



Universiteit
Leiden
The Netherlands

High-throughput profiling of small molecules using mass spectrometry

Raterink, R.J.

Citation

Raterink, R. J. (2014, October 1). *High-throughput profiling of small molecules using mass spectrometry*. Retrieved from <https://hdl.handle.net/1887/28970>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/28970>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/28970> holds various files of this Leiden University dissertation.

Author: Raterink, Robert-Jan

Title: High-throughput profiling of small molecules using mass spectrometry

Issue Date: 2014-10-01

Chapter 3

Rapid metabolic screening of early zebrafish embryogenesis based on direct-infusion-nanoESI-FTMS

ABSTRACT

Single zebrafish eggs were rapidly profiled using high resolution nanoelectrospray direct infusion mass spectrometry with limited sample preparation and without separation. The analysis time per sample is around 1 minute. Using this approach the different developmental stages of zebrafish eggs can be characterized by their active metabolites. Five different development stages with distinct metabolic fingerprints could clearly be observed when untargeted analysis is performed and the data are plotted using principal component analysis (PCA). Using this approach early embryogenesis is followed with a time resolution of 1 hour and 102 features proved relevant. Of these, significant number of putatively identified compounds has not been reported earlier to have any association with early zebrafish embryogenesis yet. The onset of gene expression and the increase in energy requirement is reflected by the measured metabolome complementing earlier reported transcriptomics studies from a systems biology point of view. By deholking and dechorionation eggs at two early developmental stages, we were able to observe distinct changes in localized metabolism.

Based on

R.J. Raterink*, F. M. Kloet*, J. Li, N. A. Wattel, M. J. M. Schaaf, H. P. Spaink, R. Berger, R. J. Vreeken, and T. Hankemeier, "Rapid metabolic screening of early zebrafish embryogenesis based on direct infusion-nanoESI-FTMS", *Metabolomics*, vol. 9, no. 4, pp. 864–873, 2013.

*shared first authorship

INTRODUCTION

In the past decade the Zebrafish (*Danio Rerio*) has become a popular vertebrate model system for studying human development and disease[1]–[3]. This is because the development of the zebrafish is very similar to the embryogenesis in higher vertebrates, including humans. But unlike mammals, zebrafish develop from a fertilized egg to an adult outside the female in a transparent egg. This makes it possible to observe developing embryos from the single cell to the entire organism level[4]. The development time of the embryo is fast (after about 2 days most common vertebrate specific body features can be seen including brain, eyes, ears and all internal organs) and the number of offspring is large (100-200 eggs per mating). Moreover, because they are small, available in large numbers and maintained at low cost, zebrafish embryos are ideal as model systems in high-throughput whole organism preclinical drug screening and toxicology studies as less drug is required, a larger number of animals can be used and less ethical issues are associated[5], [6]. Therefore, the zebrafish embryo can bridge the gap between cell assays and rodent assays.

The external and rapid development as well as transparency of zebrafish embryos is ideal for observable phenotype-based screening at cellular and organ-tissue level using live imaging, for example using a complex object parametric analyzer (COPAS)[7]. Although these phenotypic changes are adequate for specific drug-induced biological responses, they don't reveal system-wide responses and are not sufficient in elucidating the mode of action or potential toxicity of drugs[6]. With the ever increasing interest in studying biological systems in a holistic manner (systems biology) the necessity for delivering qualitative and quantitative data of complete biological systems for which zebrafish offer many advantages becomes clear[8]. Metabolomics is a powerful tool within systems biology and investigates the complex interactions of the metabolism and metabolic networks[9]–[11]. One of the greatest strengths of metabolomics is the ability to capture a molecular snapshot of metabolites as reactants, intermediates or products of (enzyme-mediated) biochemical reactions. Metabolomics complements genomics, transcriptomics and proteomics for metabolites are in a unique position as they are building blocks for all other biochemical structures including proteins (amino acids), genes and transcripts (nucleotides) and cell walls (lipids) [12], [13].

So far, the number of metabolomics studies on zebrafish embryos is limited[14]–[18]. Most of these studies included chromatography which is time consuming and therefore less suitable for high-throughput metabolic screening. Recent studies outlined the power of high-resolution direct infusion mass spectrometry (HR-DI-MS) metabolomics on other complex samples[19]–[22]. Most of these studies were performed using flow injection electrospray ionization (ESI)-MS and therefore did not explore the advantages of nanoESI over normal ESI with respect to ionization efficiency and ion suppression effects[23].

In this paper we describe a rapid metabolic profiling method of zebrafish eggs, based on lysis of single zebrafish eggs and subsequent HR-nanoESI-DI-MS analysis. A clear metabolic distinction between developmental stages in early embryogenesis is described and up-and down regulation of some important primary metabolites is shown. By de yolking and dechorionating the embryos, we were able to highlight localized and time resolved metabolism.

MATERIALS AND METHODS

Chemicals and materials

Methanol was from Biosolve (Valkenswaard, The Netherlands). Water was obtained from a Millipore high purity water dispenser (Billerica, MA, USA). All solvents were HPLC grade. The labeled amino acids ("Cell Free" amino acid mix (20 AA) (U-13C, 98%+; U-15N, 98%))

were bought from Cambridge Isotope Laboratories (Andover, MA, USA) and added to the samples at a concentration of 1 $\mu\text{g}/\text{mL}$. Reserpine (also used as internal standard) was supplied by Fluka (Buchs, Switzerland) and added to the samples at a concentration of 500 ng/mL .

Whole embryo experiments

Two wild type (strain A/B) parent groups were maintained under standard zebrafish aquarium conditions. Since many cell differentiation and phenotypic processes occur within 48 hours after fertilization, the following 5 early developmental stages were chosen: the 4-cell (1 hpf), 64-cell (2 hpf), 1k (3 hpf), 50% Epiboly (5 hpf) and the 18-somites stage (18 hpf) For each of the five developmental stages 8 eggs were analyzed separately in triplicate.

Lysation protocol for the whole embryo

- 1) The egg was optically selected under a microscope in the relevant developmental stage and pipetted to a 1.5 mL eppendorf tube. As much as possible of the egg water was removed.
- 2) The egg was washed 3 times with 1 mL demineralized H_2O .
- 3) 100 μL (9:1, v/v methanol:water) including the internal standards (1 $\mu\text{g}/\text{mL}$ labeled amino acids+ 500 ng/mL reserpine) was added. Immediately the sample was snap-frozen in liquid nitrogen for 2 minutes. This step should quench metabolism and precipitate the proteins.
- 4) After snap-freezing the sample was sonicated for 2 minutes to lyse and to homogenize the sample and visually inspected to confirm homogeneity. In case of non-homogeneity the sample was snap-frozen and sonicated a second time.
- 5.) Precipitated proteins were spun down by centrifuging the lysate at 16.1 krcf at 0 °C. The supernatant (80 μL) was used for DI-MS analysis.

Deyolking and dechorionating experiments

Two developmental stages 1kcell (3 hpf) and somite (18 hpf) of the same wild type were measured (pool of $n=10$) in triplicate.

Dechorionating

- 1) The chorion of the egg was removed mechanically by a tweezer and was put in a 1.5 mL eppendorf tube. Using this methodology only the membrane of the chorion could be analysed. The same lysis protocol as described above was used however, instead of 100 μL , now 1 mL of solvent + IS (since it was a pool of 10 egg instead of a single egg) was added and no washing was done (since it would remove parts of the sample).

Deyolking

- 1) After dechorionating the egg only consists of the yolk with its cells (see Figure 1). The yolk was removed by adding a Ringer solution to the egg and pipetting it up-and down a few times.
- 2) Subsequently, the sample was centrifuged for 5 minutes at 0.8 krcf (to segregate the cells) and the supernatant (including the yolk) was pipetted away. This way only the cells of the embryo could be analyzed[4].
- 3) 1 mL of IS was added to the pallet and the same lysis protocol was used as described above.

Non-fertilized egg (pool of $n=10$, in triplicate)

- 1) First the non-fertilized egg (only 3 hpf stage) was dechorionated resulting in only the yolk with its fluid. This way also the fluid in the yolk could be analyzed.
 - 2) 1 mL of IS was added and the same lysis protocol was used as described above.
- In the deyolking experiments again also whole-eggs were included as a reference.

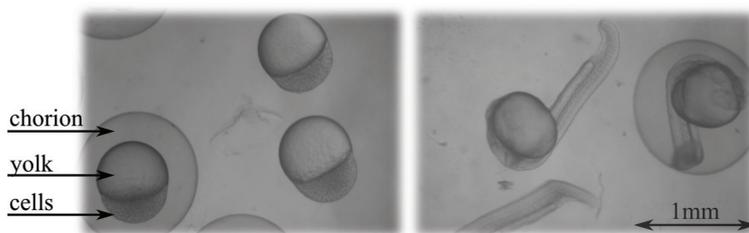


Figure 1: Photomicrograph of zebrafish embryos: (left) 1k-cell stage and (right) somite stage. In both stages some of the embryos were dechorionated in order to perform the de yolking.

MS Analysis

The analyses were performed by DI-nanoESI-MS in the positive ion mode using the automated Advion NanoMate Triversa system (type 'A' chip) coupled to a LTQ-FT Ultra (Thermo Fisher Scientific). Eppendorf 96 well plates were used on which all the samples were randomly distributed. Of each sample 5 μ L was infused using a pressure of 0.2 psi and an electrospray voltage of 1.48 kV in the positive ion mode. Mass spectra were recorded using three scan ranges containing 20 scans: 50-250; 250-500; 500-1000 m/z (in this order) at a resolution of 100.000. Separate scan ranges instead of one full scan range was chosen in order to enhance sensitivity. The MS was tuned with inlet capillary temperature of 120 $^{\circ}$ C, capillary voltage of 35V and the tube lens voltage of 50 V. For tandem MS as well as for the de yolking experiments an LTQ-Orbitrap XL (Thermo Fisher Scientific) was used with inlet capillary temperature of 120 $^{\circ}$ C, capillary voltage of 25 V and the tube lens voltage of 80 V.

Data processing

The first 10 scans (approximately 10 seconds) of every sample were averaged using XCalibur software (version 2.0.7; Thermo Fischer). These average scans were stored in separate files. In some samples no spray or no stable spray was obtained, i.e. no average scan could be made. These samples were discarded from further processing. Using a resolution of 100.000 (at 400 m/z) each sample typically resulted in 5 to 10 thousand unique features in the average spectrum. In contrast to hyphenated MS data, the drawback of DI-MS data is that all masses co-elute. Aligning features over samples now is solely based on accurate mass. To align the masses across the different samples we created mass-bins of very small sizes (0.0003 Da) and assigned the different masses to the nearest bin. For a mass range of 50 to 250 m/z this generated approximately 650.000 mass bins. All data analysis was done using Matlab (version 2011a 64bit; MathWorks). In order for the data to be analyzed in Matlab the Xcalibur averaged spectra files were converted to mzXML format (ReAdW version 4.3.1).

RESULTS AND DISCUSSION

Our primary goal was to explore the possibility of discerning developmental stages in zebrafish embryogenesis by applying HR-nanoESI-DI-MS analysis on lysed eggs. Therefore, reproducible and high-quality MS data should be generated and preferably an automated data processing tool is required. The implementation of these (automated) tools, viz., acquisition and processing strategy, forms an integral part of the proposed method. After binning, empty and almost empty features were removed. This reduced the number of features from \sim 650.000 to \sim 30.000. To reduce analytical variation as much as possible, the data was normalized by selecting the optimal internal standard for each compound from the mix of internal standards. This selection was based on the minimization of the RSD of the response of replicate measure-

ments[24]. Beckmann *et al.* [21] normalized the data by using the total ion current, however with our data this would reduce group differences which is not desirable. To enable statistical interpretation, features that showed consistency per group were selected [25]. Features were considered consistent if they were either present or absent in all samples for a group. However, as we were interested in changes between developmental stages, we allowed for, at most, one missing feature in the replicates of a particular developmental stage. This step further reduced the number of possible features to ~5.000. As we anticipate differences between the samples to be apparent in the most abundant metabolites for that sample, we focused on the top 100 features per sample reducing the number of features of interest to 200. Out of these 200 features, 102 showed significant differences in concentration between the different developmental stages. As observed in other complex samples in other research[20], the lowest mass range (50-250 m/z) showed a clear clustering of the different developmental stages (see Figure 3). This would indicate that at least small molecules relate to metabolic variations between the ages of the embryo.

Robustness/validation of the analytical method

Metabolomics often deals with differential analysis of fingerprints[20] to highlight biomarkers. Especially, when a large number of analytes is taken into consideration absolute quantitation is only reported in some cases. In order to be able to give some indication of the analytical rigor of the method, the RSD of repeated measurements, for example using QC samples is often reported. In the absence of QC samples, the RSD was calculated for each feature using a RSD approach based on multiple sets of replicated measurements[26] and classified in one of the six possible classes (0-0.1,0.1-0.25,0.25-0.35,0.35-0.5,0.5-1 and >1). The result of this inter-assay reproducibility for the 102 features is displayed in Figure 2. As can be seen almost half of these features showed an RSD lower than 25%.

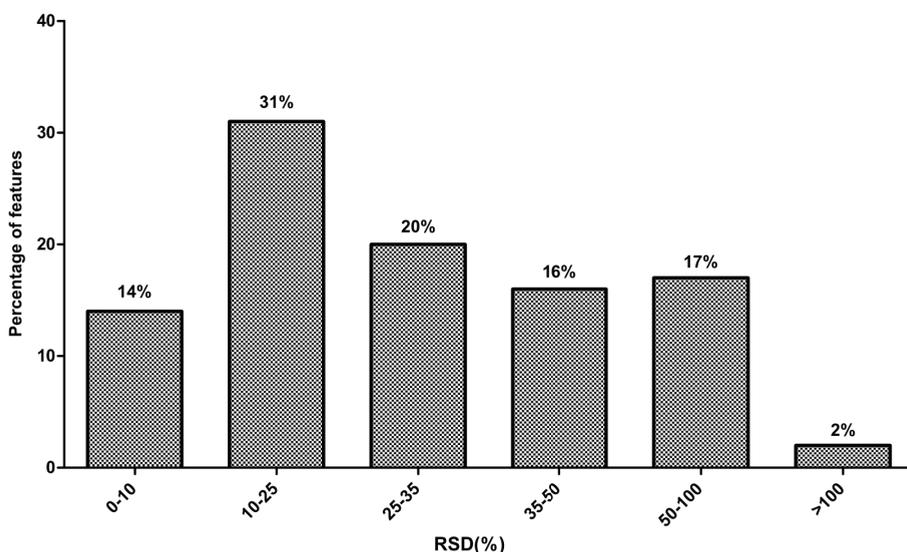


Figure 2: Histograms showing the technical reproducibility of the measurements. 45% of the 102 features has an RSD <25%.

One could argue that by focusing on the top N features too much information is being discarded. For example, tryptophan, hypoxanthine, carnosine, methionine, aspartic acid, propionylcarnitine, dimethyllysine, methyllysine, acetyllysine and serine were discarded (see Supporting Figure 1). However, including lesser abundant features invariably led to higher RSD values for these new features. This could be explained by ion suppression effects: as we measured crude, complex samples containing various compound classes and compound sizes as a result of lysis of the whole embryo. As a comparison: yeast has an estimated 1100 metabolites which is expected to be significantly less than the metabolome of a whole zebrafish embryo[13]. This together with using nanoESI lead to quite some sample-loss due to, e.g., clogging of the nanoESI emitter, which resulted in a loss of replicates. Nevertheless, by discarding the missing variables from the dataset, still 102 features were extracted which revealed the developmental stage differences (Figure 3). (FDA suggests that for LC-MS profiling an RSD of 20% is acceptable[13]).

Identification of metabolites

Out of the significant 102 features 27 separate metabolites (36 features; including all adducts) were putatively identified (see Table 1). Identification was based on (1) search of HR-MS data against HMDB (Human Metabolome project Data Base) and (2) searching MS/MS data against HMDB comparison to standards. After the first HMDB search 171 possible structures were found of which 21 were unique and 150 isomers. From those 150 remaining isomers, 6 were identified using tandem MS and subsequent database search or comparison to standards.

Table 1: RSD values of the 27 (putatively) identified features. The features that were not found in the cited references were ticked as 'new' in the last column.

Feature	M+H	M+Na	M+K	M+NH ₄	New
Acetylcarnitine	0.33	0.29	n.d.	n.d.	
Alanine	n.d.	n.d.	0.31	n.d.	
Arginine	0.11	n.d.	n.d.	n.d.	✓
Asparagine	n.d.	n.d.	0.07	n.d.	
Aspartylphosphate/acetylaspartic acid	n.d.	n.d.	0.16	n.d.	✓
Carnitine	0.28	0.27	n.d.	n.d.	✓
Creatine	0.26	0.19	0.09	n.d.	
Dimetylgarginine	0.11	n.d.	n.d.	n.d.	✓
Dopamine/vanillylamine	n.d.	n.d.	0.16	n.d.	✓
FAPy-adenine	n.d.	n.d.	0.2	n.d.	✓
Glutamine	n.d.	n.d.	0.06	n.d.	
Histidine	0.19	n.d.	0.22	n.d.	
Homoserine/threonine	n.d.	n.d.	0.34	n.d.	
Indoleacetic acid	0.05	n.d.	n.d.	n.d.	✓
Inositol cyclic phosphate	0.16	n.d.	n.d.	n.d.	
Iso/nor leucine	0.14	0.12	0.09	n.d.	
Iso valeraldehyde	n.d.	n.d.	n.d.	0.2	✓
Lysine	0.12	n.d.	n.d.	n.d.	
Phenylalanine	n.d.	0.13	0.09	n.d.	
Phosphoethanolamine	n.d.	n.d.	0.2	n.d.	✓
Proline	n.d.	n.d.	0.12	n.d.	
Quinone	n.d.	0.37	n.d.	n.d.	✓
Safrole	n.d.	0.33	n.d.	n.d.	✓
Spermine dialdehyde	0.25	n.d.	n.d.	n.d.	✓
Tyrosine	n.d.	n.d.	0.07	n.d.	
Valine	n.d.	0.2	0.13	n.d.	
Vinylacetylglycine	0.08	n.d.	n.d.	n.d.	✓

The features that were not found in the cited references were ticked as 'new' in the last column

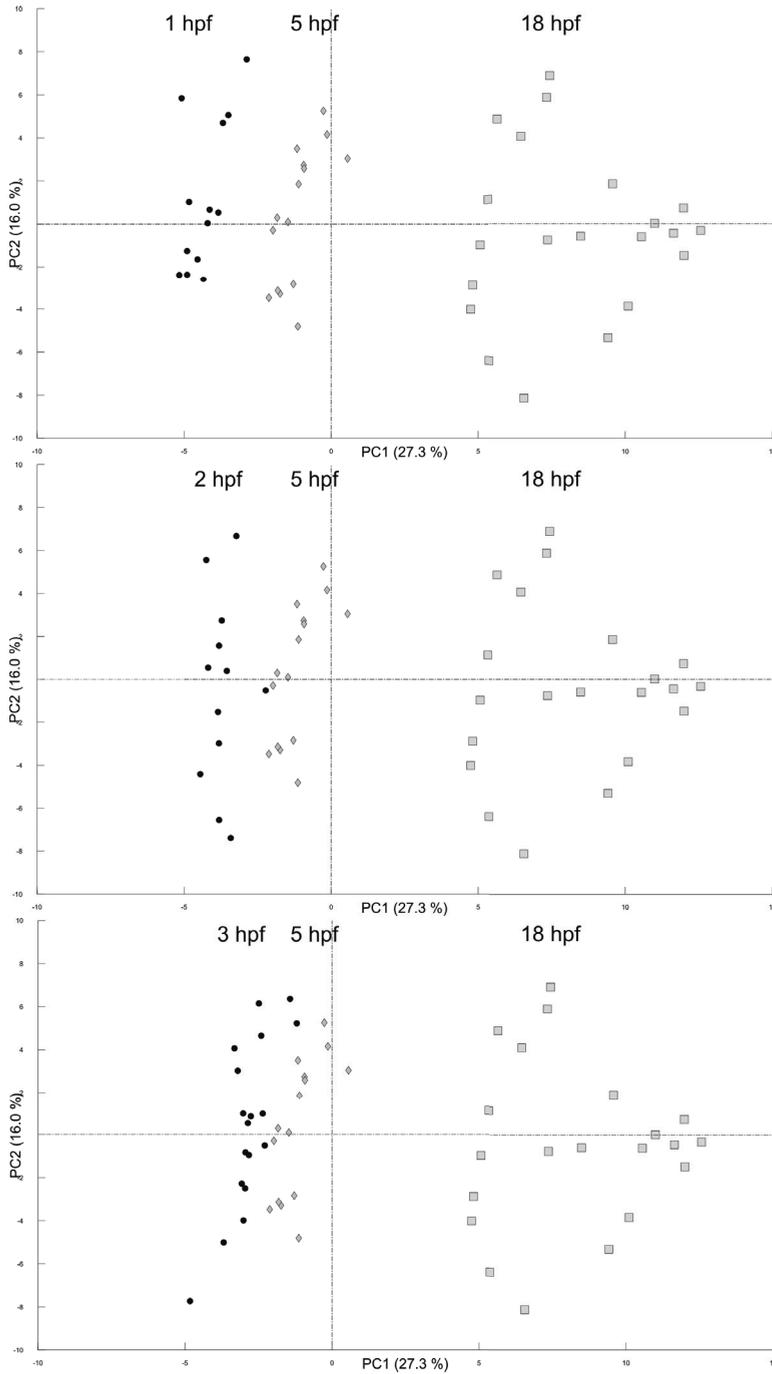


Figure 3: PCA score plots indicating a visible difference between 5 different developmental stages. For these plots only one PCA model was created using all samples and only those features with an RSD < 30%. The blocks represent the 18 hpf stage, the diamonds the 5 hpf stage. The circles represent the 1, 2 and 3 hpf respectively.

Table 1 shows all the putatively identified features, with their associated RSD values. Remarkably, the potassium adducts nearly always shows the best RSD, followed by the sodium adducts and finally the protonated molecule. As our samples were not acidified, alkali adducts were not suppressed. The reason for the pre-dominant potassium adducts is most likely due to relative high potassium concentration inside the cell as opposed to the relative high concentration of sodium outside living cells. From several of the 27 (putatively) identified features the boxplots with their up-or down regulation are shown in Figure 4. Still 63% of the 102 significant features remained unidentified. This confirms that further research is required to expand the zebrafish metabolome database in order to increase identification using DI-MS methods[16].

Biological relevance

In order to evaluate the possibility of using this technology for early fingerprinting of zebrafish embryos, different stages after fertilization were analyzed using this approach. 1, 2, 3, 5 and 18 hours post fertilization (hpf) were tested. The PCA in Figure 3 shows the somite stage (18 hpf) as the most distinctive group as well as having the most within-group variation. The fact that this stage is most distinctive from the other groups can be explained by the fact that different tissue types and organs (i.e. nervous system, skin, blood, and heart) begin to form at this stage. Since small time differences result already in differences of organ formation this will lead to biological variations in the different samples, explaining why this time point has the most variation in the metabolome measurements. Thus, as Chen *et al.* [16] remarked, the differences in the metabolome between the early embryonic stages may reflect the embryological properties of the cells. These results show that the zebrafish embryonic metabolome reflects differentiation. With our method we were able to observe a post fertilization time-trend with a time resolution of 1 hpf. The metabolic shift between 1, 2, 3 and 5 hpf can be clearly observed indicating that in early embryogenesis the metabolome changes quickly and significantly. As can be seen, the metabolome of the first three stages (1, 2 and 3 hpf) closely resemble each other. This suggests that zygotic gene transcription only begins at the onset of the midblastula transition (3 hpf)[27].

The boxplots in Figure 4 allows us to go more into depth regarding the biology during early embryogenesis. Interestingly, the increase in concentration of dimethylarginine can be a result of the enhanced methylation and acetylation metabolism of lysine and arginine which play an important role in histone activity and gene expression[28][29]. When we searched for other methylated and acetylated forms of lysine and arginine by discarding the described missing data cleaning step (resulting in a larger data subset) we also found increasing trends for dimethyllysine, methyllysine and acetyllysine (Supporting Figure 1). Although we are aware these analytes did not make it to the best 102 features, these findings correlate to the onset of gene expression at 3 hpf.

The increased concentration of spermine dialdehyde could reflect the turnover of the polyamine spermine which plays a role in normal and neoplastic growth as well as the (uni)directional transport of molecules by GAP junctions which is an important process in early embryogenesis[30][31].

It can also be observed that the concentrations of several amino acids and biogenic amines are increasing during early embryogenesis. Apparently the embryos are able to release amino acids from storage proteins to provide the cells with building blocks and energy already in an early stage. Some of these reveal more than a 10-fold change going from 1 hpf to 18 hpf. In the 18 hpf stage the rise of most of the amino acids is the largest. This could be explained by the

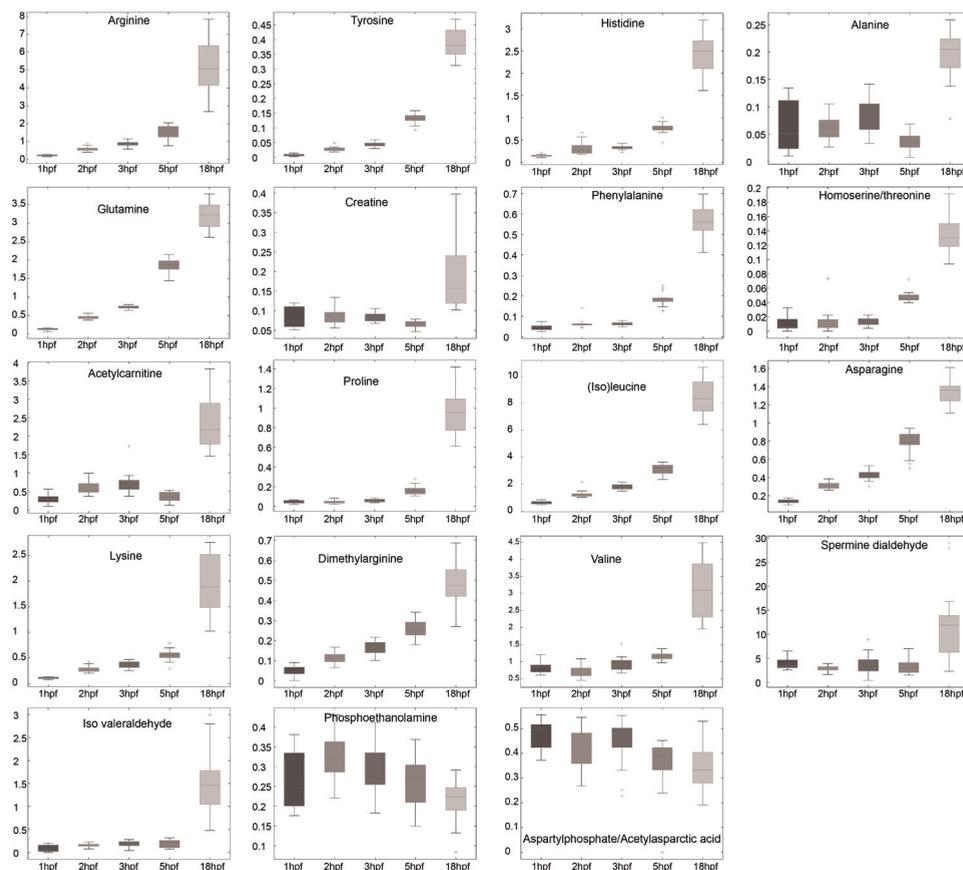


Figure 4: Boxplots of the up- and down regulation of some of the (putatively) identified metabolites through the five different developmental stages (for arg, his, actylcarnitine, lys, dimethylarg and spermine dialdehyde the $(M+H)^+$ plot is shown, for tyr, glu, creatine, phe, pro, leu, val, ala, homoserine, asparagine, phophoethanolamine, acetylaspartic acid the $(M+K)^+$ plot is shown, and for isovaleraldehyde the $(M+NH_4)^+$ plot is shown. The y-axis represents the ratio of the analyte/optimal internal standard.

increase in energy requirement for evoked muscle contraction starting at around 18 hpf[32]. Isovaleraldehyde, vinylacetylglucine could reflect the degradation of branched chain amino acids like valine and (iso)leucine, both associated with the increase in energy requirement. The concentration of hypoxanthine (see supplemental figure) is decreasing which can be explained by the enhanced DNA/RNA synthesis via xanthine, as also indicated by the decrease of guanine.

Some metabolites, like acetylcarnitine and creatine, show up- as well as- down regulation within the five developmental stages. The observed trends of almost all of our identified features is supported by the observations of previous publications[15][17]. Moreover, features that are indicated as new in table 1(arginine, acetylaspartic acid, carnitine, dimethylarginine, dopamine, FAPy-adenine, indoleacetic acid, isovaleraldehyde, phosphoethanolamine, quinone, safrole, spermine dialdehyde, vinylacetylglucine) were exclusively found with our method and not in the aforementioned references. This indicates the potential of HR-DI-MS for metabolic profiling purposes.

Deyolking

To obtain more insight in the biology of early embryogenesis, a series of deyolking experiments were performed in order to zoom in on the localization of metabolism. Figure 5 shows several bar-plots of the group means of metabolites that were also (putatively) identified in the previous section. Hypoxanthine and carnitine (metabolites which were discarded using our data cleanup steps (Supporting Figure 1)) showed a down-regulation trend during embryogenesis).

The charts confirm the same developmental trends as observed earlier. Because of the limited number of replicates that was measured (3 times) the statistical power is limited but ANOVA calculations showed that no significant difference between the zygote part and the whole egg could be detected. This could indicate that most of the metabolites and most of the metabolic conversions take place in the zygote part.

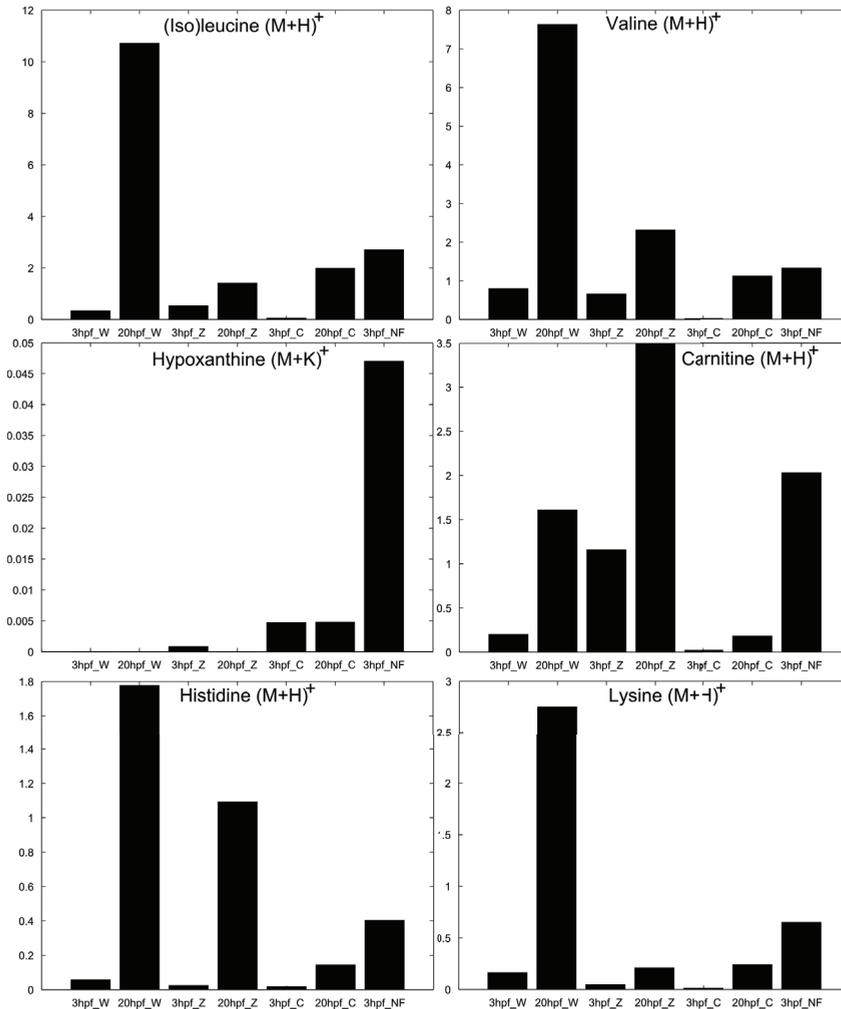


Figure 5: Bar-plots showing the mean of the different embryonic location at the age of 3 hpf and 20 hpf. W=whole embryo; C=chorion membrane; Z=zygote part; NF=dechorionated non-fertilized embryo (=yolk). The y-axis represents the ratio of the analyte/ optimal internal standard.

CONCLUDING REMARKS

Using our rapid metabolic fingerprinting method we were able to distinguish metabolic profiles of early developmental stages of zebrafish embryos. Interestingly, the onset of gene expression and the increase in energy requirement is reflected by the measured metabolome confirming that from a systems biology point of view metabolomics complements transcriptomics. After data cleanup, only those features were selected that showed consistent behavior within each developmental stage resulting in 102 features. PCA revealed that periods of 1 hour time shifts post fertilization could be differentiated from each other. In total 27 out of the 102 features were (putatively) identified. Although unambiguous identification is beyond the scope of our approach, identification on 6 of these 27 extracted features was pursued using standards and tandem MS. Several trends of the putatively identified metabolites are included and almost all of these findings are supported by previous publications. Moreover our method exclusively found several new features. By de yolking and de chorionating we showed the potential of this method to enable more in-depth studies on localization of metabolism. We conclude that HR-DI-MS is suitable for rapid metabolic profiling on zebrafish embryos. However, to improve robustness and obtain more high-quality features fast sample preparation methods are required.

ACKNOWLEDGEMENTS

This work is (co)financed by the Netherlands Metabolomics Centre (NMC) which is a part of The Netherlands Genomics Initiative/Netherlands Organization for Scientific Research and partly by the European community projects ZF-Health (FP7-Health-2009-242048).

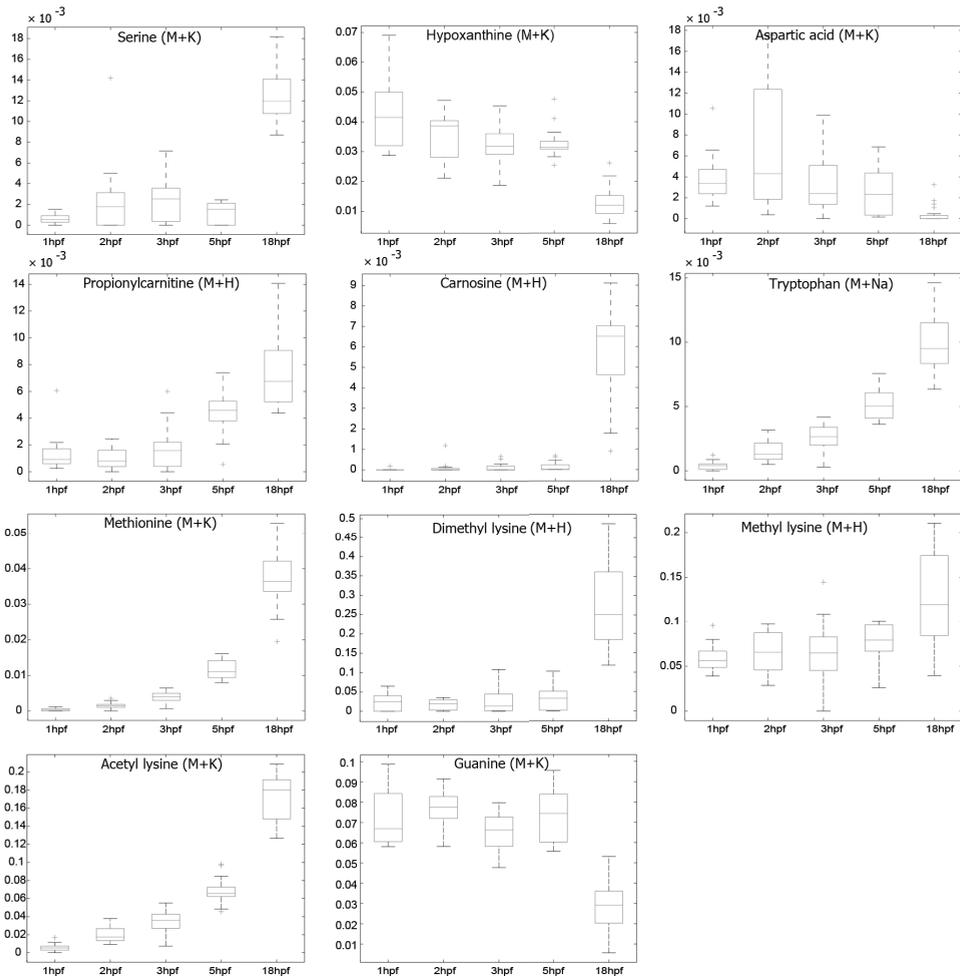
REFERENCES

- [1] S. Ali, D. L. Champagne, H. P. Spaink, and M. K. Richardson, "Zebrafish embryos and larvae: A new generation of disease models and drug screens," *Birth Defects Res. Part C Embryo Today Rev.*, vol. 93, no. 2, pp. 115–133, Jun. 2011.
- [2] A. H. Meijer and H. P. Spaink, "Host-pathogen interactions made transparent with the zebrafish model," *Curr. Drug Targets*, vol. 12, no. 7, pp. 1000–1017, 2011.
- [3] W. Goessling, T. E. North, and L. I. Zon, "New waves of discovery: modeling cancer in zebrafish," *J. Clin. Oncol.*, vol. 25, no. 17, pp. 2473–9, Jun. 2007.
- [4] M. J. M. Schaaf, W. J. A. Koopmans, T. Meckel, J. van Noort, B. E. Snaar-Jagalska, T. S. Schmidt, and H. P. Spaink, "Single-Molecule Microscopy Reveals Membrane Microdomain Organization of Cells in a Living Vertebrate," *Biophys. J.*, vol. 97, no. 4, pp. 1206–1214, 2009.
- [5] C. Pardo-Martin, T. Y. Chang, B. K. Koo, C. L. Gilleland, S. C. Wasserman, and M. F. Yanik, "High-throughput in vivo vertebrate screening," *Nat. Methods*, vol. 7, no. 8, pp. 634–636, 2010.
- [6] H. Sukardi, C. Y. Ung, Z. Gong, and S. H. Lam, "Incorporating Zebrafish Omics into Chemical Biology and Toxicology," *Zebrafish*, vol. 7, no. 1, pp. 41–52, 2010.
- [7] R. Carvalho, J. de Sonneville, O. W. Stockhammer, N. D. L. Savage, W. J. Veneman, T. H. M. Ottenhoff, R. P. Dirks, A. H. Meijer, and H. P. Spaink, "A high-throughput screen for tuberculosis progression," *PLoS One*, vol. 6, no. 2, p. e16779, Jan. 2011.
- [8] J. L. Anderson, J. D. Carten, and S. A. Farber, "Zebrafish lipid metabolism: from mediating early patterning to the metabolism of dietary fat and cholesterol," *Methods Cell Biol.*, vol. 101, pp. 111–41, Jan. 2011.
- [9] D. I. Ellis, W. B. Dunn, J. L. Griffin, J. W. Allwood, and R. Goodacre, "Metabolic fingerprinting as a diagnostic tool," *Pharmacogenomics*, vol. 8, no. 9, pp. 1243–66, Sep. 2007.
- [10] J. van der Greef, T. Hankemeier, and R. N. McBurney, "Metabolomics-based systems biology

- and personalized medicine: moving towards n = 1 clinical trials?," *Pharmacogenomics*, vol. 7, no. 7, pp. 1087–94, Oct. 2006.
- [11] E. Businge, K. Brackmann, T. Moritz, and U. Egertsdotter, "Metabolite profiling reveals clear metabolic changes during somatic embryo development of Norway spruce (*Picea abies*)," *Tree Physiol.*, vol. 32, no. 2, pp. 232–44, Feb. 2012.
- [12] S. S. Rubakhin, E. V Romanova, P. Nemes, and J. V Sweedler, "Profiling metabolites and peptides in single cells," *Nat. Methods*, vol. 8, no. 4 Suppl, pp. S20–9, Apr. 2011.
- [13] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin, "Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy," *Chem. Soc. Rev.*, vol. 40, no. 1, pp. 387–426, Jan. 2011.
- [14] E. S. Ong, C. F. Chor, L. Zou, and C. N. Ong, "A multi-analytical approach for metabolomic profiling of zebrafish (*Danio rerio*) livers," *Mol. Biosyst.*, vol. 5, no. 3, pp. 288–98, Mar. 2009.
- [15] S. Hayashi, S. Akiyama, Y. Tamaru, Y. Takeda, T. Fujiwara, K. Inoue, A. Kobayashi, S. Mae gawa, and E. Fukusaki, "A novel application of metabolomics in vertebrate development," *Biochem. Biophys. Res. Commun.*, vol. 386, no. 1, pp. 268–72, Aug. 2009.
- [16] C. Papan and L. Chen, "Metabolic fingerprinting reveals developmental regulation of metabolites during early zebrafish embryogenesis," *OMICS*, vol. 13, no. 5, pp. 397–405, Oct. 2009.
- [17] K. H. Soanes, J. C. Achenbach, I. W. Burton, J. P. M. Hui, S. L. Penny, and T. K. Karakach, "Molecular characterization of zebrafish embryogenesis via DNA microarrays and multiplat form time course metabolomics studies," *J. Proteome Res.*, vol. 10, no. 11, pp. 5102–17, Nov. 2011.
- [18] S. Hayashi, M. Yoshida, T. Fujiwara, S. Maegawa, and E. Fukusaki, "Single-embryo metabolomics and systematic prediction of developmental stage in zebrafish," *Z. Naturforsch. C.*, vol. 66, no. 3–4, pp. 191–8, Jan. 2011.
- [19] T. Fuhrer, D. Heer, B. Begemann, and N. Zamboni, "High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry," *Anal. Chem.*, vol. 83, no. 18, pp. 7074–80, Sep. 2011.
- [20] G. Madalinski, E. Godat, S. Alves, D. Lesage, E. Genin, P. Levi, J. Labarre, J.-C. Tabet, E. Ezan, and C. Junot, "Direct Introduction of Biological Samples into a LTQ-Orbitrap Hybrid Mass Spectrometer as a Tool for Fast Metabolome Analysis," *Anal. Chem.*, vol. 80, no. 9, pp. 3291–3303, 2008.
- [21] M. Beckmann, D. Parker, D. P. Enot, E. Duval, and J. Draper, "High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry," *Nat. Protoc.*, vol. 3, no. 3, pp. 486–504, 2008.
- [22] J. Draper, A. J. Lloyd, R. Goodacre, and M. Beckmann, "Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review," *Metabolomics*, vol. 9, no. S1, pp. 4–29, Jul. 2012.
- [23] A. El-Faramawy, K. W. M. Siu, and B. A. Thomson, "Efficiency of nano-electrospray ionization," *J. Am. Soc. Mass Spectrom.*, vol. 16, no. 10, pp. 1702–1707, 2005.
- [24] F. M. van der Kloet, I. Bobeldijk, E. R. Verheij, and R. H. Jellema, "Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping," *J. Proteome Res.*, vol. 8, no. 11, pp. 5132–5141, 2009.
- [25] O. Hrydziuszkowski and M. R. Viant, "Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline," *Metabolomics*, vol. 8, no. S1, pp. 161–174, Oct. 2011.
- [26] D. L. Massart, B. G. Vandeginste, L. M. Buydens, P. J. Lewi, J. Smeyers-Verbeke, and S. De Jong, *Handbook of Chemometrics and Qualimetrics*. Elsevier Science Inc., 1998, p. 1616.
- [27] D. A. Kane and C. B. Kimmel, "The zebrafish midblastula transition," *Development*, vol. 119, no. 2, pp. 447–456, 1993.
- [28] A. Shilatifard, "Chromatin Modifications by Methylation and Ubiquitination: Implications

- in the Regulation of Gene Expression,” *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 243–269, 2006.
- [29] Y.-C. Wang and C. Li, “Evolutionarily conserved protein arginine methyltransferases in non-mammalian animal systems,” *FEBS J.*, vol. 279, no. 6, pp. 932–45, Mar. 2012.
- [30] A. E. Pegg, “Spermidine/spermine-N-1-acetyltransferase: a key metabolic regulator,” *Am. J. Physiol. Metab.*, vol. 294, no. 6, pp. E995–E1010, 2008.
- [31] M. Watanabe, D. Watanabe, and S. Kondo, “Polyamine sensitivity of gap junctions is required for skin pattern formation in zebrafish,” *Sci. Rep.*, vol. 2, p. 473, Jan. 2012.
- [32] L. Saint-Amant and P. Drapeau, “Time course of the development of motor behaviors in the zebrafish embryo,” *J. Neurobiol.*, vol. 37, no. 4, pp. 622–32, Dec. 1998.

SUPPORTING INFORMATION



Supporting Figure 1: Boxplots of some of the discarded features due to our data cleanup steps.

