Functional mapping of the zebrafish early embryo proteome and transcriptome

Alli Shaik, Asfa; Wee, Sheena; Li, Rachel Hai Xia; Li, Zhen; Carney, Tom J.; Mathavan, Sinnakaruppan; Gunaratne, Jayantha

2014

Alli Shaik, A., Wee, S., Li, R. H. X., Li, Z., Carney, T. J., Mathavan, S., et al.(2014). Functional mapping of the zebrafish early embryo proteome and transcriptome. Journal of proteome research, 13(12), 5536-5550.

https://hdl.handle.net/10356/107041

https://doi.org/10.1021/pr5005136

Journal of proteome research

Article

pubs.acs.org/jpr

Functional Mapping of the Zebrafish Early Embryo Proteome and Transcriptome

Asfa Alli Shaik,[†,⊥] Sheena Wee,[†,⊥] Rachel Hai Xia Li,[†] Zhen Li,[‡] Tom J. Carney,[†,§] Sinnakaruppan Mathavan,[‡] and Jayantha Gunaratne*[,†,§,∥]

[†]Institute of Molecular and Cell Biology, Agency for Science Technology and Research, 61 Biopolis Drive, 138673, Singapore

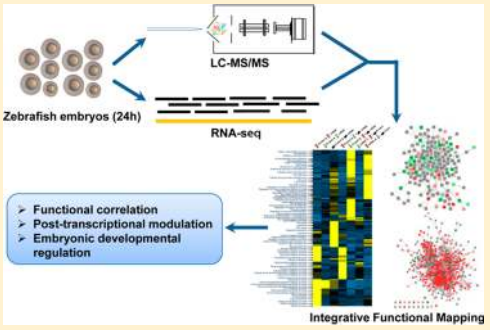
[‡]Genome Institute of Singapore, Agency for Science Technology and Research, 60 Biopolis Street, 138672, Singapore

[§]Lee Kong Chian School of Medicine, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore

[∥]Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Drive, 117597, Singapore

Supporting Information

ABSTRACT: Zebrafish is a popular system for studying vertebrate development and disease that shows high genetic conservation with humans. Molecular level studies at different stages of development are essential for understanding the processes deployed during ontogeny. Here, we performed comparative analysis of the whole proteome and transcriptome of the early stage (24 h post-fertilization) zebrafish embryo. We identified 8363 proteins with their approximate cellular abundances (the largest number of zebrafish embryo proteins quantified thus far), through a combination of thorough deyolking and extensive fractionation procedures, before resolving the peptides by mass spectrometry. We performed deep sequencing of the transcripts and found that the expressed proteome and transcriptome displayed a moderate correlation for the majority of cellular processes. Integrative functional mapping of the quantified genes demonstrated that embryonic developmental systems differentially exploit transcriptional and post-transcriptional regulatory mechanisms to modulate protein abundance. Using network mapping of the low-abundance proteins, we identified various signal transduction pathways important in embryonic development and also revealed genes that may be regulated at the post-transcriptional level. Our data set represents a deep coverage of the functional proteome and transcriptome of the developing zebrafish, and our findings unveil molecular regulatory mechanisms that underlie embryonic development.



KEYWORDS: *Zebrafish, proteomics, transcriptomics, early embryo, development, functional mapping*

INTRODUCTION

The zebrafish (*Danio rerio*) is an attractive experimental model organism for exploring the molecular mechanisms of vertebrate development.[1] Mutant phenotypes commonly emulate the pathology/phenotype of human diseases and disorders, hence making it a premier model for genetic and phenotypic analysis.[2−4] Genetic screens have also broadened our understanding of the various factors that control cell differentiation and fate as well as organogenesis, allowing us to chart the sequential events involved during the transition from embryo to adult.[5] There are more than 26 000 coding regions in the zebrafish genome, many of which are orthologous to those in humans.[6,7] To understand the complex interplay among the expressed genes, large-scale analysis that extensively captures the expression variation at the mRNA and protein levels is important.[8−11] A recent systematic analysis identified a total of over 56 000 transcripts, including alternative splice variants, during zebrafish embryogenesis.[12] However, the maximum number of proteins that have been identified to date has been limited to only 5267 and 8475 in embryos and adults, respectively.[13,14]

Although transcript abundance can provide valuable information on the status of the cell at any point in time, proteins are the fundamental biological effectors that orchestrate key events within the cells. Gene expression patterns derived from large-scale transcriptomics, including those involving microarrays and RNA-seq, have been routinely used to estimate protein abundance. However, only a modest correlation has been observed between mRNA and protein levels across different species from yeast through higher eukaryotes.[15−18] A recent comparative transcriptomic and proteomic study in the late-stage zebrafish embryo suggests that such differences between transcript and protein levels may underlie important translational and post-translational regulatory mechanisms.[19] Hence, the need for a thorough representation of the proteome is increasingly recognized.

Currently, mass spectrometry-based shotgun proteomics is the only available high-throughput method for identification and quantification of the whole proteome. In recent years,

Received: May 26, 2014

   dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

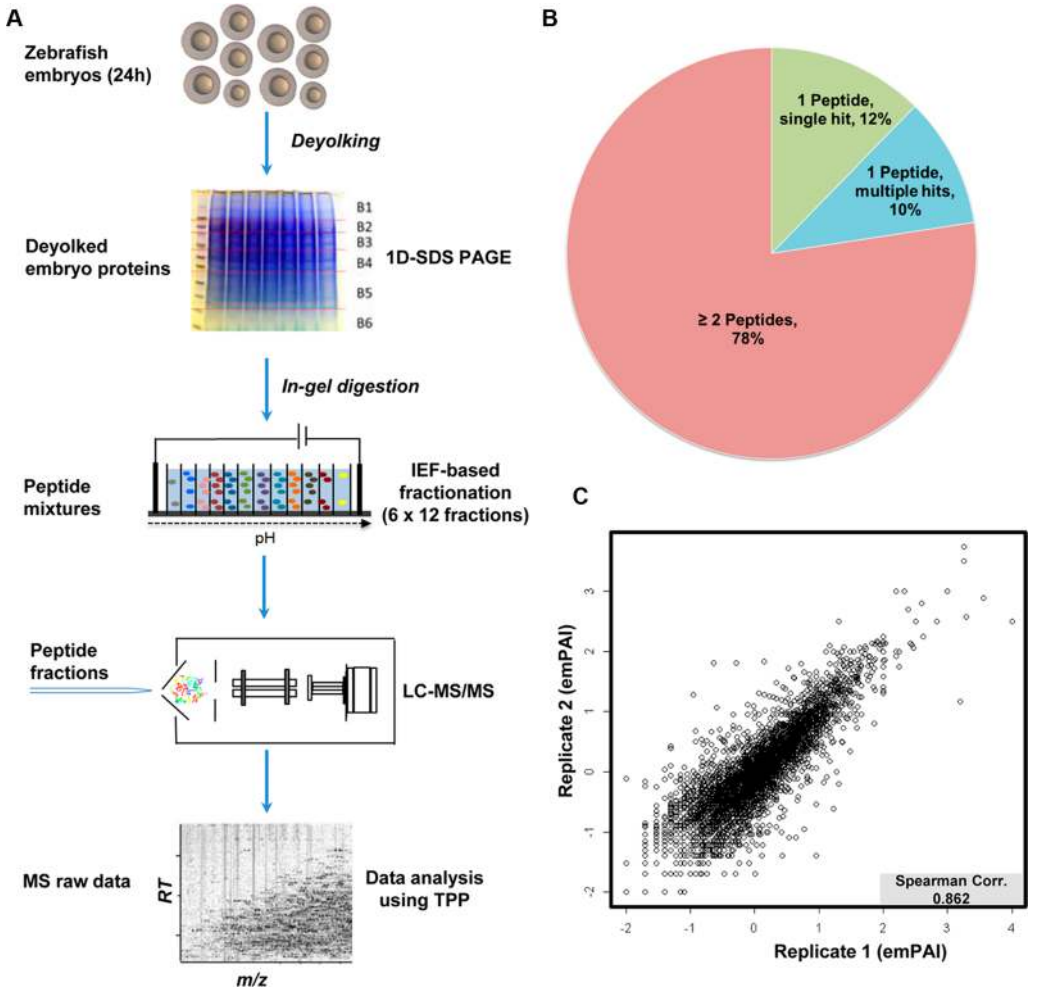


Figure 1. Zebrafish embryo proteome analysis. (A) Deep proteome analysis workflow consisted of extensive fractionation of deyolked zebrafish embryo proteins through SDS-PAGE (protein level) and isoelectric focusing (IEF; peptide level). LC−MS/MS data was processed using the Trans-Proteomic Pipeline (TPP). (B) Summery of identified protein clusters based on number of unique peptides. (C) Correlation of the quantified protein abundances (emPAI) between the two biological replicates, each of which was characterized by 72 fractions run in two technical replicates.

global proteomic profiling of zebrafish adults and embryos has been extensively carried out, primarily to understand developmental processes as well as to recapitulate disease mechanisms.[10,20−24] With zebrafish being recognized as a powerful model for chemical toxicity and drug safety assessment, proteomics-based methods along with transcriptomics are increasingly applied for large-scale system-wide studies.[25−29] Although applications of quantitative proteomic approaches have been limited in zebrafish, recent studies used stable isotope-labeled zebrafish for studying cardiac morphogenesis and profiling various organs in the adult,[30,31] indicating the possibility of performing large-scale quantitative proteomics studies in zebrafish in the future.

Despite the gaining popularity of proteomic studies in zebrafish, thorough protein identification is largely dependent on sample complexity and the dynamic range of the proteins within the sample. However, in the case of zebrafish, this endeavor is even more challenging owing to the high proportion of yolk proteins, particularly during the early stages of development.[32] Hence, most of the proteomic studies in zebrafish have been performed in late-stage embryos or adults.[33] Early embryonic stages are highly dynamic in nature and are marked by events that accompany cell differentiation and morphogenesis.[34] A thorough representation of the proteome during these stages is essential to map the key biological events that occur during embryogenesis.

Here, we report a comprehensive map of the quantitative proteome profile of early stage zebrafish (24 h post-fertilization embryo) containing 8363 proteins, the highest number of proteins reported so far for early embryonic stages. We establish that the protein functions are linked to their abundances, wherein high-abundance proteins are predominantly associated with cellular core functions and low-abundance proteins perform regulatory functions that mediate development. The high coverage proteome was also compared to the corresponding transcriptome derived from the same early embryonic stages to provide a comprehensive functional map of the quantified proteome and transcriptome. The integrative approach identified biological processes that are modulated differently by transcript and protein levels in the early stage zebrafish embryo.

MATERIALS AND METHODS

Sample Preparation

Adult zebrafish were maintained on a 14 h light/10 h dark cycle at 28 °C in the AVA (Singapore) certified IMCB zebrafish facility. Zebrafish embryos were obtained through crosses of

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

TL/AB hybrid parents, and zebrafish were raised at 28 °C in zebrafish embryo medium. At 24 h post-fertilization (hpf), approximately 1000 embryos were dechorionated with Pronase and subsequently washed extensively in embryo medium. Deyolking was performed largely as per Link et al.,[35] using three washes in calcium-free Ringer's solution with mechanical disruption through a flame-narrowed glass Pasteur pipet. The resulting cell pellet was lysed by brief sonication in CSH buffer (50 mM Tris-HCl, 250 mM NaCl, 1 mM EDTA, 1% NP40) supplemented with protease inhibitors (Roche). Insoluble material was removed by centrifugation. Protein concentration was determined using the BCA protein assay, reducing agent compatible (Thermo Fisher Scientific).

SDS-PAGE and In-Gel Digestion

Five-hundred micrograms of lysate, obtained from two biological replicates, was separated on a SDS-PAGE gel from 3.5 to 260 kDa using a NuPAGE 4−12% Bis-Tris 1.0 mm, 10 well gel (Invitrogen). The gel was cut into 6 bands as shown in Figure 1A, and each band was excised to 1.5 × 1.5 mm^2 using a scalpel. In-gel digestion was carried out as previously described with minor modifications.[36] Briefly, the gel pieces were washed with 3 mL of 50 mM ammonium bicarbonate. Reduction was carried out by the addition of 10 mM DTT, covering the gel pieces, and incubation for 30 min at 56 °C; alkylation was performed with the addition of 55 mM iodoacetamide and incubation for 20 min in the dark at room temperature. The gel pieces were destained with 6 mL of 50% acetonitrile/25 mM ammonium bicarbonate for 10 min. Six milliliters of 100% acetonitrile was used to shrink the gel pieces, which was done twice for 10 min each. One to two milliliters of 13 ng/μL sequencing-grade trypsin (Promega) was added to each well, and the gel pieces were allowed to swell for 60 min at 4 °C before enough 25 mM ammonium bicarbonate was added to cover the gel pieces. The samples were incubated for 3 h at 37 °C. All supernatants were collected by centrifugation. One and a half milliliters of 5% formic acid was added to each well followed by 1.5 mL of 100% acetonitrile for peptide extraction. Both steps were repeated.

Off-Gel Isoelectric Focusing

Off-gel isoelectric focusing (IEF) was carried out using a 3100 OFFGEL fractionator (Agilent) as described in the manufacturer's manual with slight modifications. In short, 13 cm IPG strips for the pH range of 3−10 (GE Healthcare) were used, resulting in 12 peptide fractions. The concentration of glycerol and IPG buffer pH 3−10 (GE Healthcare) was halved in the peptide OFFGEL stock solution as described previously.[37] The voltage gradient during the run was 250 V for the first hour followed by a gradient of up to 1000 V over the next 2 h and 1000 V for an additional hour. Then, the voltage was increased up to 3000 V over the next 7 h and held at 3000 V until a total voltage of 20 000 V hours was reached. After the run, 30 μL of 1% TFA was added to each well for acidification.

NanoHPLC−ESI−MS/MS

Peptides resulting from the different fractionation methods were desalted using self-packed C18 StageTips.[38] The C18 StageTip was conditioned with 100 μL of methanol followed by 100 μL of 0.1% formic acid at 6000*g* for 2 min. The extracted peptides were loaded onto the C18 StageTips and washed with 100 μL of 0.1% formic acid. The peptides were eluted with 60 μL of 0.1% formic acid/80% acetonitrile. All eluents were dried using a SpeedVac and reconstituted in 12 μL of 0.1% formic acid. A total of 144 IEF fractions (72 fractions per biological replicate) were analyzed in duplicate using an EASY-nLC (Proxeon) coupled to a LTQ Velos (Thermo Fisher Scientific). Samples were directly loaded at 400 nL/min onto a PicoFrit column (HALO, C18, 90 Å, 2.7 μm, 75 um (i.d.) × 100 mm length) (New Objectives). The HPLC gradient was created using buffer A consisting of 2% acetonitrile/0.1% formic acid and buffer B consisting of 80% acetonitrile/0.1% formic acid: buffer B was increased from 0 to 8% over the first 4 min, followed by an increase to 25% over the next 58 min, an increase to 45% over the subsequent 32 min, an increase to 70% over the following 10 min, and an increase to 100% over the next 3 min. This condition was maintained for 5 min. Buffer B was then decreased to 5% over the subsequent 3 min and retained at 5% for another 5 min. This results in a HPLC gradient run of a total of 120 min. The flow rate was 250 nL/min for the first 104 min and 400 nL/min for the last 16 min. MS analysis was online-coupled to the LC using a LTQ Velos with the following settings: MS scans ranging from 300 to 1600 m/z, AGC target of 3×10^4, and maximum injection time of 10 ms. The 10 most intense ions with an ion intensity above 1000 and a charge state excluding one were sequentially isolated to a maximum AGC target value of 4×10^4 for a maximum of 100 ms and fragmented by collision induced dissociation (CID) using a normalized collision energy of 35%. A dynamic exclusion list was applied using an exclusion list size of 500, one repeat count, a repeat duration of 45 s, an exclusion duration of 90 s, and mass widths of 1.0 (low) and 1.5 (high). Expiration count was set to 3, and its S/N threshold, to 3.0.

Data processing and emPAI Calculation

All Velos raw data were first converted to peak lists in the centroid mzXML file format and then to the mgf file format. The conversion was performed with ReAdW.exe (version 4.0.2), which is part of the Trans-Proteomic Pipeline (TPP) (version 4.4.0).[39]

A target-decoy database was compiled using Sequence Reverser (part of MaxQuant v1.0.13.13) with the ipi.danre.v3.85.fasta downloaded from ftp://ftp.ebi.ac.uk/pub/databases/IPI/last_release/current/ and 262 contaminant sequences in Sequence Reverser. The final database containing 81 476 sequence entries was searched on the mgf peak list files using Mascot (version 2.3).

Mascot search parameters were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues at two peptide termini), two missed cleavages were allowed, carbamidomethylation (C) was set as fixed modification, and acetyl (protein N-term) and oxidation (M) were set as variable modifications. Peptide charge was set to 2+, 3+, and 4+. Mass tolerance of the precursor ion and the fragment ions was set at 2 and 0.5 Da, respectively.

All of the mascot search outputs were combined in TPP (version 4.4.0). First, mascot outputs (dat file) were converted to pepxml file format. Then, PeptideProphet with a minimum length of 7aa, a probability of 0.9, and an accurate mass model was applied. iProphet was used to integrate all of the PeptideProphet results.[40] Finally, proteins were assembled with PeptideProphet on the iProphet results with a minimum probability of 0.9.

The PeptideProphet output with peptide count was used to calculate the relative protein abundances. Relative protein abundances were calculated using the emPAI algorithm as described by Ishihama et al.[41]

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

Analysis of Detected Protein Bias

The IPI identifiers of all of the detected proteins were mapped to the corresponding UniProt IDs and subjected to protein parameter analysis in ExPASy (http://www.expasy.ch). The ProtParam tool was used to calculate the protein length and pI values. For mapping the chromosomal bias of the detected proteins, the Entrez database (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA) and the annotations from the Zv9, as implemented in the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTables), were used.

RNA-Seq Analysis

Zebrafish embryos (wild-type AB line) were collected and incubated at 27 °C. Synchronously developing embryos were collected at 24 hpf and frozen in liquid nitrogen. The frozen embryos were used for RNA extraction. Total RNA was extracted using TRIzol reagent (Invitrogen, USA). RNA concentration was determined using a NanoDrop 2000 (Thermo Scientific), and 60 μg of total RNA was used as starting material. The integrity of the RNA samples was determined using an Agilent RNA 6000 Nano chip on an Agilent 2100 Bioanalyzer. The RNA sample with RIN > 9.0 was selected for mRNA purification using the MicroPoly(A) purist kit (Ambion, USA). Five-hundred nanograms of mRNA was applied for RNA library construction with the solid total RNA-seq kit (ABI, USA) according to the manufacturer's instructions. RNA was sequenced in a SOLiD3 (ABI) platform, generating 50 bp single-ended reads. We generated about 40 million tags for this library and mapped them to the genome (ZV9). The RNA-seq reads were mapped in a strand-specific manner to the reference seq genes (RefSeq), and the expression is presented as reads per kilobase of exon per million reads mapped (RPKM).

Comparison with RNA-Seq Data

The complete RNA-seq data comprised 10 101 transcripts with RPKM abundance greater than 2. The IPI identifiers of the quantified proteome were mapped to Entrez nucleotide identifiers and subsequently to the corresponding genes. Five-thousand two-hundred and fifty four IPI identifiers from the total 8363 quantified proteins could be successfully mapped to a corresponding transcript from RNA-seq. Some of the IPI identifiers could be mapped to more than one gene identifier. Excluding the events of alternative splice variants, a total of 5084 different protein-coding genes could be mapped. The anatomical enrichment of the quantified proteome and transcriptome was carried out using DAVID.[42]

GO Pathway Analysis-Based Clustering for Protein and Transcript Groups

The high and low protein abundance groups were identified on the basis of the quantile density distribution of the emPAI and RPKM values across the quantified proteome and transcriptome, respectively. Genes in the top 20% quantile (upper) with respect to the abundance values were categorized into high-abundance groups (very high + high cluster), and those in the bottom 20% (lower) comprised the low-abundance groups (very low + low cluster). The remaining genes were considered to be expressed at moderate levels. For analyzing the enrichment across the high- and low-abundance proteins in accordance with the GO terms, biological process, molecular function, and cellular component (BINGO), as implemented in Cytoscape, was used.[43] The enrichment was done using hypergeometric testing followed by Benjamini–Hochberg false discovery rate correction. The frequency of over-representation in the high- or low-abundance groups was calculated by comparing against the enrichment across the whole quantified proteome. For integrative proteomic and transcriptomic analysis based on GO categorization (biological process), the genes were first categorized according to their abundance values into different groups. The GO enrichment along with their p values was obtained for each of the groups and then filtered to retain only those groups that were significantly enriched (p value <0.05) in at least one of the analyzed groups. The filtered p values were log-transformed and z-score normalized before being subjected to hierarchical clustering based on Euclidean distance and average linkage.

MicroRNA Prediction

Genes that displayed low protein levels and high mRNA levels were analyzed for potential microRNA (miRNA) regulation using TargetScanFish version 6.2.[44,45] Only those predicted miRNA families with a target score (total context score) ≤ -0.3 were considered to be reliable. For genes with multiple 3′ UTRs (untranslated region), the predictions were specifically carried out on those that are curated to be expressed at 24 hpf developmental stage.

Network Analysis

The proteins in the high- or low-abundance groups that could be distinctly mapped to corresponding transcripts were used for network reconstructions. Protein–protein interactions, as implemented in reactome functional interaction (reactome FI) in Cytoscape visualization software and in GeneGO MetaCore, were used to unravel the connectivity between the protein groups.[46,47] Zebrafish shares many orthologous genes and pathways with other vertebrate species and hence the human orthologous proteins corresponding to the zebrafish genes were identified from ZFIN (http://zfin.org/) and InParanoid (http://inparanoid.sbc.su.se/) databases and used for network analysis.[48] Only interactions between the quantified proteins were retained, and other linker candidates (not in our data set) were excluded. The direct interactions from MetaCore were downloaded and parsed into Cytoscape as a SIF (simple interaction file) network. This was combined with the functional interaction network derived from reactome FI, and subsequent pathway enrichment was performed for the combined network. The densely connected regions in the network were identified using molecular complex detection (MCODE) algorithm.[49] The highest-ranking modules were extracted and visualized. MCODE could not be successfully applied to the low-abundance protein groups owing to the less-dense nature of the network. Hence, the clusters were visualized purely on the basis of their significant pathway enrichment.

RESULTS

Extensive Analysis of Zebrafish Embryo Proteome

To generate an extensive map of the zebrafish proteome, lysates were obtained from deyolked embryos representing the 24 h post-fertilization (hpf) developmental stage and resolved on a 1D SDS-PAGE gel. Deyolking ensures that high-abundance yolk proteins that would otherwise interfere with the deep mining of zebrafish embryos are depleted.[33] After deyolking, the sample still consists of a complex mixture of proteins. In order to reduce this complexity, we carried out extensive fractionation of the deyolked protein mixture using 1D SDS-

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX–XXX

PAGE protein level fractionation followed by off-gel isoelectric focusing (IEF) of the tryptic peptides from each gel band, as these methods were observed to be the best among the other evaluated approaches in our previous study[13] (Figure 1A). A total of 72 IEF fractions obtained from two biological replicates each were finally subjected to LC−MS/MS analysis, summing to a total of 288 runs including those of the technical replicates. MS/MS data analysis was performed using the Trans Proteomics Pipeline, and a minimum probability of 0.9 was set for confident peptide assignment. The mass error in parts per million (ppm) for precursor ions of all identified peptides is shown in Figure S1 in the Supporting Information. After assembling the proteins using ProteinProphet, we identified a total of 8363 different proteins including splice variants (at a false discovery rate < 1.2%), which, to our knowledge, is by far the most comprehensive proteome map of the zebrafish embryo (Tables S1 and S2 in the Supporting Information). This translates to an improvement of more than 2-fold in proteome coverage in comparison to that in our previous report.[13] Of these proteins, 6475 proteins (78%) were detected by at least two peptides. About 10% of the proteins (857) were detected via the same single peptide multiple times, whereas about 12% (1031) were detected once by a single peptide (Figure 1B). The median number of peptides identified per protein was 4, and the corresponding tandem spectra detected per protein was 17. The median sequence coverage per protein was found to be 13.504%.

The coverage of each protein attained by the corresponding peptide matches can be used to estimate the abundance of the identified proteins. To calculate the approximate abundance of the proteins in the zebrafish embryo, we used the exponentially modified protein abundance index (emPAI) algorithm, which normalizes the number of sequenced peptides per protein by the number of theoretically observable peptides of the protein.[41] In order to conduct a comprehensive study of the zebrafish proteome, we explored the possibility of studying all detected proteins, including those that were detected in only one biological replicate. The emPAI-based semiquantitative protein abundance values showed a high correlation between the two biological replicates (Spearman's correlation, 0.862) (Figure 1C). We, therefore, merged the data from both biological replicates for subsequent analysis (Table S3 in the Supporting Information).

The ranked distribution of all identified proteins allowed for evaluation of individual protein contribution to the total mass. It is revealed that 97.2% of the total protein abundance is contributed by the most abundant 25% of identified proteins (Figure 2A). Ninety percent of the total quantified proteome is within a range of $\log_2$ emPAI between 2.82 and −3.64 around the median abundance value. For further analysis, the quantified proteins were categorized into five quantiles based on their protein abundance values. Accordingly, the upper quantile constituted the very high and high categories, representing greater than the 90th percentile and 80−90% quantile, respectively, and the lower quantile comprised the very low and low categories, representing less than the 10th percentile and 10−20% quantile, respectively, of the estimated protein abundance values (Figure S2 in the Supporting Information). The quantile corresponding to 20−80% was considered to be moderate.

Assessing the physiochemical features across all of the identified proteins revealed that proteins of shorter length (<100 amino acids (aa)) are more abundant than proteins that

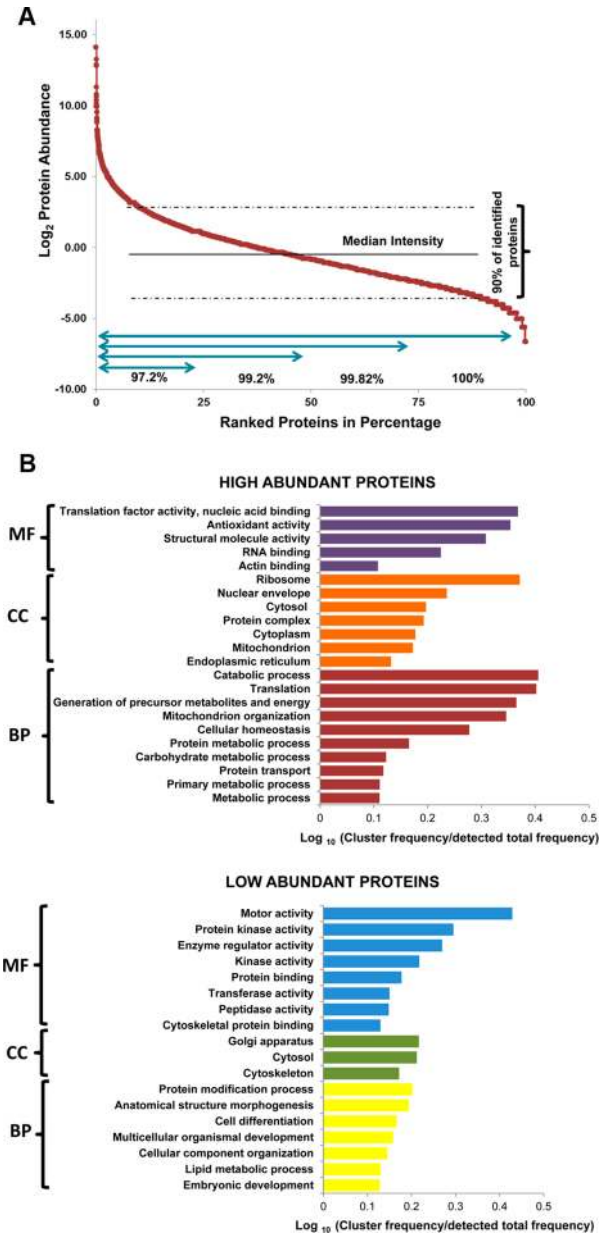


Figure 2. Quantitative analysis of expressed proteins. (A) Ranked protein abundances from highest to lowest across the global quantified proteome. The contribution of each of the ranked quantiles to the total quantified embryonic protein abundance is indicated. (B) Enrichment based on GO categorization in each of the high- and low-abundance protein groups. Frequency corresponds to the preferential enrichment against the enrichment of total detected proteins. The GO categories are represented as MF (molecular function), BP (biological process), and CC (cellular component).

are over 1500 aa in length (Figure S3A and Table S4 in the Supporting Information). On dissecting the length distribution within the individual clusters, we noticed that a majority of the highly abundant proteins (very high + high) are shorter than 450 aa (Figure S4A,B in the Supporting Information). The bias toward shorter length is also reported for highly abundant transcripts.[50] This serves as an efficient means of minimizing energy cost, and the short proteins, generally in high abundance, play key roles in various cellular process including signaling, cell−cell communication, and other basic metabolic processes.[51] The pI distribution, on the other hand, showed a

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

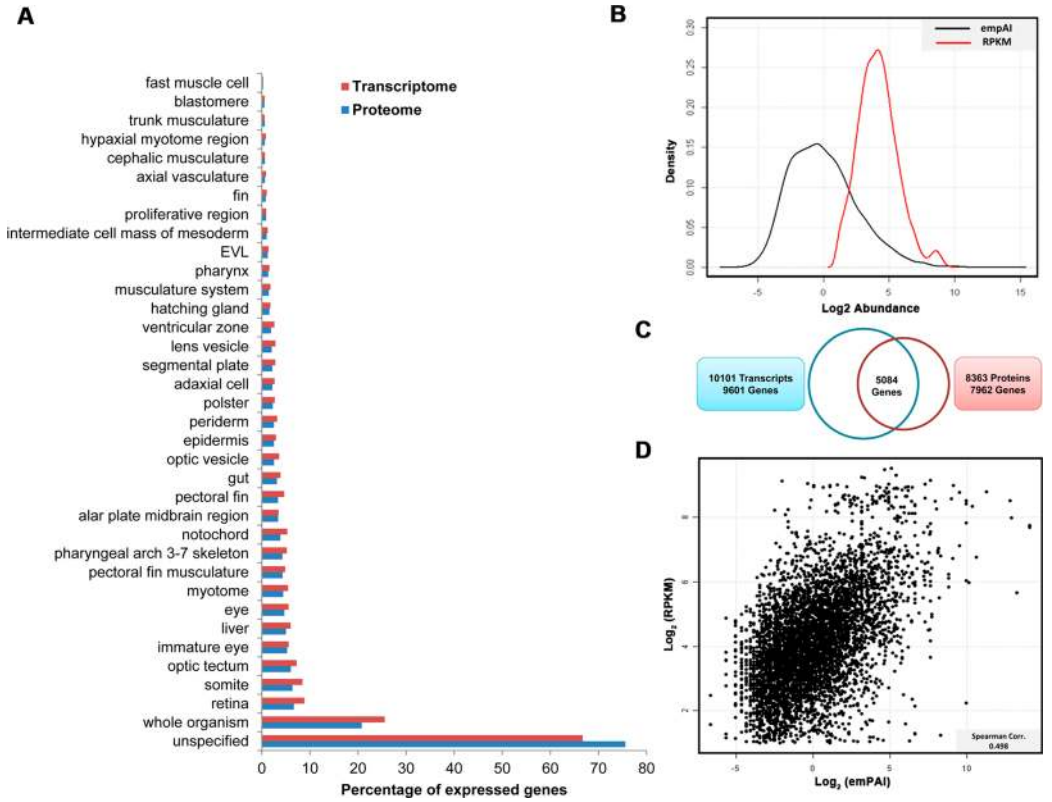


Figure 3. Comparison of quantified proteins and transcripts. (A) Distribution of enrichment of the quantified proteins and transcripts to distinct anatomical structures in zebrafish. A large proportion of the genes remained unmapped. (B) Distribution based on protein and transcript abundances as measured by empAI and RPKM, respectively, for those proteins with corresponding quantified transcripts. (C) Venn diagram of the number of genes quantified at the protein and mRNA levels and mapped to different protein-coding regions. (D) Density scatter plot of transcript versus protein abundances. The emPAI and RPKM are represented in log scale, and the Spearman correlation score is indicated.

drop in abundance for proteins with midrange pI values (Figure S3B and Table S4 in the Supporting Information). However, comparison of the pI value distribution between the upper and lower quantile displayed similar trends across the pI range, with an under-representation of proteins with basic pI values (Figure S4C,D in the Supporting Information).

To map the chromosomal distribution of the quantified proteome, the identified proteins were traced back to their genomic loci by mapping against the annotated protein-coding genes (Table S3 in the Supporting Information). However, only about 78.3% of the detected proteins could be successfully mapped to each of the 25 chromosomes (Table S5 in the Supporting Information). Most of the identified proteins mapped to genomic regions in chromosome 5, and the least number of identified proteins were represented by chromosomes 4 and 24 (Figure S5 in the Supporting Information). This is in agreement with the genome data that demonstrated relatively fewer protein-coding genes in chromosome 4.[6] On comparing the coverage of the identified proteins to the total protein-coding genes annotated for each of the chromosomes, we observed a maximum coverage of ~31.5% for chromosome 19, closely followed by ~30.4% for chromosome 5 (Figure S6A in the Supporting Information). The least coverage (~14.7%) was observed for chromosome 4, suggesting that the genes present here are not protein coding or do not express at 24 hpf. Intriguingly, chromosome 4 is unique in the large number of noncoding RNAs and repeats that it harbors as well as the presence of a large family of genes that are specific to *D. rerio*.[6] Of note, chromosomes 3, 6, 11, and 19 are particularly enriched in abundantly expressed (very high + high cluster) protein-coding genes (Figure S6B in the Supporting Information).

To obtain functional insights into the biological processes and cellular organization that are active at this developmental stage, we performed GO-slim analysis for the high- and low-abundance protein groups. Enrichment was performed by hypergeometric testing, and significant GO categories were identified in each group. The cluster frequencies of each of the significant GO categories were used to calculate the frequency of over-representation with respect to the overall quantified proteome, as shown in Figure 2B. We observed that proteins related to basic metabolic functions, primarily translation-related processes, are the most significantly enriched in the highly abundant protein cluster. Also, processes related to protein transport and organelles involved in trafficking, including the endoplasmic reticulum and nuclear envelope, are abundantly enriched. As observed in other systems, regulatory proteins associated with kinase activity, enzyme regulation, and protein binding have lower expression levels in the zebrafish embryo. At 24 hpf, the embryo is still in a very early stage of development, and important morphogenetic features including pigmentation, the cardiac tube, and fin fold begin to appear.[34] Accordingly, we observed that the lower quantile proteins are enriched in functions relating to cell differentiation, structure morphogenesis, and embryonic development. While metabolic processes related to protein and carbohydrates are functionally enriched, lipid metabolic processes are low in abundance.

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

Concordance with Transcript Abundance

It is widely appreciated that the regulation of a protein's level occurs at multiple levels beyond RNA transcription. To determine the extent of such regulatory systems, we sought to correlate our proteome data with the transcriptome of embryos at the same stage; hence, we performed RNA-seq on the 24 hpf embryos to determine transcript abundance (Table S6 in the Supporting Information). The RPKM measure obtained from the RNA-seq is a representation of transcript abundance. By including only those transcripts with an abundance greater than 2 and subsequently mapping the reads to the zebrafish reference genome (ZV9), a total of 10 101 unique transcripts were obtained. Some of the transcripts included alternative splice variants of the same genes. Thus, a total of 9601 different protein-coding genes were successfully identified. The different protein-coding genes showed a similar distribution among the 25 chromosomes as that from the quantified proteome. The maximum number of identified transcripts mapped to chromosome 5, similar to that observed for the proteome, and the least number of genes mapped to chromosome 24 (Figure S5 and Table S5 in the Supporting Information).

The quantified proteome and transcriptome data showed similar percentages across the different anatomical enrichment categories (Figure 3A). A majority of the transcripts and proteins (~65% for transcripts and ~75% for proteins), however, were not annotated to any specific anatomical feature and hence the distinct roles of these genes in the development of zebrafish remain to be explored. In fact, many of the known morphological developments that occur at 24 hpf, including the development of the retina, fin, and myotome, are represented with higher percentages of proteins, suggesting that the deyolking and extensive fractionation have enabled a thorough representation of the proteome.[34]

On comparing the transcriptome and proteome data on the basis of gene annotations, we observed that some of the IPI (International Protein Index) identifiers mapped to more than one transcript (Table S7 in the Supporting Information). In all, we identified a corresponding transcript for a total of 5254 proteins in our quantified proteome. On comparing the distribution of abundances for the overall quantified transcriptome, we observed that no proteins were identified for a considerable number of transcripts in the lower abundance range (Figure S7A,B in the Supporting Information). We also noticed that the distribution of protein abundance is broader than that of the corresponding transcript abundance values, although both of the abundance distributions share the same general shape (Figure 3B). Altogether, there was a 64% overlap between our proteome and transcriptome data on the basis of common protein-coding genes, excluding references of alternative slice variants (Figure 3C).

The RPKM and emPAI values are a proxy for the cellular abundance of transcripts and proteins, respectively, at a given point in time; hence, we analyzed the correlation between these two measurements. We observed a moderate correlation between the RPKM-based transcript abundance and emPAI-based protein abundance (Spearman correlation, 0.498) (Figure 3D). The level of correlation obtained is comparable to that observed previously in other organisms including human (Spearman correlation, 0.6), *Drosophila* (Spearman correlation, 0.66), *Caenorhabditis elegans* (the Spearman correlation, 0.59), and yeast (Spearman correlation, 0.58).[15−17] Although proteins modulate key events within the cells, up- or downregulation of mRNA from large-scale transcriptomic studies is directly associated with protein expression levels based on the assumption that there is high correlation between transcript and protein abundances. Comparative studies performed in various organisms, however, suggest that the correlation coefficients generally range between 0.3 and 0.6, highlighting that protein levels are regulated beyond transcription.[52,53]

We next analyzed the concordance between the transcript and protein abundance across functional categories that were arbitrarily grouped to represent core cellular and regulatory and developmental functions (Figure 4A,B). Translational process-

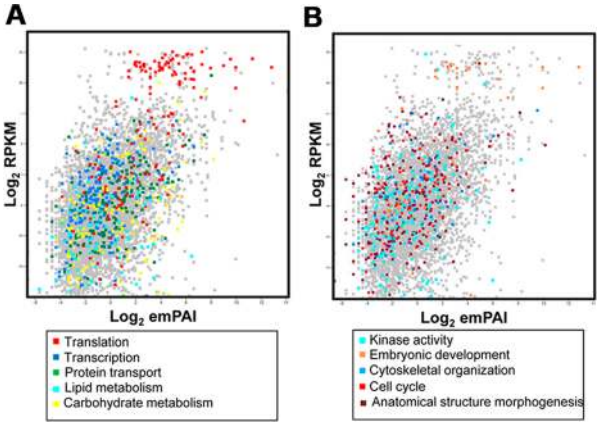


Figure 4. Functional correlation of protein and mRNA levels. (A) Scatter plot of the mRNA and protein abundances across cellular core functions and (B) regulatory and developmental functions based on GO terms. Significantly enriched groups based on GO categorization are shown.

related genes were found at the extreme end of the distribution with highest the correlation (Spearman correlation, 0.683), suggesting that these genes have elevated expression at both the transcript and protein levels (Figure S8 in the Supporting Information). The transcriptional machinery proteins, on the other hand, showed moderate abundance at the protein level but were more elevated on the transcript scale. Those belonging to carbohydrate metabolism, although generally considered to be abundant, spanned over almost the entire distribution with a moderate correlation. The lipid metabolic process-associated genes were frequently of low abundance, and this category was the least correlated. Focusing on the proteins important in development and signal transduction, we found that with the exception of a few proteins at the top end of the distribution that are associated with embryonic development the rest had moderate-to-low protein expression levels (Figure 4B and Figure S9 in the Supporting Information). This suggests that processes relating to cell organization and development are modulated more at the mRNA level than at the protein level at this early stage of development.

Further investigating the cellular compartmentalization across the distribution, we noticed that the ribosomal proteins form one tight cluster at the top end of the distribution and a second additional cluster at moderate expression levels (Figure S10A,B in the Supporting Information). The nucleus, represented with the maximum number of proteins, extended over a large range traversing the entire range of abundance distribution. Although we noticed a slight bias for organelles like mitochondria and endoplasmic reticulum when based on

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

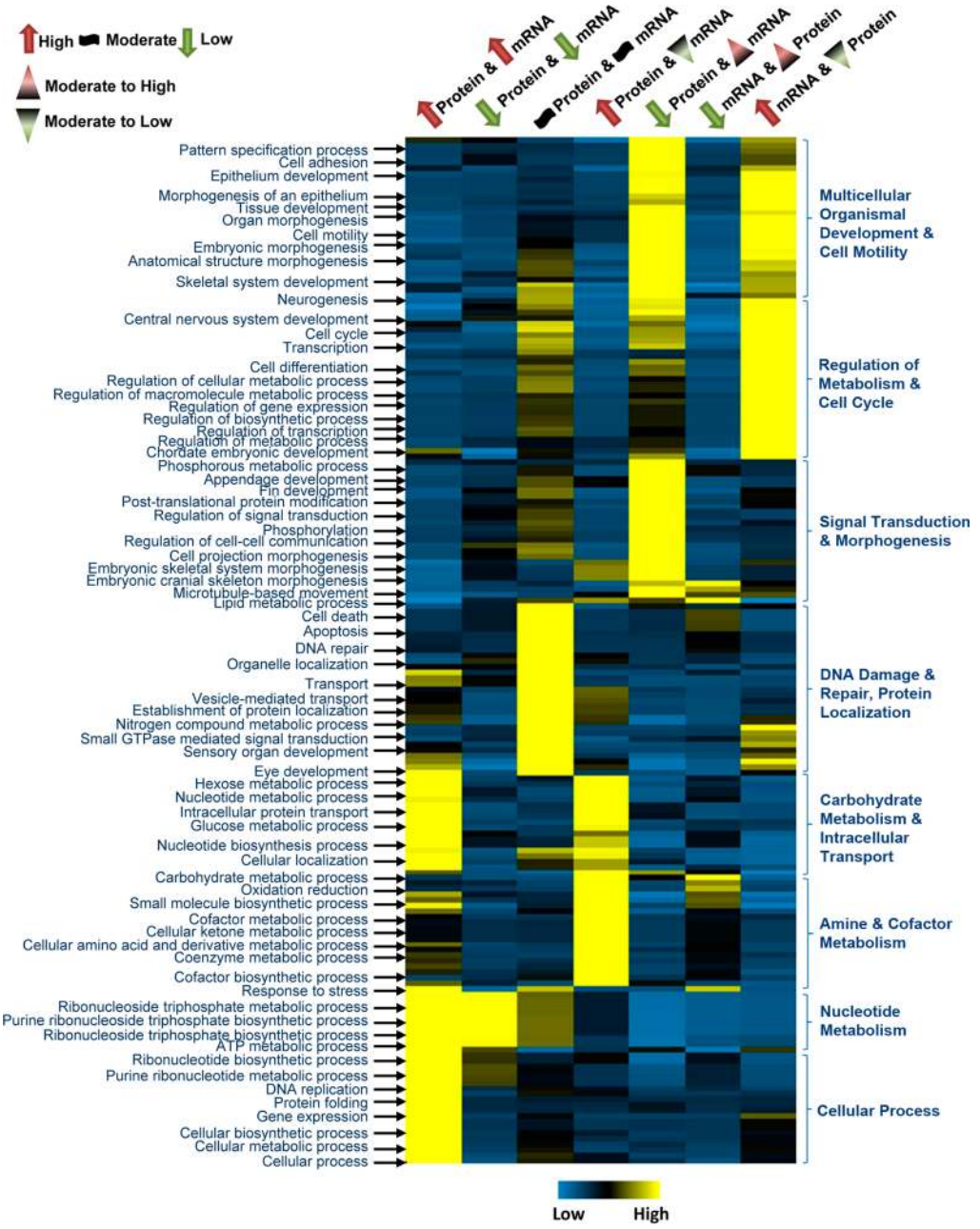


Figure 5. Functional modulation of protein and transcripts in the developing embryo. Proteins and mRNAs were grouped into seven groups based on abundance values as follows: high protein and mRNA, low protein and mRNA, moderate protein and mRNA, high protein and moderate-to-low mRNA, low protein and moderate-to-high mRNA, low mRNA and moderate-to-high, and high mRNA and moderate-to-low protein. The clustered GO biological process terms enriched in at least one of the seven groups are depicted on the heat map. The red arrow corresponds to high abundance of quantified proteins or transcripts, the green arrow, low, and the black symbol, moderate. The shaded red triangle corresponds to moderate-to-high protein or transcript abundances, and the shaded green triangle represents moderate-to-low protein or transcript abundances. High (yellow) and low (blue) in the heat map represent statistical over- or under-representation, respectively.

the quantified proteome, no such distributional bias was observed for the quantified transcriptome.

Distinct Functional Regulation of Proteins and Transcripts

Embryonic development is associated with highly regulated processes that need to be precisely controlled at both the mRNA and protein levels. The analysis above provided a functional overview of the transcriptome and proteome across the entire distribution of cellular abundance. To delineate biological processes that are modulated at different levels of mRNA and protein abundances, we performed a combined hierarchical clustering of the observed transcriptome and proteome. We categorized genes based on their mRNA or protein level into various abundance groups (detailed below) and performed GO enrichment (biological process) by hypergeometric testing on all of the individual groups. Those GO categories that were significantly enriched in at least one of the groups were retained, and one-way hierarchical clustering was performed after normalizing the obtained p values across all of the groups for each category. Such a heat map allowed us to distinctly identify processes that are regulated by different levels

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

of mRNA or protein abundance (Figure 5 and Figure S11 in the Supporting Information).

We first categorized the quantified 5254 genes (including splice variants) into four groups that represented only the upper and lower quantile as follows: high protein and high mRNA, high protein and low mRNA, low protein and high mRNA, and low protein and low mRNA. After performing hierarchical clustering on the four groups, we observed that genes associated with core cellular functions such as energy metabolism, protein transport, and cellular biosynthetic processes are fine-tuned by high levels of both mRNAs and their corresponding proteins (Figure S11 in the Supporting Information). The regulatory (post-translational modification) and developmental process (embryo morphogenesis, nervous system development) are modulated by genes with high mRNA expression and low protein levels. For the remaining two groups, very few GO categories passed the p value threshold, and these often coincided with enriched categories in the high protein and high mRNA and low protein and high mRNA groups.

To gain further functional insights on the entire range of the quantified proteome and transcriptome, we additionally categorized the quantified 5254 genes to include genes from the high-, low-, and moderate-abundance groups. For this purpose, we identified seven groups based on their abundances as follows: high protein and mRNA, low protein and mRNA, moderate protein and mRNA, high protein and moderate-to-low mRNA, low protein and moderate-to-high mRNA, low mRNA and moderate-to-high protein, and high mRNA and moderate-to-low protein (Figure 5). We observed that processes associated with core cellular functions such as metabolism, protein transport, and cellular biosynthetic processes are generally modulated by high protein levels, whereas the mRNA levels show considerable variation. The genes associated with nucleoside metabolic and biosynthetic processes display both high and low levels of mRNA and protein. Interestingly, genes associated with eye development have high mRNA and protein levels, with the optic system being in the mid-to-late phase of its development. Other cellular process including folding and DNA replication are also modulated by high expression at both levels. Functions associated with DNA damage and repair, cell death, protein localization, and sensory organ development exhibit moderate abundance of expressed genes.

Genes that displayed a low protein level but moderate-to-high transcript abundance encoded for those functionally important proteins that are involved in regulatory (post-translational modifications), signal transduction, migratory, and developmental processes. We observed that genes associated with the development of the fin, cartilage, and embryonic skeletal system are highly represented within this group. We note that all of these tissues have not yet initiated their developmental programs and thus these may represent poised conditions. Other processes, such as those associated with the development of the central nervous system, and important morphogenetic events, such as pattern specification process, cell projection organization, and appendage development, are also characterized by differential levels of mRNA and protein.

Genes that displayed a high mRNA level but moderate-to-low protein level were associated with tissue and organ development and regulation of cellular processes. Specifically, processes associated with the regulation of primary metabolic and biosynthetic processes, gene expression, mRNA processing, and cell cycle display significantly high mRNA expression levels and low protein levels. Functions pertaining to microtubule-based movement, lipid metabolic processes, and stress response show enrichment in the low transcript and moderate-to-high protein level group, possibly indicative of long-lived proteins with sentinel roles.

Although proteins are the ultimate biological effectors, we observed that most of the embryonic morphogenetic events and signal transduction processes are modulated by low protein and high mRNA levels, in contrast to the core cellular functions that are modulated by high protein and transcript levels. To assess if any of these low-abundance proteins could be regulated by putative miRNAs, we carried out miRNA target prediction using TargetScanFish 6.2, which predicts gene targets based on conserved sites (7-mer or 8-mer) that match the seed region within a miRNA. It is predicted that many such low-abundance proteins are specifically regulated by miRNA families in the zebrafish embryo (Table S8 in the Supporting Information). This underscores that finer control of protein levels through miRNA regulation or high turnover is constantly in action to ensure proper and coordinated development.

Network Mapping of Quantified Proteome and Transcriptome

Potential interactions among proteins expressed at high and low abundance may modulate important functional processes within the cell. To unveil the functional connectivity among the high- and low-abundance proteins, we constructed protein−protein interaction networks based on curated information from the reactome pathways and GeneGO MetaCore. The reactome functional interaction (reactome FI) combines protein−protein interactions from various organisms alongside curated pathway maps to provide with a high-quality pathway-informed interaction resource, and MetaCore is an expert-curated reliable data resource for protein interactions, primarily focusing on human, rat, and mouse pathways.[46,47] The zebrafish shares many orthologous proteins with those from the mammalian groups and hence these resources may be useful in deriving possible interactions among the quantified zebrafish proteins.

High-Abundance Protein Interactome. A large number of the proteins present in the high-abundance protein groups shared direct interactions with each other and resulted in a closely connected network (653 nodes and 6881 edges) (Figure S12A in the Supporting Information). While a majority of the corresponding transcripts also displayed high to very high abundance, we noticed that a few genes were modulated differently at the protein and mRNA levels and that the rest remained at moderate-abundance levels. Using MCODE, the tightly connected clusters were identified (Figure S12B in the Supporting Information). Of the 10 significant clusters identified with a minimum number of five nodes, the highest-ranking cluster (80 proteins) belonged to Translation. The other clusters were primarily associated with core cellular functions including transcription and RNA transport, oxidative phosphorylation, protein folding, DNA repair, and carbohydrate metabolism. The very low transcript abundance of one of the splicing factor proteins that forms the U2 small nuclear ribonucleoprotein complex (U2snRNP), SF3B3, leads us to speculate that this protein may be highly stable or efficiently translated. RPS6KA1 is the other important protein whose abundance is different at the protein and mRNA levels. The gene encodes for a serine/threonine kinase that is involved in

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

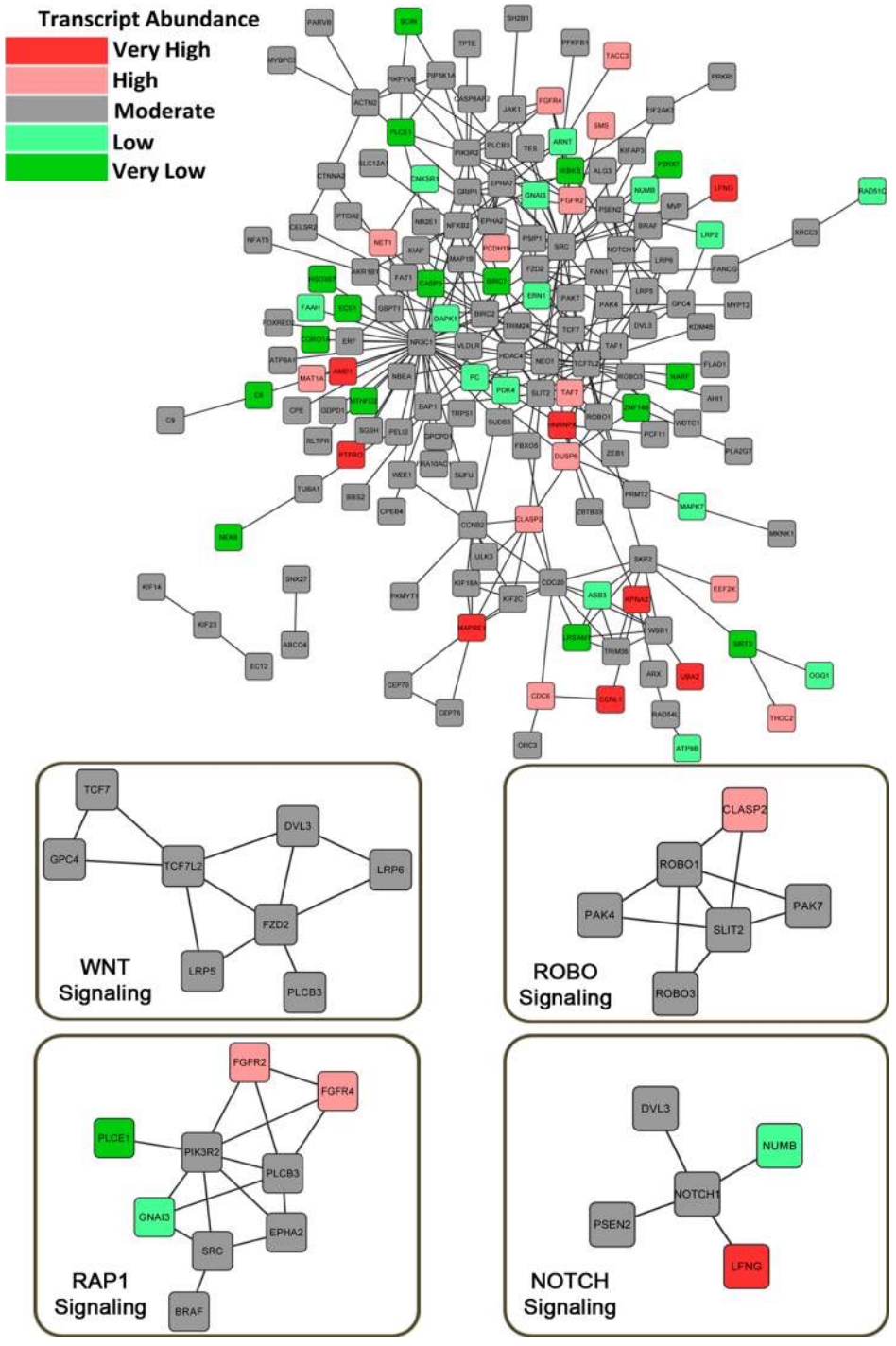


Figure 6. Protein−protein interaction network among low-abundance proteins in the embryo. Direct protein−protein interactions between the low-abundance protein groups identified using reactome FI and MetaCore are represented. The corresponding transcript abundances are indicated by nodes of different colors, with red representing high RPKM values, green representing low RPKM values, and gray representing mRNA expressed at moderate abundances. The significantly enriched signaling pathways are visualized individually.

various signal transduction processes, including MAPK and the nutrient sensing mammalian target of rapamycin (mTOR).[54] Poor correlations between the mRNA and protein abundances may point to more control at the translational or post-translational level. Assessing the degree of variation of the mRNA expression at different time points during development may provide us with clues on the influence of transcriptional or translational control for the poorly correlated genes.

Low-Abundance Protein Interactome. In contrast to the high number of proteins that could interact directly within the higher abundance groups, only 163 low-abundance proteins could be potentially mapped onto a direct interaction network (Figure 6). The clustering coefficient was low (0.156), and the average number of neighbors for each node was only around 3. Hence, dense clusters, as observed in the high-abundance protein network, could not be visualized. The number of noncorrelated genes in terms of protein and mRNA abundance

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

is high in the mapped network. While only about 20% of the genes were expressed at low levels that were on par with the protein abundance, 13% of the genes had high transcript abundance, and the rest had moderate mRNA levels. Protein interactions significantly enriched within the low-abundance protein groups primarily correspond to cell signaling pathways. WNT and Notch signaling, which are important modulators of growth and cell fate decisions in the developing embryo, are highly represented. The Slit-Robo pathway, which is involved in axon guidance and heart tube formation, and the Rap1 signaling pathway, which is associated with cell adhesion and junction formation and is also crucial for heart development in zebrafish, are also significantly enriched.[55,56] The mapped signaling pathways also include genes that are differentially regulated at the transcriptional and translational levels. The observed discordance in expression of several of these genes may possibly point to important regulatory processes at the level of signal transduction and embryonic development

DISCUSSION

In this study, we present a deep proteome map of a selected stage of early embryonic development in zebrafish. We identified >8000 proteins, which is by far the most extensively measured zebrafish embryo proteome. In addition, we also performed transcriptome mining by RNA-seq to provide an informative comparative map of the quantified proteome and transcriptome during embryonic development. Such an integrative approach links the transcript abundance to protein levels and ultimately highlights specific developmental processes that are modulated by changes in the mRNA or protein species. Although similar studies have been performed previously in other organisms,[16,17,57,58] we unveil, for the first time, large-scale combinatorial functional mapping in zebrafish.

Previous transcriptomics and proteomic studies are limited by overall coverage.[52,58] It may be possible that some of the proteins are masked by high-abundance proteins, restricting their detection limits, or that the peptides are not amenable to detection by mass spectrometry. Also, regardless of the transcript levels, some of the mRNAs may not be efficiently translated, resulting in low or undetectable protein levels. We noticed that for a majority of the genes expressed at low levels the corresponding proteins could not be identified. Nevertheless, we could successfully map ∼64% of our quantified proteome to a corresponding mRNA level. Our proteome data identified a paucity of coding sequences on chromosome 4, confirming the analysis of the genome.[6] Moreover, our analysis revealed that there was only a moderate correlation between the protein and transcript levels for most of the cellular processes, consistent with observations in other biological systems.[15,17,58] Although such low correlations may be attributed to technical discrepancies, it is also suggestive of an intricate functional regulatory mechanism that operates to maintain proper levels of transcripts and proteins. There is continuing debate on the concordance of transcript and protein abundances, and the precise mechanisms that act at the post-transcriptional level remain to be elucidated.[57,59,60] In fact, cross-species comparisons suggest that orthologous protein levels correlate better than the corresponding transcript abundances, hinting that the mechanisms to achieve a particular protein abundance evolve rapidly[16] and may include utilization of altered protein stability, translational efficiency, and ribosomal occupancy to achieve the final protein abundance.[61−63] Interestingly, a genome-scale study established that genes that showed minimum variation at the mRNA level through the cell cycle had poor correlation with final protein levels, whereas those that displayed large variation had a high degree of concordance.[64] Analyzing the concerted variation in transcript and protein abundances through the different developmental stages may expose more insights on the orchestration of the transcriptional and translational machineries in zebrafish.

As observed previously in other biological systems, we find that the most abundant proteins are of considerably shorter length and are often associated with central pathways and core processes, including energy and carbohydrate metabolism, translation, and transport.[51,65] For such core metabolic processes, the correlation between the protein and mRNA levels was considerably higher. The genes associated with these processes indeed exhibit a highly conserved coexpression pattern and are also highly correlated at the protein level across species.[16,66] The low-abundance protein groups are primarily regulatory in function and are involved in signal transduction, phosphorylation, and other protein modifications. In spite of the low abundance, higher eukaryotes have a large fraction of the protein mass dedicated to regulatory functions.[67] Furthermore, low-abundance species generally show high sequence variability across species.[16] These proteins, although present in low abundance, are associated with various important processes of development, including cell differentiation and anatomical structure morphogenesis, in our quantified proteome. Of note, it has been shown that regulatory proteins display varied expression levels between different human cell lines.[68] This supports the notion that these regulatory proteins are potent regulators of cell identity and behavior.

While we observed that most of the core metabolic and biosynthetic processes are modulated by proteins that are highly abundant, specific developmental processes are marked by genes showcasing high transcript but low protein levels. It is known that some transcripts are not efficiently translated, are differentially degraded, and/or are stalled during the process of translation. Those cohorts of transcripts, which showed abundant transcripts and low protein levels, might be subjected to the process of stalling during translation. Ribosome profiling experiments identified the existence of differential translational efficiencies of these transcripts.[69,70]

Anatomical developments associated with morphogenesis of embryonic skeletal systems, pattern formation, and neuronal differentiation, for example, are fine-tuned by low protein and high transcript levels. Such systems are yet to initiate full developmental programs at 24 hpf, and we speculate that they may be held in a poised state. As the embryo develops, some of these processes may be primarily modulated by the levels of proteins, transcripts, or both. Network mapping highlighted instances of several genes involved in signaling pathways that had low protein and high mRNA expressions. The low protein abundance may be attributed to reduced stability of the regulatory proteins or may point to post-transcriptional regulation of protein abundance. miRNAs have emerged as important modulators of post-transcriptional regulation, and it is estimated that approximately 30% of the mammalian coding genes are regulated by them.[71] Indeed, most miRNAs in zebrafish are primarily expressed from segmentation stage onward, and some miRNAs have been shown to regulate different processes during development.[72−75] This leads us to speculate that many of the low-abundance (or even undetected) proteins may have been subjected to miRNA

dx.doi.org/10.1021/pr5005136 | J. Proteome Res. XXXX, XXX, XXX−XXX

regulation to allow for the tightly regulated initiation of differentiation. Of note, we observed that many of the critical proteins involved in development, including LFNG,[76] FGFRs,[77] and CTNNB2,[78] are predicted to be regulated by miRNAs (Table S8 in the Supporting Information). Such restriction of mRNA translation underscores the potency of the derived protein products.

From our network mapping, we identified a particular modulator of the Notch pathway, Lunatic Fringe (LFNG), which has low protein and high mRNA levels. The Notch signaling cascade plays a major role in the establishment of the neural crest and binary fate decisions in the neural tube and elsewhere, and it also regulates somitogenesis in developing embryos.[79−81] Formation of the somites (embryonic segments of the vertebrate body) is regulated by oscillatory expression of genes in the segmentation clock that define the spatial pattern, and Notch functions to synchronize the segmentation clock.[82] Such synchronization depends on tight control at the level of mRNA half-life and translational efficiency. Interestingly, LFNG is one such oscillatory gene whose expression is post-transcriptionally regulated by miR-125a-5p (miR-125 was also predicted by TargetScanFish 6.2) for proper somite formation in chick embryo.[83,84] In zebrafish, LFNG is expressed within the presomitic mesoderm and is important for the formation of segment boundaries in the somites and hindbrain.[76,85] Our proteome quantification was performed using embryos at 24 hpf, around which time the boundary between the midbrain and hindbrain forms (22−24 hpf).[86−88] It is tempting to speculate that such regulatory processes may have resulted in the observed differences in the gene and protein expression levels of LFNG. Similar dissection of the network for other protein modules, like FGFR, CLASP2, and DUSP6, may provide functional insights into the differential regulation of the transcriptional and translational machineries during zebrafish embryogenesis.

■ CONCLUSIONS

The in-depth comparative and functional mapping presented here highlights the usefulness of integrative proteomics and transcriptomics to unveil molecular mechanisms regulating early embryogenesis in zebrafish. We particularly highlight differential modulation of various morphogenetic events during embryogenesis. We believe that such an exhaustive approach over the entire time course of development is likely to uncover many novel mechanisms and various levels of gene expression control during embryogenesis and provide a valuable resource for systems biology-based modeling in the future.

■ ASSOCIATED CONTENT

Supporting Information

Table S1: List of all identified peptides and proteins from the zebrafish embryo proteome. The data was retrieved using Trans-Proteomic Pipeline analysis. Table S2: List of identified proteins from the zebrafish embryo proteome. This table enlists the spectral matches, sequence coverage, and peptide counts for each protein. Table S3: Quantification of detected proteins. The table enlists the calculated abundances as emPAI values and information on chromosomal mapping for each protein. Table S4: Assessment of length and pI for all quantified proteins. Table S5: Chromosomal distribution of detected proteins and transcripts. This table enlists the total number of genes annotated for each chromosome and those that could be distinctly mapped from the quantified proteome and transcriptome to individual chromosomes. Table S6: Deep sequencing of transcripts from the zebrafish early embryo. Table S7: Mapping of quantified proteins with corresponding transcripts. This table enlists the 5254 proteins, including instances of splice variants, for which a corresponding transcript was identified using deep sequencing of the zebrafish embryo. Table S8: miRNA prediction for low-abundance proteins. This table enlists the sites and the total context score for each predicted miRNA. Figure S1: Mass error in parts per million (ppm) of precursor ions of all identified peptides. Figure S2: Density distribution of protein abundances as represented by emPAI. Figure S3: Bias analysis of protein length and pI. Figure S4: Distribution of length and pI in the high- and low-abundance protein clusters. Figure S5: Distribution of mapped genes from the proteome and transcriptome across different chromosomes. Figure S6: Coverage and enrichment of mapped genes from the proteome across different chromosomes. Figure S7: Density distribution of transcript abundances as represented by RPKM values. Figure S8: Correlation of transcriptome and proteome for core cellular processes. Figure S9: Correlation of transcriptome and proteome for regulatory and developmental processes. Figure S10: Correlation of protein and mRNA levels for different subcellular compartments. Figure S11: Functional modulation of proteins and transcripts in the early stage embryo. Figure S12: Protein−protein interaction network among the high-abundance proteins. This material is available free of charge via the Internet at http://pubs.acs.org. The protein data set information was uploaded to the PeptideAtlas database and is available at ftp://PASS00444:HP4768xss@ftp.peptideatlas.org/.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (65) 6586 9689. Fax: (65) 6779 1117. E-mail: jayanthag@imcb.a-star.edu.sg.

Author Contributions

⊥A.A.S. and S.W. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Ler Siok Ghee for assistance with mass spectrometry and Hyung Won Choi for critically reading the manuscript. The work was supported by the Agency for Science, Technology and Research (A*STAR), Singapore. Animal experiments were conducted according to guidelines of the Agri-Food and Veterinary Authority of Singapore.

■ ABBREVIATIONS

AA, amino acids; BP, biological process; CC, cellular component; DAVID, Database for Annotation, Visualization and Integrated Discovery; emPAI, exponentially modified protein abundance index; ESI, electrospray ionization; FDR, false discovery rate; GO, gene ontology; hpf, hours post fertilization; HPLC, high-performance liquid chromatography; IEF, isoelectric focusing; IPI, international protein index; LC−MS/MS, liquid chromatography−tandem mass spectrometry; LFNG, lunatic fringe; LTQ, linear trap quadrupole; MF, molecular function; miRNA, microRNA; mRNA, messenger RNA; mTOR, mammalian target of rapamycin; PPM, parts per

million; RPKM, reads per kilobase of exon per million reads mapped; SIF, simple interaction file; TPP, Trans-Proteomic Pipeline; UTR, untranslated region; ZV9, zebrafish genome assembly

REFERENCES

(1) Driever, W.; Stemple, D.; Schier, A.; Solnica-Krezel, L. Zebrafish: genetic tools for studying vertebrate development. *Trends Genet.* **1994**, *10*, 152−9.

(2) Zon, L. I. Zebrafish: a new model for human disease. *Genome Res.* **1999**, *9*, 99−100.

(3) Ward, A. C.; Lieschke, G. J. The zebrafish as a model system for human disease. *Front. Biosci.* **2002**, *7*, d827−33.

(4) Lieschke, G. J.; Currie, P. D. Animal models of human disease: zebrafish swim into view. *Nat. Rev. Genet.* **2007**, *8*, 353−67.

(5) Stemple, D. L.; Driever, W. Zebrafish: tools for investigating cellular differentiation. *Curr. Opin. Cell Biol.* **1996**, *8*, 858−64.

(6) Howe, K.; Clark, M. D.; Torroja, C. F.; Torrance, J.; Berthelot, C.; Muffato, M.; Collins, J. E.; Humphray, S.; McLaren, K.; Matthews, L.; McLaren, S.; Sealy, I.; Caccamo, M.; Churcher, C.; Scott, C.; Barrett, J. C.; Koch, R.; Rauch, G. J.; White, S.; Chow, W.; Kilian, B.; Quintais, L. T.; Guerra-Assuncao, J. A.; Zhou, Y.; Gu, Y.; Yen, J.; Vogel, J. H.; Eyre, T.; Redmond, S.; Banerjee, R.; Chi, J.; Fu, B.; Langley, E.; Maguire, S. F.; Laird, G. K.; Lloyd, D.; Kenyon, E.; Donaldson, S.; Sehra, H.; Almeida-King, J.; Loveland, J.; Trevanion, S.; Jones, M.; Quail, M.; Willey, D.; Hunt, A.; Burton, J.; Sims, S.; McLay, K.; Plumb, B.; Davis, J.; Clee, C.; Oliver, K.; Clark, R.; Riddle, C.; Elliot, D.; Threadgold, G.; Harden, G.; Ware, D.; Mortimore, B.; Kerry, G.; Heath, P.; Phillimore, B.; Tracey, A.; Corby, N.; Dunn, M.; Johnson, C.; Wood, J.; Clark, S.; Pelan, S.; Griffiths, G.; Smith, M.; Glithero, R.; Howden, P.; Barker, N.; Stevens, C.; Harley, J.; Holt, K.; Panagiotidis, G.; Lovell, J.; Beasley, H.; Henderson, C.; Gordon, D.; Auger, K.; Wright, D.; Collins, J.; Raisen, C.; Dyer, L.; Leung, K.; Robertson, L.; Ambridge, K.; Leongamornlert, D.; McGuire, S.; Gilderthorp, R.; Griffiths, C.; Manthravadi, D.; Nichol, S.; Barker, G.; Whitehead, S.; Kay, M.; Brown, J.; Murnane, C.; Gray, E.; Humphries, M.; Sycamore, N.; Barker, D.; Saunders, D.; Wallis, J.; Babbage, A.; Hammond, S.; Mashreghi-Mohammadi, M.; Barr, L.; Martin, S.; Wray, P.; Ellington, A.; Matthews, N.; Ellwood, M.; Woodmansey, R.; Clark, G.; Cooper, J.; Tromans, A.; Grafham, D.; Skuce, C.; Pandian, R.; Andrews, R.; Harrison, E.; Kimberley, A.; Garnett, J.; Fosker, N.; Hall, R.; Garner, P.; Kelly, D.; Bird, C.; Palmer, S.; Gehring, I.; Berger, A.; Dooley, C. M.; Ersan-Urun, Z.; Eser, C.; Geiger, H.; Geisler, M.; Karotki, L.; Kirn, A.; Konantz, J.; Konantz, M.; Oberlander, M.; Rudolph-Geiger, S.; Teucke, M.; Osoegawa, K.; Zhu, B.; Rapp, A.; Widaa, S.; Langford, C.; Yang, F.; Carter, N. P.; Harrow, J.; Ning, Z.; Herrero, J.; Searle, S. M.; Enright, A.; Geisler, R.; Plasterk, R. H.; Lee, C.; Westerfield, M.; de Jong, P. J.; Zon, L. I.; Postlethwait, J. H.; Nusslein-Volhard, C.; Hubbard, T. J.; Roest Crollius, H.; Rogers, J.; Stemple, D. L.; Begum, S.; Lloyd, C.; Lanz, C.; Raddatz, G.; Schuster, S. C. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **2013**, *496*, 498−503.

(7) Collins, J. E.; White, S.; Searle, S. M.; Stemple, D. L. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res.* **2012**, *22*, 2067−78.

(8) Vesterlund, L.; Jiao, H.; Unneberg, P.; Hovatta, O.; Kere, J. The zebrafish transcriptome during early development. *BMC Dev. Biol.* **2011**, *11*, 30.

(9) Mathavan, S.; Lee, S. G.; Mak, A.; Miller, L. D.; Murthy, K. R.; Govindarajan, K. R.; Tong, Y.; Wu, Y. L.; Lam, S. H.; Yang, H.; Ruan, Y.; Korzh, V.; Gong, Z.; Liu, E. T.; Lufkin, T. Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet.* **2005**, *1*, 260−76.

(10) Lucitt, M. B.; Price, T. S.; Pizarro, A.; Wu, W.; Yocum, A. K.; Seiler, C.; Pack, M. A.; Blair, I. A.; Fitzgerald, G. A.; Grosser, T. Analysis of the zebrafish proteome during embryonic development. *Mol. Cell. Proteomics* **2008**, *7*, 981−94.

(11) Lin, Y.; Chen, Y.; Yang, X.; Xu, D.; Liang, S. Proteome analysis of a single zebrafish embryo using three different digestion strategies coupled with liquid chromatography−tandem mass spectrometry. *Anal. Biochem.* **2009**, *394*, 177−85.

(12) Pauli, A.; Valen, E.; Lin, M. F.; Garber, M.; Vastenhouw, N. L.; Levin, J. Z.; Fan, L.; Sandelin, A.; Rinn, J. L.; Regev, A.; Schier, A. F. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **2012**, *22*, 577−91.

(13) Lossner, C.; Wee, S.; Ler, S. G.; Li, R. H.; Carney, T.; Blackstock, W.; Gunaratne, J. Expanding the zebrafish embryo proteome using multiple fractionation approaches and tandem mass spectrometry. *Proteomics* **2012**, *12*, 1879−82.

(14) Singh, S. K.; Rakesh, K. S.; Ramamoorthy, K.; Saradhi, A. V. P.; Idris, M. M. Proteome Profile of Zebrafish Brain Based on Gel LC−ESI MS/MS Analysis. *J. Proteomics Bioinf.* **2010**, *3*, 135−142.

(15) Gunaratne, J.; Schmidt, A.; Quandt, A.; Neo, S. P.; Sarac, O. S.; Gracia, T.; Loguercio, S.; Ahrne, E.; Xia, R. L.; Tan, K. H.; Lossner, C.; Bahler, J.; Beyer, A.; Blackstock, W.; Aebersold, R. Extensive mass spectrometry-based analysis of the fission yeast proteome: the Schizosaccharomyces pombe PeptideAtlas. *Mol. Cell. Proteomics* **2013**, *12*, 1741−51.

(16) Schrimpf, S. P.; Weiss, M.; Reiter, L.; Ahrens, C. H.; Jovanovic, M.; Malmstrom, J.; Brunner, E.; Mohanty, S.; Lercher, M. J.; Hunziker, P. E.; Aebersold, R.; von Mering, C.; Hengartner, M. O. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* **2009**, *7*, e48.

(17) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Paabo, S.; Mann, M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **2011**, *7*, 548.

(18) Groh, K. J.; Nesatyy, V. J.; Segner, H.; Eggen, R. I.; Suter, M. J. Global proteomics analysis of testis and ovary in adult zebrafish (*Danio rerio*). *Fish Physiol. Biochem.* **2011**, *37*, 619−47.

(19) Palmblad, M.; Henkel, C. V.; Dirks, R. P.; Meijer, A. H.; Deelder, A. M.; Spaink, H. P. Parallel deep transcriptome and proteome analysis of zebrafish larvae. *BMC Res. Notes* **2013**, *6*, 428.

(20) Gebriel, M.; Prabhudesai, S.; Uleberg, K. E.; Larssen, E.; Piston, D.; Bjornstad, A. H.; Møller, S. G. Zebrafish brain proteomics reveals central proteins involved in neurodegeneration. *J. Neurosci. Res.* **2014**, *92*, 104−15.

(21) Singh, S. K.; Saxena, S.; Meena Lakshmi, M. G.; Saxena, P.; Idris, M. M. Proteome profile of zebrafish *Danio rerio* olfactory bulb based on two-dimensional gel electrophoresis matrix-assisted laser desorption/ionization MS/MS analysis. *Zebrafish* **2011**, *8*, 183−9.

(22) Kessels, M. Y.; Huitema, L. F.; Boeren, S.; Kranenbarg, S.; Schulte-Merker, S.; van Leeuwen, J. L.; de Vries, S. C. Proteomics analysis of the zebrafish skeletal extracellular matrix. *PLoS One* **2014**, *9*, e90568.

(23) Posner, M.; Hawke, M.; Lacava, C.; Prince, C. J.; Bellanco, N. R.; Corbin, R. W. A proteome map of the zebrafish (*Danio rerio*) lens reveals similarities between zebrafish and mammalian Crystallin expression. *Mol. Vision* **2008**, *14*, 806−14.

(24) De Souza, A. G.; MacCormack, T. J.; Wang, N.; Li, L.; Goss, G. G. Large-scale proteome profile of the zebrafish (*Danio rerio*) gill for physiological and biomarker discovery studies. *Zebrafish* **2009**, *6*, 229−38.

(25) Wang, M.; Chan, L. L.; Si, M.; Hong, H.; Wang, D. Proteomic analysis of hepatic tissue of zebrafish (*Danio rerio*) experimentally exposed to chronic microcystin-LR. *Toxicol. Sci.* **2010**, *113*, 60−9.

(26) Hanisch, K.; Kuster, E.; Altenburger, R.; Gundel, U. Proteomic signatures of the zebrafish (*Danio rerio*) embryo: sensitivity and specificity in toxicity assessment of chemicals. *Int. J. Proteomics* **2010**, *2010*, 630134.

(27) Gundel, U.; Benndorf, D.; von Bergen, M.; Altenburger, R.; Kuster, E. Vitellogenin cleavage products as indicators for toxic stress in zebra fish embryos: a proteomic approach. *Proteomics* **2007**, *7*, 4541−54.

(28) Hung, M. W.; Zhang, Z. J.; Li, S.; Lei, B.; Yuan, S.; Cui, G. Z.; Man Hoi, P.; Chan, K.; Lee, S. M. From omics to drug metabolism and high content screen of natural product in zebrafish: a new model for

discovery of neuroactive compound. *J. Evidence-Based Complementary Altern. Med.* **2012**, *2012*, 605303.
(29) Kuznetsova, G. P.; Larina, O. V.; Petushkova, N. A.; Kisrieva, Y. S.; Samenkova, N. F.; Trifonova, O. P.; Karuzina, I. I.; Ipatova, O. M.; Zolotaryov, K. V.; Romashova, Y. A.; Lisitsa, A. V. Effects of fullerene C60 on proteomic profile of Danio rerio fish embryos. *Bull. Exp. Biol. Med.* **2014**, *156*, 694−8.
(30) Konzer, A.; Ruhs, A.; Braun, H.; Jungblut, B.; Braun, T.; Kruger, M. Stable isotope labeling in zebrafish allows in vivo monitoring of cardiac morphogenesis. *Mol. Cell. Proteomics* **2013**, *12*, 1502−12.
(31) Nolte, H.; Konzer, A.; Ruhs, A.; Jungblut, B.; Braun, T.; Kruger, M. Global protein expression profiling of zebrafish organs based on in vivo incorporation of stable isotopes. *J. Proteome Res.* **2014**, *13*, 2162−74.
(32) Thomas, R. J. Yolk distribution and utilization during early development of a teleost embryo (*Brachydanio rerio*). *J. Embryol. Exp. Morphol.* **1968**, *19*, 203−15.
(33) Link, V.; Shevchenko, A.; Heisenberg, C. P. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **2006**, *6*, 1.
(34) Kimmel, C. B.; Ballard, W. W.; Kimmel, S. R.; Ullmann, B.; Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **1995**, *203*, 253−310.
(35) Link, V.; Shevchenko, A.; Heisenberg, C.-P. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **2006**, *6*, 1.
(36) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2007**, *1*, 2856−2860.
(37) Hubner, N. C.; Ren, S.; Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *PROTEOMICS* **2008**, *8*, 4862−4872.
(38) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2*, 1896−1906.
(39) Pedrioli, P. G. Trans-proteomic pipeline: a pipeline for proteomic analysis. *Methods Mol. Biol.* **2010**, *604*, 213−38.
(40) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **2011**, *10*, M111 007690.
(41) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4*, 1265−72.
(42) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44−57.
(43) Maere, S.; Heymans, K.; Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448−9.
(44) Lewis, B. P.; Burge, C. B.; Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **2005**, *120*, 15−20.
(45) Grimson, A.; Farh, K. K.; Johnston, W. K.; Garrett-Engele, P.; Lim, L. P.; Bartel, D. P. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **2007**, *27*, 91−105.
(46) Wu, G.; Feng, X.; Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **2010**, *11*, R53.
(47) Wu, G.; Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome biology* **2012**, *13*, R112.
(48) Ostlund, G.; Schmitt, T.; Forslund, K.; Kostler, T.; Messina, D. N.; Roopra, S.; Frings, O.; Sonnhammer, E. L. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **2010**, *38*, D196−203.
(49) Bader, G. D.; Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* **2003**, *4*, 2.
(50) Coghlan, A.; Wolfe, K. H. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast* **2000**, *16*, 1131−45.
(51) Frith, M. C.; Forrest, A. R.; Nourbakhsh, E.; Pang, K. C.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Bailey, T. L.; Grimmond, S. M. The abundance of short proteins in the mammalian proteome. *PLoS Genet.* **2006**, *2*, e52.
(52) Tian, Q.; Stepaniants, S. B.; Mao, M.; Weng, L.; Feetham, M. C.; Doyle, M. J.; Yi, E. C.; Dai, H.; Thorsson, V.; Eng, J.; Goodlett, D.; Berger, J. P.; Gunter, B.; Linseley, P. S.; Stoughton, R. B.; Aebersold, R.; Collins, S. J.; Hanlon, W. A.; Hood, L. E. Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol. Cell. Proteomics* **2004**, *3*, 960−9.
(53) Clarke, C.; Henry, M.; Doolan, P.; Kelly, S.; Aherne, S.; Sanchez, N.; Kelly, P.; Kinsella, P.; Breen, L.; Madden, S. F.; Zhang, L.; Leonard, M.; Clynes, M.; Meleady, P.; Barron, N. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* **2012**, *13*, 656.
(54) Anjum, R.; Blenis, J. The RSK family of kinases: emerging roles in cellular signalling. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 747−58.
(55) Fish, J. E.; Wythe, J. D.; Xiao, T.; Bruneau, B. G.; Stainier, D. Y.; Srivastava, D.; Woo, S. A Slit/miR-218/Robo regulatory loop is required during heart tube formation in zebrafish. *Development* **2011**, *138*, 1409−19.
(56) Dong, W.; Yang, Z.; Yang, F.; Wang, J.; Zhuang, Y.; Xu, C.; Zhang, B.; Tian, X. L.; Liu, D. Suppression of Rap1 impairs cardiac myofibrils and conduction system in zebrafish. *PloS one* **2012**, *7*, e50960.
(57) Juschke, C.; Dohnal, I.; Pichler, P.; Harzer, H.; Swart, R.; Ammerer, G.; Mechtler, K.; Knoblich, J. A. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome biology* **2013**, *14*, r133.
(58) Ghazalpour, A.; Bennett, B.; Petyuk, V. A.; Orozco, L.; Hagopian, R.; Mungrue, I. N.; Farber, C. R.; Sinsheimer, J.; Kang, H. M.; Furlotte, N.; Park, C. C.; Wen, P. Z.; Brewer, H.; Weitz, K.; Camp, D. G., II; Pan, C.; Yordanova, R.; Neuhaus, I.; Tilford, C.; Siemers, N.; Gargalovic, P.; Eskin, E.; Kirchgessner, T.; Smith, D. J.; Smith, R. D.; Lusis, A. J. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS genetics* **2011**, *7*, e1001393.
(59) Olivares-Hernández, R.; Usaite, R.; Nielsen, J. Integrative analysis using proteome and transcriptome data from yeast to unravel regulatory patterns at post-transcriptional level. *Biotechnol. Bioeng.* **2010**, *107*, 865−875.
(60) Robles, M. S.; Cox, J.; Mann, M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet.* **2014**, *10*, e1004047.
(61) Arava, Y.; Wang, Y.; Storey, J. D.; Liu, C. L.; Brown, P. O.; Herschlag, D. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3889−94.
(62) Pratt, J. M.; Petty, J.; Riba-Garcia, I.; Robertson, D. H.; Gaskell, S. J.; Oliver, S. G.; Beynon, R. J. Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 579−91.
(63) Tuller, T.; Waldman, Y. Y.; Kupiec, M.; Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 3645−50.
(64) Greenbaum, D.; Colangelo, C.; Williams, K.; Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **2003**, *4*, 117.
(65) Warringer, J.; Blomberg, A. Evolutionary constraints on yeast protein size. *BMC Evol. Biol.* **2006**, *6*, 61.

(66) Bergmann, S.; Ihmels, J.; Barkai, N. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol.* **2004**, *2*, E9.
(67) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **2011**, *7*, 549.
(68) Lundberg, E.; Fagerberg, L.; Klevebring, D.; Matic, I.; Geiger, T.; Cox, J.; Algenas, C.; Lundeberg, J.; Mann, M.; Uhlen, M. Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **2010**, *6*, 450.
(69) Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev. Genet.* **2014**, *15*, 205–13.
(70) Bhushan, S.; Meyer, H.; Starosta, A. L.; Becker, T.; Mielke, T.; Berninghausen, O.; Sattler, M.; Wilson, D. N.; Beckmann, R. Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol. Cell* **2010**, *40*, 138–46.
(71) Lewis, B. P.; Shih, I. H.; Jones-Rhoades, M. W.; Bartel, D. P.; Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **2003**, *115*, 787–98.
(72) Wienholds, E.; Kloosterman, W. P.; Miska, E.; Alvarez-Saavedra, E.; Berezikov, E.; de Bruijn, E.; Horvitz, H. R.; Kauppinen, S.; Plasterk, R. H. MicroRNA expression in zebrafish embryonic development. *Science* **2005**, *309*, 310–1.
(73) Ketley, A.; Warren, A.; Holmes, E.; Gering, M.; Aboobaker, A. A.; Brook, J. D. The miR-30 microRNA family targets smoothened to regulate hedgehog signalling in zebrafish early muscle development. *PLoS One* **2013**, *8*, e65170.
(74) Stahlhut, C.; Suarez, Y.; Lu, J.; Mishima, Y.; Giraldez, A. J. miR-1 and miR-206 regulate angiogenesis by modulating VegfA expression in zebrafish. *Development* **2012**, *139*, 4356–64.
(75) Bazzini, A. A.; Lee, M. T.; Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **2012**, *336*, 233–7.
(76) Prince, V. E.; Holley, S. A.; Bally-Cuif, L.; Prabhakaran, B.; Oates, A. C.; Ho, R. K.; Vogt, T. F. Zebrafish lunatic fringe demarcates segmental boundaries. *Mech. Dev.* **2001**, *105*, 175–80.
(77) Fürthauer, M.; Van Celst, J.; Thisse, C.; Thisse, B. Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development* **2004**, *131*, 2853–2864.
(78) Bellipanni, G.; Varga, M.; Maegawa, S.; Imai, Y.; Kelly, C.; Myers, A. P.; Chu, F.; Talbot, W. S.; Weinberg, E. S. Essential and opposing roles of zebrafish beta-catenins in the formation of dorsal axial structures and neurectoderm. *Development* **2006**, *133*, 1299–309.
(79) High, F. A.; Zhang, M.; Proweller, A.; Tu, L.; Parmacek, M. S.; Pear, W. S.; Epstein, J. A. An essential role for Notch in neural crest during cardiovascular development and smooth muscle differentiation. *J. Clin. Invest.* **2007**, *117*, 353–63.
(80) Ferjentsik, Z.; Hayashi, S.; Dale, J. K.; Bessho, Y.; Herreman, A.; De Strooper, B.; del Monte, G.; de la Pompa, J. L.; Maroto, M. Notch is a critical component of the mouse somitogenesis oscillator and is essential for the formation of the somites. *PLoS Genet.* **2009**, *5*, e1000662.
(81) Weinmaster, G.; Kintner, C. Modulation of notch signaling during somitogenesis. *Ann. Rev. Cell Dev. Biol.* **2003**, *19*, 367–95.
(82) Terry, A. J.; Sturrock, M.; Dale, J. K.; Maroto, M.; Chaplain, M. A. J. A spatio-temporal model of Notch signalling in the zebrafish segmentation clock: conditions for synchronised oscillatory dynamics. *PLoS One* **2011**, *6*, e16980.
(83) Serth, K.; Schuster-Gossler, K.; Cordes, R.; Gossler, A. Transcriptional oscillation of lunatic fringe is essential for somitogenesis. *Genes Dev.* **2003**, *17*, 912–25.
(84) Riley, M. F.; Bochter, M. S.; Wahi, K.; Nuovo, G. J.; Cole, S. E. Mir-125a-5p-mediated regulation of Lfng is essential for the avian segmentation clock. *Dev. Cell* **2013**, *24*, 554–61.
(85) Sato, Y.; Yasuda, K.; Takahashi, Y. Morphological boundary forms by a novel inductive event mediated by Lunatic fringe and Notch during somitic segmentation. *Development* **2002**, *129*, 3633–44.
(86) Graeden, E.; Sive, H. Live imaging of the zebrafish embryonic brain by confocal microscopy. *J. Visualized Exp.* **2009**, e1217.
(87) Gutzman, J. H.; Graeden, E. G.; Lowery, L. A.; Holley, H. S.; Sive, H. Formation of the zebrafish midbrain-hindbrain boundary constriction requires laminin-dependent basal constriction. *Mech. Dev.* **2008**, *125*, 974–83.
(88) Lowery, L. A.; Sive, H. Initial formation of zebrafish brain ventricles occurs independently of circulation and requires the nagie oko and snakehead/atp1a1a.1 gene products. *Development* **2005**, *132*, 2057–67.