs in *D. rerio* and *O. latipes*. (A) Venn diagrams representing the genes associated with SPARRs identified in zebrafish and

medaka. (B) Frequency distribution of medaka genes (in %) associated with either all H3K27ac peaks (ALL), OPARRs, SPARRs, and cSPARRs according to the number of H3K27ac regulatory regions included in their vicinity. Genes associated with cSPARRs harbor more H3K27ac peaks than any other subset of genes. Similar results were obtained with OPARRs, SPARRs and cSPARRs identified in zebrafish (data not shown). (C) Venn diagrams representing orthologous genes associated with SPARRs and OPARRs in medaka and zebrafish. Note that a significant fraction of genes associated with SPARRs are associated with OPARRs in both species.

Figure S4. Assay validation using transgenic mouse embryos. (A) Tissue-specific expression patterns driven by representative human SPARRs homologs, as reported in the VISTA Enhancer Browser database. (B) Orthologous regions in Xenopus and zebrafish of one of the human SPARRs homologs, hs1327, drive similar expression patterns in each model organism.

Figure S5. SPARRs inter-species validation using transgenic assays. Mouse embryos show tissue-specific expression patterns, driven by representative human SPARRs homologs, as reported in the VISTA Enhancer Browser database. The orthologous regions of these human SPARRs were tested in zebrafish (A, B). All the regions drive similar expression patterns in mice and zebrafish. A group of them, hs73, hs1315 and hs1327 were further validated in medaka (B), where they also drive similar expression patterns to those observed in mice and zebrafish. fb, forebrain; hb, hindbrain; mb, midbrain; sc, spinal chord; nc, neural crest. The white discontinuous circles mark the eyes.

Figure S6. Validation of H3K27ac peaks by qPCR. Relative input recovery (in %) for representative H3K27ac peaks was measured from three independent ChIP experiments in medaka and zebrafish. Specific primers for four different peaks of each category (SPARRs, OPARRs from medaka and OPARRs from zebrafish) were assayed. Error bars correspond to standard deviation values (n=3).

Figure S7. H3K27ac ChIP-seq reproducibility. (A) UCSC tracks corresponding to both replicate samples of H3K27ac ChIP-seq show nearly identical profiles. (B) The

correlation, in 1 kb windows throughout the genome, between the number of reads from the two replicate yields a Pearson correlation coefficient of 0.97.

Movie S1. The time-lapse analysis shows spontaneous contractions of the trunk and the tail in 24 hpf zebrafish embryos, but not in stage 24 medaka embryos.

Table S1. Comparative staging of representative anatomical landmarks achieved during development in zebrafish (*D. rerio*) and/or medaka (*O. latipes*) embryos.

Table S2. Comparative transcriptomic analysis. The table includes:
- A list of Ensembl IDs for orthologous genes in the four analyzed species. *D. rerio*, *O. latipes*, *X. tropicalis* and *M. musculus*.
- A list of zebrafish orthologous genes (Ensembl ID) showing their differential expression vs medaka. Log Fold Change (logFC); P Values and Fold Discovery Rate (FRD) are shown.
- A list of tissue-specific genes in zebrafish (Ensembl ID and ZFIN gene symbol are shown), as annotated in ZFIN.
- Lists including significantly enriched Gene Ontology terms (Biological Process Function) according to DAVID package for differentially expressed genes (up or down-regulated) in medaka vs zebrafish.
- Lists including significantly enriched Gene Ontology terms (Biological Process Function) according to PANTHER package for differentially expressed genes (up or down-regulated) in medaka vs zebrafish.

Table S3. Hatching/delivering time, generation time and spawn size is listed for the four species analyzed. *D. rerio*, *O. latipes*, *X. tropicalis* and *M. musculus*

Table S4. Comparative epigenomic analysis. Lists of H3K27ac peaks obtained from medaka and zebrafish. The table includes:
- Summary of the number of peaks obtained for each category.
- H3K27ac peaks from medaka also acetylated in zebrafish (medaka SPARRs).
- H3K27ac peaks from zebrafish also acetylated in medaka (zebrafish SPARRs).
- H3K27ac peaks from medaka not acetylated in zebrafish (medaka OPARRs).
- H3K27ac peaks from zebrafish not acetylated in medaka (zebrafish OPARRs).

Each list is composed of five columns separated by tabs: chromosome, peak start, peak end, peak name and -10*log10 (p-value).

Table S5. H3K27ac peaks from fish that also operate as enhancers in mammals. The table includes:
- Number and definition of the different types of peaks considered.
- Zebrafish SPARRs conserved in human tested in mouse transgenesis assays in Vista Enhancer Browser.
- Medaka SPARRs conserved in human tested in mouse transgenesis assays in Vista Enhancer Browser.
In both lists, the position of the orthologous region in human (chromosome: start-end) links to the webpage with pictures of transgenic mice. Highlighted in green are the peaks that overlap an H3K27ac region in at least one human differentiated-cell type (Xie et al., 2013).

Table S6. Analysis of genes associated with cSPARR regions. The table includes:
- A list of medaka and zebrafish orthologous genes associated with cSPARR regions. Ensembl IDs, gene symbols and gene descriptions are provided.
- Two lists with significantly enriched Gene Ontology terms (Biological Process Function) for genes associated with all H3K27Ac peaks in medaka and zebrafish.
- Lists with significantly enriched Gene Ontology terms (Biological Process Function) according to DAVID and PANTHER analyses, and protein domains (Interpro) for genes associated with cSPARR regions.

Supplementary Methods and references.

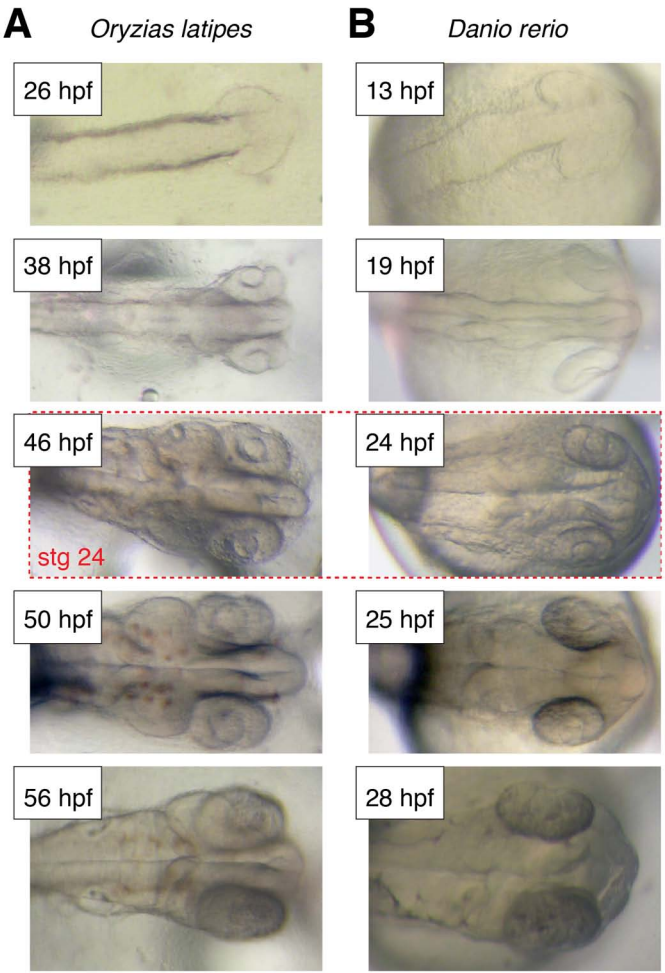
Transient transgenesis assays. Genomic fragments containing the zebrafish and medaka orthologous regions were amplified with primers described below. PCR fragments were subcloned in PCR8/GW/TOPO vector and transferred, through recombination using Gateway technology, to the ZED destination vector for zebrafish transgenesis [1], or to a I-SceI transgenesis specific vector for medaka [2]. To generate zebrafish transgenic embryos, we used Tol2 transposon/transposase method [3]. Medaka transgenic embryos were generated by the I-SceI endonuclease method [2].

Transient transgenic embryos were observed and photographed under a fluorescence scope.

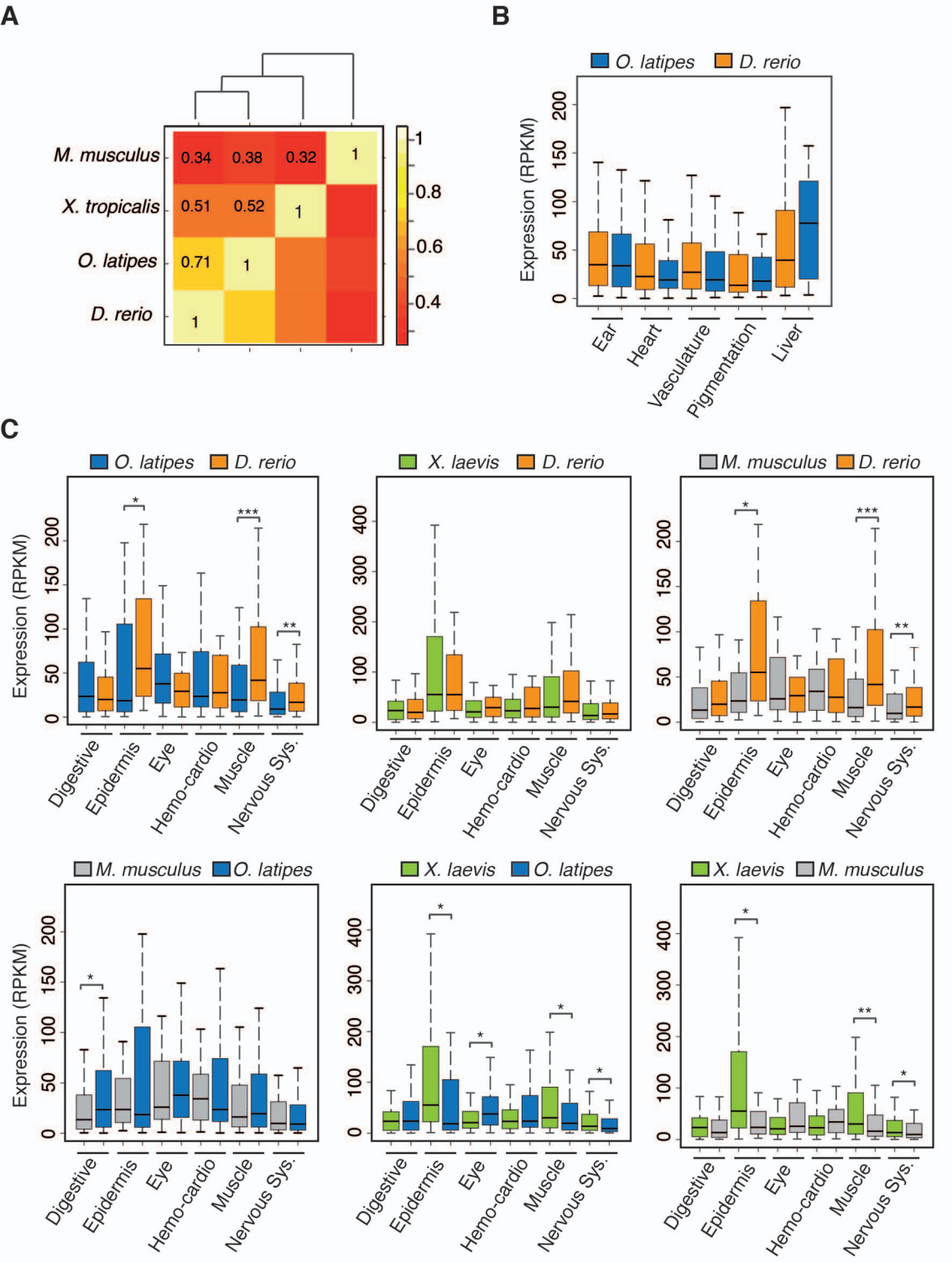
Primers used:
Z1327_F: GGCCTGTTTGGTTTTCG
Z1327_R: GCCCCGACCCACAC
Z1315_F: TGACTTGCCTAATTACCCTAACC
Z1315_R: GGACTTAAGTAGTTCCATCTTCAACC
Z625_F: ATGCTTCTGGTGGCATAACC
Z625_R: TCGACAGAAGACCAGAACACC
Z619_F: AAATGTGTTGTGGCATCTCC
Z619_R: CAGGTTTGGAAACGTATGAGG
Z969_F: ATGTGCCTGTCAGAGTGTGC
Z969_R: AAATATTAATTACTGCCATCATGACC
Z73_F: TACCACCCTCAGCCTTAACC
Z73_R: GGCAGTTCCCATTTACTTGC
M1327_F: GAGGATGCATCTCACTTCTGC
M1327_R: AGAAGCGTCCACAAGACAGC
M1315_F: CCATGCAATATTTGACAACTGG
M1315_R: TGGGACCTACAATGCAAACC
M73_F: AGCAGGGAGGATTAGGAAGG
M73_R: CAGTAAACAGATTCACACCAATCC

1. Bessa J, Tena JJ, de la Calle-Mustienes E, Fernandez-Minan A, Naranjo S, et al. (2009) Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* 238: 2409–2417.
2. Thermes V, Grabher C, Ristoratore F, Bourrat F, Choulika A, Wittbrodt J, and Joly JS (2002). I-SceI meganuclease mediates highly efficient transgénesis in fish. *Mech. Dev*. 118, 91–98.
3. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, et al. (2004) A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell* 7: 133–144.

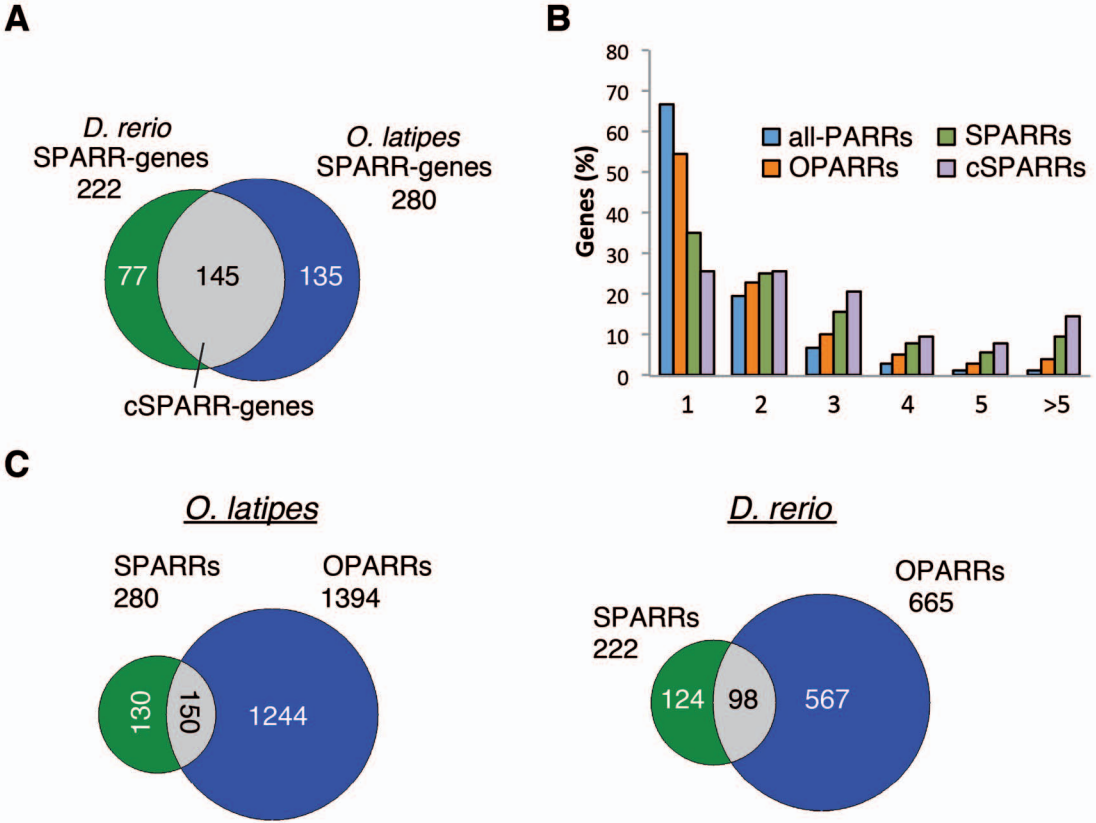
Tena_FigS1


A
Oryzias latipes
B
Danio rerio
C
D
26 hpf
38 hpf
46 hpf
stg 24
50 hpf
56 hpf
13 hpf
19 hpf
24 hpf
25 hpf
28 hpf
O. latipes
t=0 sec
t=10 sec
t=20 sec
D. rerio
t=0 sec
t=10 sec
t=20 sec

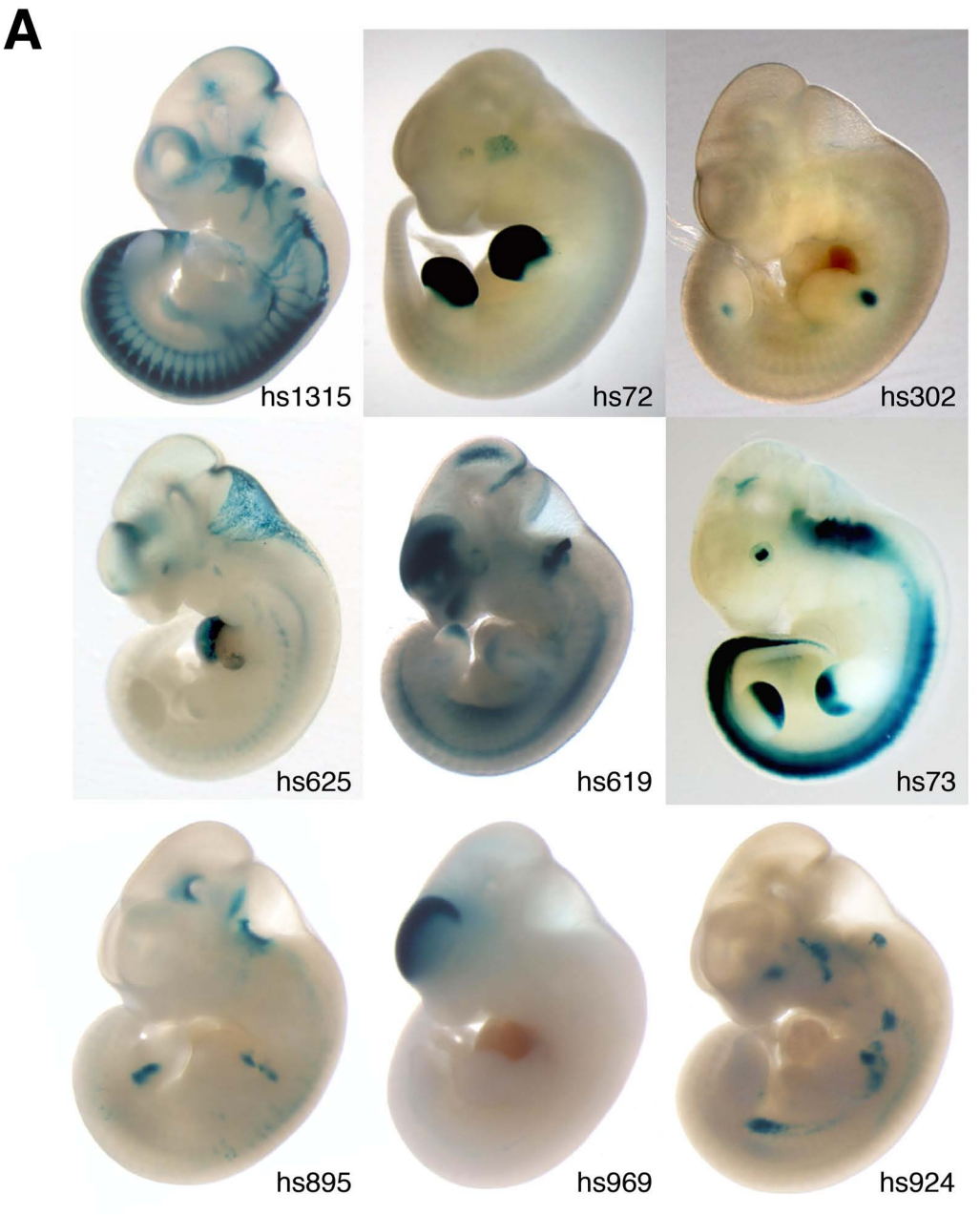
Tena_FigS2


A
M. musculus
X. tropicalis
O. latipes
D. rerio
0.34
0.38
0.32
1
0.51
0.52
1
0.71
1
1
1
0.8
0.6
0.4
B
O. latipes
D. rerio
Expression (RPKM)
0
50
100
150
200
Ear
Heart
Vasculature
Pigmentation
Liver
C
O. latipes
D. rerio
X. laevis
D. rerio
M. musculus
D. rerio
M. musculus
O. latipes
X. laevis
O. latipes
X. laevis
M. musculus
Expression (RPKM)
Digestive
Epidermis
Eye
Hemo-cardio
Muscle
Nervous Sys.

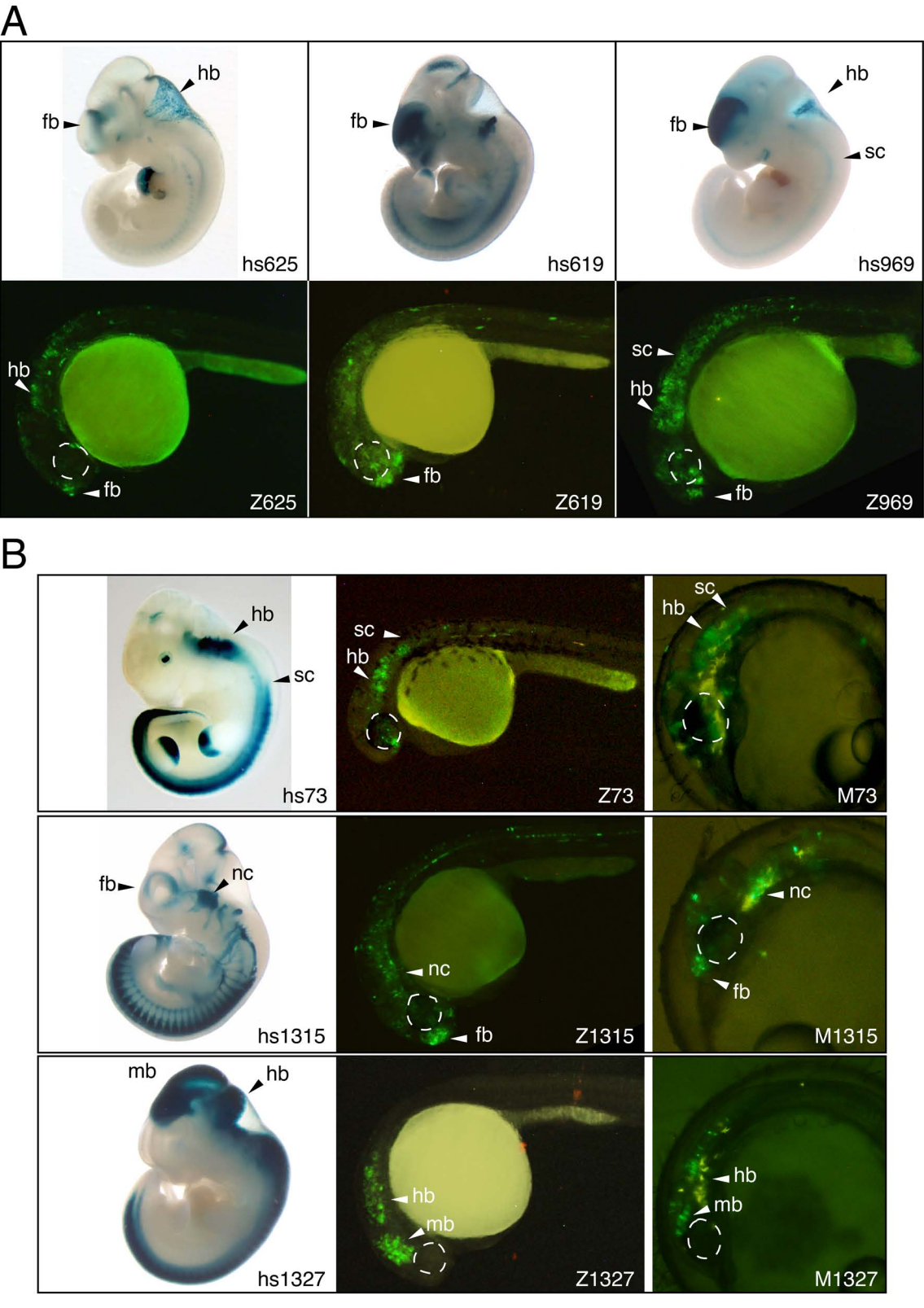
Tena_FigS3


A
D. rerio
SPARR-genes
222
O. latipes
SPARR-genes
280
77
145
135
cSPARR-genes
B
Genes (%)
80
70
60
50
40
30
20
10
0
all-PARRs
SPARRs
OPARRs
cSPARRs
1
2
3
4
5
>5
C
O. latipes
SPARRs
280
OPARRs
1394
130
150
1244
D. rerio
SPARRs
222
OPARRs
665
124
98
567

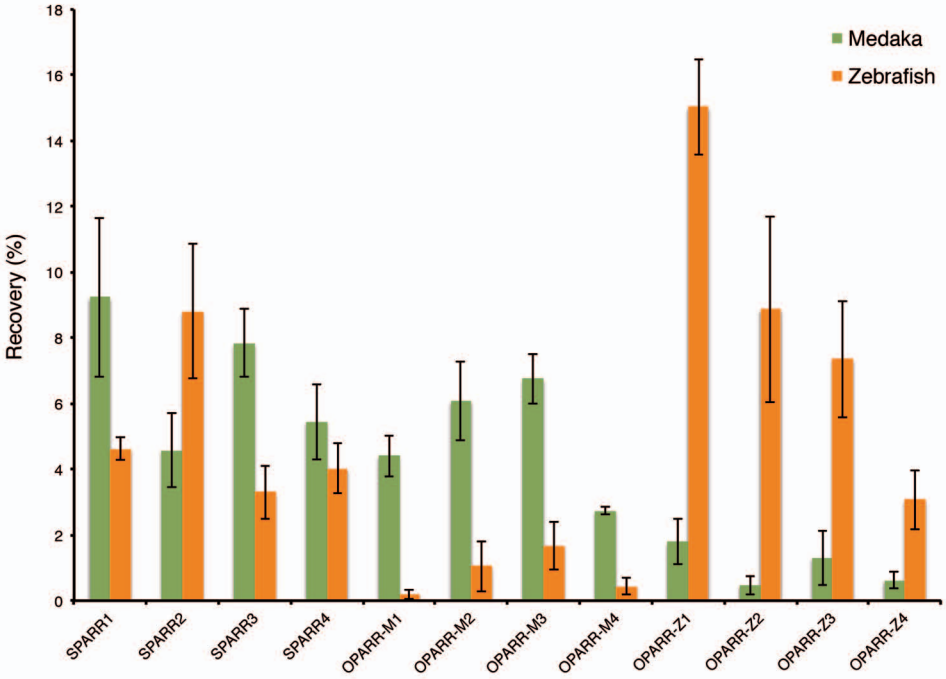
Tena_FigS4


A
hs1315
hs72
hs302
hs625
hs619
hs73
hs895
hs969
hs924
B
xt1327
dr1327
hs1327

Tena_FigS5

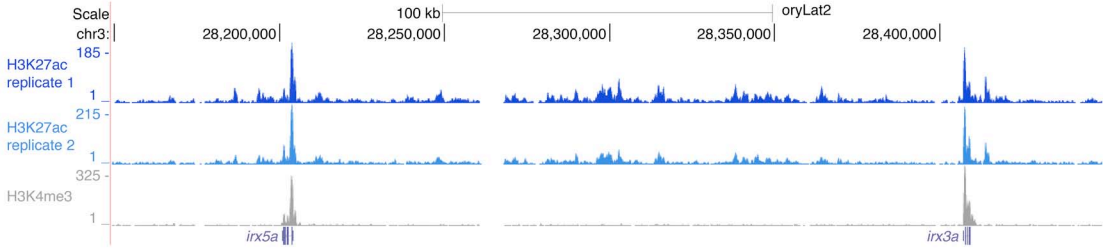

A
hb
fb
hs625
fb
hs619
hb
fb
sc
hs969
hb
fb
Z625
fb
Z619
sc
hb
fb
Z969
B
hb
sc
hs73
sc
hb
Z73
sc
hb
M73
fb
nc
hs1315
nc
fb
Z1315
nc
fb
M1315
mb
hb
hs1327
hb
mb
Z1327
hb
mb
M1327

Tena_FigS6


Medaka
Zebrafish
Recovery (%)
0
2
4
6
8
10
12
14
16
18
SPARR1
SPARR2
SPARR3
SPARR4
OPARR-M1
OPARR-M2
OPARR-M3
OPARR-M4
OPARR-Z1
OPARR-Z2
OPARR-Z3
OPARR-Z4

Tena_FigS7

A



B