Supporting Information

Özbudak et al. 10.1073/pnas.0909375107

SI Materials and Methods

Short Time-Series Expression Miner Analysis. The present probeset list was filtered further to include only the probesets in the top 50% variance. No further filtering was done in short time-series expression miner (STEM) (1) based on expression levels. The left and right side samples were used as duplicates. The average value of each probeset was used to obtain its profile. The $\log_2$ expression values were normalized to the first time point. Fifty profiles were used. Maximum step size was selected to be two (4-fold change in between consecutive time steps). All possible permutations were done. Bonferroni (5%) correction method was selected. The number of genes belonging to each profile is shown within each panel (Fig. 3*B*) and the names of the genes with Ensembl IDs are provided (Dataset S1). Gene Ontology (GO) statistics enrichment was done for the genes belonging to significant profiles (Dataset S1). Clusters were color-coded based on their similarity to each other with respect to enriched GO terms (Fig. 3*B*).

Annotation and Analysis Strategy in Manteia. Probeset descriptions were downloaded from Affymetrix. Corresponding gene models from the National Center for Biotechnology Information (NCBI) and Ensembl were found using the mapping provided by Affymetrix and Ensembl. The probeset mapping to unigene cDNAs and RefSeq sequences provided by Affymetrix was used to compute missing links to gene models. Gene models and their functional annotation including GO were downloaded from Ensembl and NCBI. This annotation is enhanced by inferring to the fish gene models the annotation of their corresponding human and mouse orthologs. Orthology was computed using Inparanoid (2). GO statistics were computed using the hypergeometric distribution similar to Beissbarth et al. (3). *P* values were corrected for multiple testing using the Benjamini-Hochberg false-discovery rate (FDR) method (4). Manteia is accessible at http://research.stowers-institute.org:8000/Manteia.

GO Enrichment Statistics. GO enrichment statistics were done using Manteia. Briefly, the Ensembl IDs of the differentially expressed gene lists were entered into the Manteia database and GO enrichment statistics were obtained for each spatial position separately. For the GO statistics, the reference list was selected as the genes existing on the zebrafish array. Only annotated genes were used for both the reference list and the query list. GO enrichment was done only in the "biological processes" subbranch of the GO tree. The GO terms that were lower than level 4 or had fewer than four genes or were higher than 10% FDR were filtered out. The GO lists are provided in Dataset S1. After using RMA (Robust Multi-Array) normalization (rather than Microarray Suite 5.0), GO statistical analysis resulted in similar GO terms enriched in the anterior and posterior PSM (Dataset S1).

Transcription Factor Data-Mining. The transcription factor list was obtained by selecting genes belonging to the GO term "transcriptional regulatory activity." In total, 447 genes were expressed at different levels in the tail-bud region. The names of the transcription factors are provided in Dataset S1.

Measurements of Total Protein Levels. Bio-Rad Protein Assay (Bradford) was used to measure the total protein levels. Total protein levels were measured for cells located at three different spatial domains: the posterior presomitic mesoderm (PSM), the last five somites formed (including the axial tissues inbetween), and somites 6 to 10 (including the axial tissues inbetween). The numbers of experiments were 5, 4, and 5, respectively. Twelve to 15 somite-stage embryos were dissected. For each experiment, corresponding sections from 15 embryos were pooled into 10 μL of Nonidet P-40 lysis buffer [50 mM Tris-Cl pH 8.0, 150 mM NaCl, 1% Nonidet P-40, 1 mM PMSF, 1× protease inhibitors mixture tablets, Roche (catalog number 11 836 145001)]. Tubes were frozen on dry ice and thawed; 7 μL of the sample was added into 1 mL of Protein Assay Dye Reagent Concentrate (catalog number 500-0006). The absorbance at 595 nm was measured. The absorbance levels were in the linear range as determined by BSA (BSA) protein-absorbance standard curve. The absorbance values were normalized by DNA quantities obtained by qPCR. The ratios of the normalized total protein levels were obtained by dividing anterior segments by the posterior one. Genomic DNA was purified from the 2.5 μL of the remaining lysate. qPCR was performed using intronic primers corresponding to *mespb* and *tbx6* genes. For each sample, 12 qPCR experiments were completed (two different primer sets, three triplicates for each primer set at two different input DNA concentrations). The average cycle threshold (Ct) values were used to calculate the DNA ratio between the anterior and posterior samples. The average normalized ratios of protein levels were plotted with error bars representing SEM.

Measurements of ATP Levels. ATPlite Luminescence ATP Detection Assay System (Perkin-Elmer) was used to measure the ATP levels. ATP levels were measured for cells located in two different spatial domains: the posterior PSM and the last two somites formed (including the axial tissues in between). Twelve to 15 somite-stage embryos were dissected. For each experiment, corresponding sections from a single embryo were placed in a tube. Tubes were frozen on dry ice and thawed. Next, the ATPlite protocol was followed with the following exceptions: the centrifugation steps were omitted and only 35 μL of the 50-μL cell lysate was used in each experiment. The ratios of the ATP levels were obtained by dividing the ATP levels in somites by that of the tail bud. Genomic DNA was purified from the 12.5 μL of the remaining lysate. qPCR was performed by using intronic primers corresponding to *mespb* and *tbx6* genes. For each sample, 12 qPCR experiments were performed (two different primer sets, three triplicates for each primer set at two different input DNA concentrations). The average Ct values were used to calculate the DNA ratio between the anterior and posterior samples. In total, 14 experiments were completed. The average normalized ATP ratio was plotted with error bars representing SEM.

Measurements of Cytochrome-C Oxidase Activity. Cytochrome-C Oxidase Assay Kit (Sigma) was used. Cytochrome-C oxidase activity was measured for cells located in two different spatial domains: the posterior PSM and the last five somites formed (including the axial tissues in between). Twelve to 14 somite-stage embryos were dissected. For each experiment, corresponding sections from seven embryos were placed in a tube. Tubes were frozen in liquid nitrogen and thawed. Samples were processed according to the manufacturer's instructions. The ratios of the cytochrome-C oxidase activities were obtained by dividing the cytochrome-C oxidase activities in somites by that of the tail bud. Genomic DNA was purified from the remaining half of the lysate. qPCR was performed using intronic primers corresponding to *mespb* and *tbx6* genes. For each sample, six qPCR experiments were performed (two different primer sets, three triplicates for each primer set). The average Ct values were used to calculate the DNA ratio between the anterior and posterior samples. In

total, eight experiments were completed. The average normalized ratio of cytochrome-C oxidase activity was plotted with error bars representing SEM.

MitoTracker Staining. Twelve to 15 somite-stage embryos were dissected and incubated with 25 nM MitoTracker Orange CM-H_2TMROS (M7511; Molecular Probes-Invitrogen) in L-15 (21083; Gibco Invitrogen) medium for 2 h. The stained embryos were briefly fixed with 4% paraformaldehyde. Fluorescent pictures were taken using a Leica DM microscope. The profile of the intensity was quantified in ImageJ (http://rsb.info.nih.gov/ij/) program. In total, six embryos were analyzed and the picture of a representative one is shown in Fig. 4 *C* and *D*. The line trajectory is selected in the center of the mesoderm to prevent possible dye-trapping effects on the boundaries.

Measurements of Hydrogen Peroxide Levels. H_2O_2 levels were measured by using the Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Molecular Probes; A22188). H_2O_2 levels were measured for cells located at three different spatial domains: the posterior PSM, the last five somites formed (including the axial tissues in between), and somites 6 to 10 (including the axial tissues in between). Twelve to 14 somite-stage embryos were dissected. For each experiment, corresponding sections from 15 embryos were placed in a tube. Tubes were frozen on dry ice and thawed. One-fifth of the lysate was saved for qPCR DNA measurements. The remainder was used to measure H_2O_2 levels according to the manufacturer's instructions. The ratios of the H_2O_2 levels were obtained by dividing the H_2O_2 levels in somites by that of the tail bud. Genomic DNA was purified from the remainder of the lysate. qPCR was performed by using intronic primers corresponding to *mespb* and *tbx6* genes. For each sample, six qPCR experiments were performed (two different primer sets, three triplicates for each primer set). The average Ct values were used to calculate the DNA ratio between the anterior and posterior samples. In total, five experiments were done. The average normalized H_2O_2 ratios were plotted with error bars representing SEM.

RT-PCR Measurements. The tail-bud region and the left and right last two somites formed (LS12 or RS12) of 12 to 13 somite-stage zebrafish embryos were dissected as for the microarray experiments. The isolated sections were placed in TRIzol solution (Invitrogen) and frozen immediately at −80 °C. Total RNA was isolated from these sections. The RNA was further purified using Zymo Research RNA Cleanup Kit (catalog number R1015). Then, cDNA was generated using Applied Biosystems High-Capacity RNA-to-cDNA Kit. Fast SYBR Green master mix (Applied Biosystems) was used in RT-PCR reactions. Each primer set was tested for efficiency, linearity and reproducibility in RT-PCR. For each primer set, RT-PCR reactions were done in technical triplicates using cDNAs from the tail bud, the left and right last two somites formed (LS and RS) of a single embryo. Average fold-change was calculated as: FC= 2^ (Ct_{TB} - (Ct_{LS}+ Ct_{RS})/2). The FC is the ratio of the expression levels of the corresponding gene in the two last somites formed versus the tail bud. For each primer set, experiments were repeated with cDNAs from three different embryos. The average FC of the three different experiments is shown in Table S3. The sequences of the primer sets are provided in Table S4.

1. Ernst J, Bar-Joseph Z (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191.
2. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314:1041–1052.
3. Beissbarth T, Speed TP (2004) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20:1464–1465.
4. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statistic Soc B* 57:289–300.

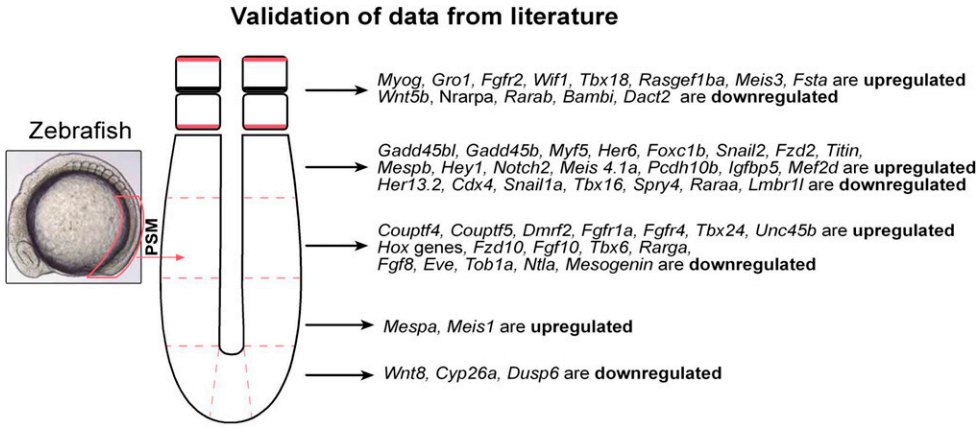

Fig. S1. Validation of data from the literature. Genes are marked, depending after which spatial point their levels are significantly increased or decreased (compared with the initial time point; tail bud) with 10% false-discovery rate and 1.5-fold change. The results can be compared with those from the ZFIN database online (http://zfin.org/cgi-bin/webdriver?MIval=aa-xpatselect.apg).

Table S1. Coefficients of correlation between the 12 samples

	SL5	SR5	L71	R71	L72	R72	L73	R73	L74	R74	L79	L59
SL5	1											
SR5	**0.975**	1										
L71	0.976	0.958	**1**									
R71	0.977	0.96	**0.99**	**1**								
L72	0.963	0.949	0.986	0.984	**1**							
R72	0.957	0.948	0.976	0.976	**0.982**	**1**						
L73	0.947	0.933	0.971	0.968	0.986	0.974	**1**					
R73	0.95	0.935	0.97	0.97	0.982	0.977	**0.988**	**1**				
L74	0.924	0.925	0.944	0.944	0.964	0.954	0.977	0.972	**1**			
R74	0.932	0.927	0.958	0.958	0.976	0.975	0.984	0.979	**0.975**	**1**		
L79	0.916	0.891	0.934	0.934	0.951	0.939	0.966	0.966	0.96	0.959	**1**	
L59	0.927	0.915	0.939	0.939	0.948	0.949	0.964	0.968	0.953	0.953	**0.97**	**1**

Pearson correlation coefficients are calculated between the 12 samples based on signal intensities from microarray data. The coefficients between the duplicate samples (in bold) are above 0.97.

Table S2. Genes associated with "Oxidative Metabolism" GO term in the up-regulated group of genes can be divided into different functional groups

Functional groups	Ensembl gene ID	Gene names
ATP synthase(Complex 5) subunits	ENSDARG00000068940	*atp5i*
	ENSDARG00000011553	*atp5f1*
	ENSDARG00000011841	*atp5l*
	ENSDARG00000019404	*atp5d*
	ENSDARG00000070819	*ATP5B*
	ENSDARG00000017775	*zgc:55970 / ATP5G3*
Complex 1 subunits	ENSDARG00000006290	*ndufs5*
	ENSDARG00000013333	*ndufa10*
	ENSDARG00000038028	*ndufa6*
	ENSDARG00000041400	*zgc:100908 / NDUFA3*
	ENSDARG00000051986	*ndufs8*
Complex 3 subunits	ENSDARG00000011146	*Uqcrb*
	ENSDARG00000059128	*Uqcrh*
Cytochrome *c*	ENSDARG00000044562	*zgc:86706 / cyc*
	ENSDARG00000038075	*cyc1*
Mitochondria membrane transporter complex subunits	ENSDARG00000029510	*timm17a*
	ENSDARG00000045146	*tomm22*
	ENSDARG00000071396	*timm23*
	ENSDARG00000025132	*tomm20*
	ENSDARG00000036721	*tomm40l*

Table S3. Fold-changes in the qPCR and microarray data

Avg qPCR FC	Microarray FC	Gene symbol
11.51	3.68	*mycA*
21.98	7.89	*mycB*
2.46	1.09	*rbm22*
3.13	1.36	*rpl13a*
1.05	1.50	*atpB*
2.08	1.54	*eIF2s1l*
1.75	1.56	*rps14*
1.38	1.56	*rplp0*
0.96	1.67	*cyc1*
1.05	1.69	*sfrs11*
1.13	1.70	*taz*
1.39	1.73	*eif5*
0.82	1.74	*atp5fl*
1.25	1.75	*eif4B*
1.11	1.75	*uqcrh*
0.82	1.86	*atp5d*
1.44	1.90	*ddx51*
1.88	1.91	*S15mt*
1.49	1.97	*timm23*
1.65	1.99	*timm17a*
1.54	1.20	*mrpl19*
2.50	2.01	*rpp14*
1.12	2.02	*atp5l*
1.40	2.03	*naca*
1.17	2.10	*tomm22*
1.54	2.13	*ddx49*
1.14	2.14	*EF1*
1.57	2.39	*rpl13*
1.27	2.50	*tomm20*
1.69	2.56	*nol5*
0.83	2.64	*mttfa*
1.58	2.80	*atp synthase lipid binding protein mitochondrial precursor*
2.09	2.80	*mrrf*
1.89	2.86	*4EBP*
5.84	2.88	*pabpc4*
2.32	2.89	*EIF3*
2.34	2.90	*eIF4A-2*
2.49	3.09	*Cell growth regulating nucleolar, lyar*
6.02	4.92	*pex5*
1.77	6.10	*eIF5A2*
2.79	6.80	*tsc1a*

qPCR results for the 41 selected genes belonging to GO terms identified in the zebrafish data set. Fold-change (FC) of gene expression in the somites versus tail bud are provided in the left column and middle column from qPCR and microarray data, respectively.

Table S4. Primer sequences that are used during qPCR

Forward (F) and reverse (R) primers	Sequence
sfrs11F	CCCCGTCCCATCTATTGGT
sfrs11R	TCCAAGTTTGGTCCTCCAAGA
ddx51F	GTCCAGGCGAACGTCAGAA
ddx51R	TGTGCTTATCAACAGTGGGATCTT
rpp14F	TCTGATGCACAGTTCAAGCAATT
rpp14R	CGCCAATCTCGCCATACAA
ddx49F	TGTTTTGAAGCCATACGAGTTCA
ddx49R	TCATTCCATCCCGACATCTGT
pabpc4F	GCTAACAAGGCGGTGGAAGA
pabpc4R	TGTCGCTCCATTTTTTTTTGTG
eIF4A-2F	TGTAGAACGAGAGGAGTGGAAGTTG
eIF4A-2R	GCGTGATTGTCAGGGTCTCA
eif2s1lF	GCATGCCCATCAAGATCAACT
eif2s1lR	TCCAGTGTGGTCGTGGTCAT
rps14F	GCATCACTGCTCTGCACATCA
rps14R	TCCAGGAGTCTTGGTTCTGTTTC
rplp0F	CCTGGCTGTCACTGTCGAAA
rplp0R	CAGGTAGGCCTTCACCTTCTCA
eif5F	GCAGCGTGTCAGACCAGTTCTA
eif5R	CCTTCCACCTTAGCGATGAGA
EIF4BF	CGGCCAAGCTCAGACCAA
EIF4BF	ACGCGACTTGTCCCCATATC
nol5F	AGCAATCTCTCTGCTGGATGATC
nol5R	TCCCTGCAGCGCATGAT
eIF4EBPF	CTGGCGGAACCAGGATCAT
eIF4EBPR	CGGGCGATCGGAGAGTT
tazF	TCAGGCATCTGGTCAATGCA
tazR	CGCATTTCCGTCGGATTC
timm23F	GAGGACTCGGGAGCATTTTTG
timm23R	CAGTCAACGGTACTCCTGAAAGC
timm17aF	TTAATAGAAGGTGCTGGAATTTTGC
timm17aR	GTGGGCCAGTTGGGAATTG
tomm22F	CGAGGAGCTGGATGAAACTATGTT
tomm22R	GCTGCCGATCGCACTGA
tomm20F	GCTGCTGGCTCAAGGAGATT
tomm20R	GCGATGGCATTGGTCAGAT
ATPBF	CCATGCAGGAAAGAATCACAAC
ATPBR	GCACATAGATAGCCTGCACAGATG
cyc1F	GCAGCCGAGCCAGAACA
cyc1R	CCACCCAACAACACCTTCAGA
atp5f1F	CAAGAAGTTTGGGCCCAGTGT
atp5f1R	GCCTTTGCTATTTTATCCGCATT
uqcrhF	ATGGTGGATCCTCTAGAGACAGTGA
uqcrhR	CGCGAGCATGAGCACAGT
atp5dF	CGAACGAAGCGATTGTCAAA
atp5dR	ACCGACAGAATTCTCAACAAACAG
atp5lF	GCACAGGCGGTGCAGAA
atp5lR	GACTGCGGCTCCGACAAGT
S15mtF	AAGGTGCAGAGGGATGAGAATG
S15mtR	TGTCAAAATCGCCACCTTGA
mrpl19F	CATCCCAGTCAATCCGTTAAAAGT
mrpl19R	GGCGCTCCCAACGTTTAGA
nacaF	AAACACACAGACGCCCACAGT
nacaR	CTCCGGTCTCATCAACCTCTTC
EF1F	GAAGCGGAACGACGGTTTT
EF1R	CAGTGTCCTGTTTGGGCTTTG
rpl13F	CGCTAAGGACGGAGTGAACAA
rpl13R	GAGGGTTCAAAATCATGCCATT
mrrfF	CAATCTCCAAAGTCTAGTTACCACAGA
mrrfR	ATACTGGCCGCAGGAGGTT
EIF3F	GAGCGGAGGACGCCTACA
EIF3R	CAGTCTCTAGTTTGGCCAGGTATG
CELL GROWTH REGULATING NUCLEOLAR / LYAR F	TGCGCCGCCTCAGTGT

Table S4. Cont.

Forward (F) and reverse (R) primers	Sequence
CELL GROWTH REGULATING NUCLEOLAR / LYAR R	TCCAGTTCTCAAACTTGGCTTTC
eIF5A2F	TGATCAGAGCGTTTGGAAACTTC
eIF5A2R	CCACTGGTGAAATCAGTATCCAGAT
pex5F	GCTCTTACGACAAGGGATATCAGTT
pex5R	TTCACCCCTTCTGCAAACG
ATP SYNTHASE LIPID BINDING PROTEIN MITOCHONDRIAL PRECURSOR F	GATCCGGTGCTGGAATCG
ATP SYNTHASE LIPID BINDING PROTEIN MITOCHONDRIAL PRECURSOR R	GACGGGTTCCTGGCGTATC
tsc1aF	TGCCTCAAGACTGACTCAGATGTAG
tsc1aR	GAATCATGGGCATTAGAGTGATCA
mttfaF	GCAGAGAGTGCGGCTTTTG
mttfaR	CAATTCCTTTTTTTTCCGAATGG
mycAF	CATTGCATTGCGTCTCGAAA
mycAR	CAGCATTTTGACACTTGTATTTAACG
mycBF	TCACGCTCACGCTGACATC
mycBR	GAGTGCCGTAGCCGTGGTAA
rbm22F	CACATCTGTTCCTTCTGGGTTAAA
rbm22R	TTTTCATGCCTGTACGGACACT
rpl13aF	CCCGTGGACCATATCACTTCA
rpl13aR	GCATACCTCTTACAGTCCTCCAGAA
beta actin1F	TGCCCCTCGTGCTGTTTT
beta actin1R	TCTGTCCCATGCCAACCA
beta actin2F	CGAGCTGTCTTCCCATCCA
beta actin2R	TCACCAACGTAGCTGTCTTTCTG
tbx6 intron	TTTCTGCTTGCTCAAACTGC
tbx6 intron	AAACTCAAGAGTTGTGTTGTTTCG
mespa intron	TTTTCCCCATCTCAGTCCAC
mespa intron	GGCCTTCGGTTACAGGGTAT

Other Supporting Information Files

Dataset S1(XLS)