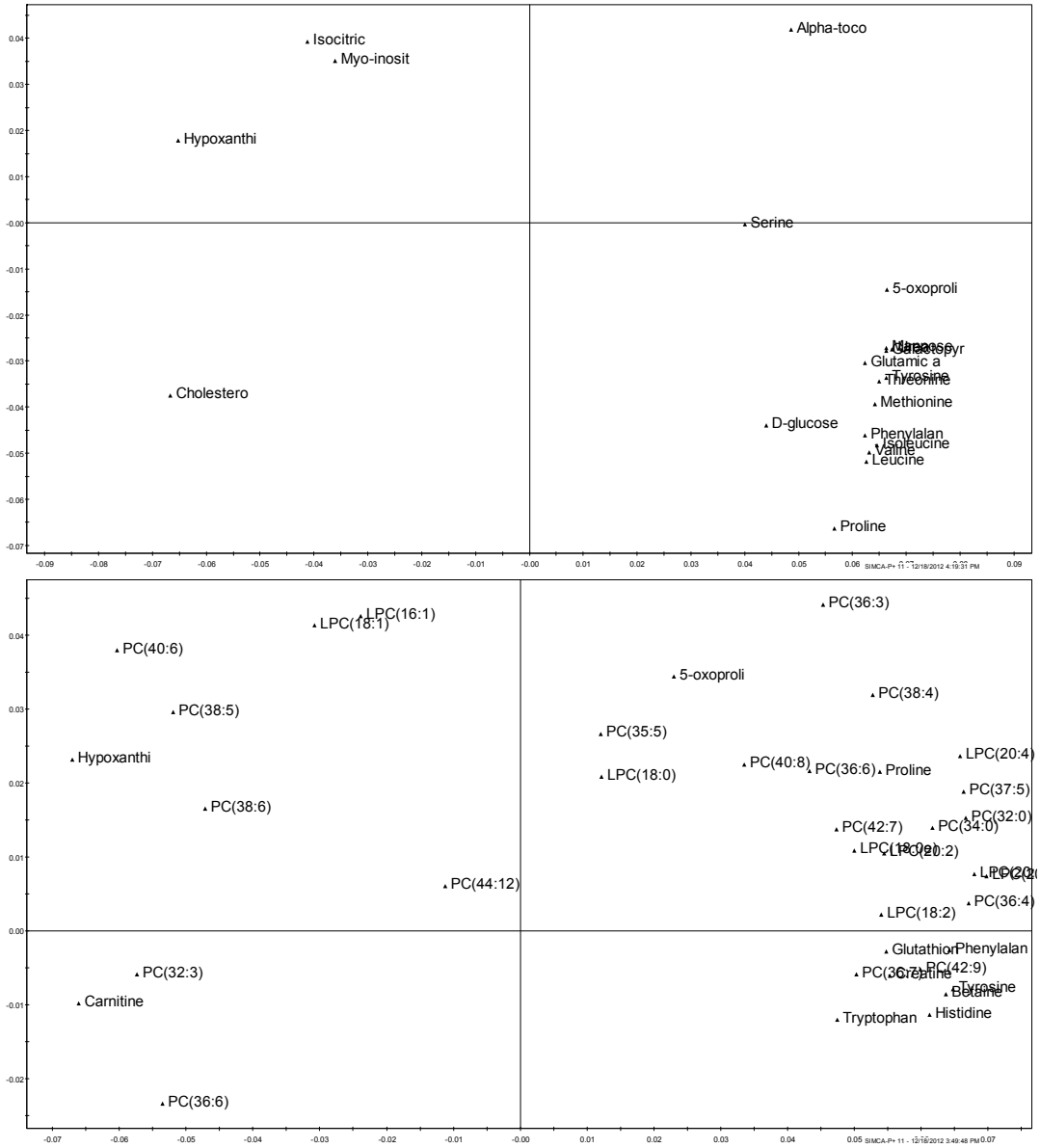
Supplementary Information



Supplementary Fig. 1 PCA loadings plot of GC-MS (top) and LC-MS (bottom) data derived from zebrafish embryogenesis samples (see main Fig 1). The horizontal axis refers to the first component (p[1]), while the vertical axis refers to second component (p[2]).

Supplementary Table 1. The variable importance in the projection (VIP) values of identified metabolites. Higher values indicate a stronger influence of the metabolite in distinguishing different time points.

GC-MS		LC-MS			
Metabolite	VIP value	Metabolite	VIP value	Metabolite	VIP value
Tyrosine	1.75	Hypoxanthine	2.40	PC(38:6)	1.55
L-methionine	1.71	Phenylalanine	2.32	PC(42:9)	1.51
Hypoxanthine	1.69	Histidine	2.22	PC(32:3)	1.41
L-valine	1.68	Tyrosine	2.19	LPC(18:1)	1.37
Iso-leucine	1.66	Creatine	2.06	LPC(16:1)	1.36
Leucine	1.59	Betaine	1.99	LPC(20:5)	1.35
Urea	1.59	LPC(20:1)	1.93	PC(40:8)	1.33
L-threonine	1.58	PC(36:4)	1.89	LPC(18:0)	1.32
L-phenylalanine	1.57	PC(32:0)	1.88	Proline	1.31
Mannose	1.53	Carnitine	1.86	PC(38:4)	1.24
Galactopyranose	1.52	LPC(18:0e)	1.83	5-oxoproline	1.23
Pyroglutamic acid	1.52	PC(40:6)	1.81	LPC(18:2)	1.16
L-proline	1.42	Tryptophan	1.79	PC(38:5)	1.11
Isocitric acid	1.41	LPC(20:4)	1.76	PC(36:6)	1.04
D-glucose	1.25	PC(42:7)	1.75	Glutathione	1.02
Glutamic acid	1.23	PC(36:6)	1.72	PC(44:12)	1.01
Alpha-tocopherol	1.20	LPC(20:2)	1.64	PC(36:3)	1.01
Serine	1.15	PC(37:5)	1.63	PC(35:5)	1.01
Cholesterol	1.12	PC(36:7)	1.62		
Myo-inositol	1.09	PC(34:0)	1.59		

Supplementary table 2. Detectable metabolite-transcript associations for embryogenesis

Metabolite	Transcript match	Reaction
Valine	• Branched chain aminotransferase 2, mitochondrial	Branched chain amino acid + 2-oxoglutarate → corresponding keto acid + L-glutamate
Leucine	• Branched chain aminotransferase 2, mitochondrial	Branched chain amino acid + 2-oxoglutarate → corresponding keto acid + L-glutamate
Isoleucine	• Branched chain aminotransferase 2, mitochondrial	Branched chain amino acid + 2-oxoglutarate → corresponding keto acid + L-glutamate
Serine	• Zgc:66171 (serine hydroxymethyltransferase 1) • Cystathionine-beta-synthase • Seryl-tRNA synthetase	• Transferase activity • L-serine + L-homocysteine ←→ L-cystathionine + H2O • Coupling of serine to tRNA
Methionine	• Methionine adenosyltransferase II, beta	• Synthesizes S-adenosylmethionine (SAM) from methionine and ATP
Glutamic acid	• Glutamate-cysteine ligase, modifier subunit • Glutamate dehydrogenase 1a • Glutamate-ammonia ligase (glutamine synthase) a	• Rate limiting step in glutathione synthesis • Oxidative deamination of glutamate to 2-oxoglutarate and free $NH4^+$ using either NAD+ or NADP+ as a co-factor • Glutamate + ATP + NH_3 → Glutamine + ADP + phosphate
Phenylalanine	• Wu:fe50f06	• Phenylalanine-tRNA ligase activity
Hypoxanthine	• Hypoxanthine guanine phosphoribosyltransferase-like	• conversion of hypoxanthine to inosine monophosphate and guanine to guanosine monophosphate.
Tyrosine	• Tyrosine hydroxylase	• L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA).
Vitellogenin	• Cathepsin F • Cathepsin L-a • Cathepsin L-b • Cathepsin D • Zgc:85774	• Proteolytic activity, including that of vitellogenin (for all cathepsins, including Zgc:85774)

Supplementary Information 1. Data pre-processing and mining

Each chromatogram obtained from GC-MS and LC-MS analysis was converted to .AIA and .xml respectively using MassHunter (Agilent). The data were further pre-processed using MZmine 2.0[14] to generate discrete peak lists containing m/z, retention time (RT) and ion abundance (peak intensity). Within MZmine 2.0, the data from different samples were combined into a single matrix by aligning peaks with the same *m/z* and RT for GC-MS and LC-MS data, respectively. The cut-offs for minimum time span are 0.01 min (GC-MS with min. height 5000.0 and tolerance of *m/z* 0.500) and 0.1 min (LC-MS with min. height 10000.0 and tolerance of *m/z* 0.500). The area of each peak was normalized to that of internal standard (FMOC-glycine) in each data column of the matrix (variables in rows; observations in columns). The data was subsequently filtered by the presence of peaks in at least 80% of each sample category[15] using Microsoft Excel.

The pre-processed GC-MS and LC-MS data were exported to SIMCA-P 11.0 (Umetrics AB, Umeå, Sweden) for analysis and visualization using multivariate statistical methods. Two multivariate analysis methods were used to find the optimal separation of clusters, namely the principal component analysis (PCA) and orthogonal projection to latent structure-discriminant analysis (OPLS-DA). For PCA & OPLS-DA, all data were mean-centered and unit variance scaled. A 7-fold cross-validation was applied to both PCA & OPLS-DA and the reliabilities of models were further rigorously validated by permutation tests.

The variable importance in the projection (VIP) values reflect the importance of terms in the OPLS-DA model both with respect to Y, i.e. its correlation to all the responses, and with respect to X (the projection). Thus, metabolites with VIP value greater than one (>1) were selected for further testing and validation using Kruskal-Wallis Test ($p < 0.05$). The selected compounds from GC-MS analysis were identified by comparison of mass spectra and retention time with those of reference standards, and those available in libraries (NIST 2005), While those from LC-MS detection were searched in the Human Metabolome Database (HMDB, www.hmdb.ca) first using ion mass and then were further identified by either MS/MS fragmentation pattern or reference standards[16].

Heatmaps of identified metabolites were generated using MultiExperiment View Version 4.8[17] and further analyzed using agglomerative hierarchical cluster analysis based on Pearson correlation distances with gene leaf optimization. The Kruskal-Wallis Test was also performed within this software package.