Genome **Biology**

Constructing a fish metabolic network model

Shuzhao Li[1,2*], Alexander Pozhitkov[1,3], Rachel A Ryan[1], Charles S Manning[1], Nancy Brown-Peterson[1], Marius Brouwer[1]

Abstract

We report the construction of a genome-wide fish metabolic network model, MetaFishNet, and its application to analyzing high throughput gene expression data. This model is a stepping stone to broader applications of fish systems biology, for example by guiding study design through comparison with human metabolism and the integration of multiple data types. MetaFishNet resources, including a pathway enrichment analysis tool, are accessible at http://metafishnet.appspot.com.

Rationale

Small fish species are widely used in ecological and pharmaceutical toxicology, developmental biology and genetics, evolutionary biology and as human disease models. Among the species commonly found in scientific literature are zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), European flounder (*Platichthys flesus*), channel catfish (*Ictalurus punctatus*), sheepshead minnow (*Cyprinodon variegatus*), mummichog (*Fundulus heteroclitus*), Atlantic salmon (*Salmo salar*), common carp (*Cyprinus carpio*), rainbow trout (*Oncorhynchus mykiss*) and swordtail (*Xiphophorus hellerii*). Each of these fish species has its own niche as a research tool. For example, *Xiphophorus* is a classic genetic model of melanomas [1,2], whereas medaka is a good model for reproductive and ecotoxicological studies [3]. Zebrafish, in particular, has risen to stardom in recent years, with a large collection of mutants and established techniques for transgenesis, expression studies, forward and reverse genetics and *in vivo* imaging [4-8]. The use of zebrafish as human disease models has also spiked significant interests [9-11]. Since small fish are currently the only vertebrate species that can be studied in high throughput, their future in modern biomedical sciences is brighter than ever [12,13].

Fish genomics is also taking off. Thus far, whole genome sequences are available for five fish species:

D. rerio, O. latipes, T. rubripes, T. nigroviridis and *G. aculeatus*. DNA microarrays have been applied to study gene expression in many more fish species [14-18]. However, fish functional genomics is far behind other model organisms. In the example of sheepshead minnows, which are used in our lab for ecotoxicology, gene annotation is poor and no pathway analysis tool is readily available for interpreting DNA microarray data. The situation is similar for other fish species, with zebrafish perhaps an arguable exception. Bioinformatic tools that fill in this gap in fish functional genomics are highly desirable [17]. Oberhardt *et al.* [19] summarized the five applications of genome-wide metabolic network models: '(1) contextualization of high-throughput data, (2) guidance of metabolic engineering, (3) directing hypothesis-driven discovery, (4) interrogation of multi-species relationships, and (5) network property discovery.' While significant interest exists for a fish metabolic network model in all five categories, the immediate and primary application of our model will be the interpretation of high throughput expression data, especially pathway analysis, which can be done either by direct mapping to metabolic genes [20,21] or via established enrichment statistics [22,23]. This model will also provide a first glance of how fish metabolism resembles human metabolism, which should be instructional for the use of fish in many research areas [24]. This proposed first generation model will serve as a reference and stepping stone to further systems investigations, helping study design and hypotheses generation. As more data become available in the future, the model can be further refined to support broader applications.

The recent completion of genome sequencing of five fish species has paved the way for constructing a genome-wide fish metabolic network model. That is, all

* Correspondence: shuzhao.li@gmail.com
[1]Gulf Coast Research Laboratory, Department of Coastal Sciences, University of Southern Mississippi, 703 East Beach Drive, Ocean Springs, MS 39564, USA
Full list of author information is available at the end of the article

metabolic enzymes can be identified from complete genomes by sequence analysis, compounds can then be associated with enzymatic activities and a metabolic network can be constructed by linking these compounds and enzymes. This type of *ab initio* construction of metabolic networks has been carried out for many unicellular organisms [19,25-30].

However, *ab initio* construction alone is not yet feasible for vertebrate metabolic networks due to their complexity. Two high-quality human metabolic network models [20,31] have been published recently. Both studies included intensive human curation and comprehensive supporting evidence, including data from model species other than human. Thus, these two 'human' models can provide critical references for constructing a genome-wide fish metabolic network model, to help overcome the limitation of *ab initio* construction. Combining the integration of existing models and *ab initio* construction from whole genomes has been the strategy for our project. A metabolic model for zebrafish exists in the KEGG database [32].

However, our genome-wide model offers a significant expansion of the KEGG zebrafish model.

We will first report the construction process of this fish metabolic network model (MetaFishNet). We then use MetaFishNet to methodically comparefish and human metabolism to identify the most and least conserved pathways. The last sections of this paper will demonstrate the application of MetaFishNet in analyzing two sets of DNA microarray data: one from zebrafish as liver cancer model in public repository, the other from sheepshead minnow exposed to cadmium in our lab.

Results and discussion
Construction of MetaFishNet
Our genome-wide fish metabolic network, MetaFishNet, adopts a conventional bipartite network structure, where enzymes and compounds are two types of nodes. The construction strategy of MetaFishNet is shown in Figure 1. Details are given in the 'Method' section and Additional file 1, while a short description follows here.

We first analyzed all cDNA sequences from five fish genomes (*D. rerio, O. latipes, T. rubripes, T. nigroviridis* and *G. aculeatus*) to create a list of all fish metabolic genes via gene ontology. From this metabolic gene list, the corresponding enzymes were identified using either orthologous relationships to human genes or similarity to consensus enzyme sequences (Table 1). Two types of metabolic reactions are included in MetaFishNet. The majority consists of reactions in reference models that can be associated with fish enzymes. The rest of the reactions were created according to relationships between inferred enzymatic activity and compounds. The reference reactions in this project are data
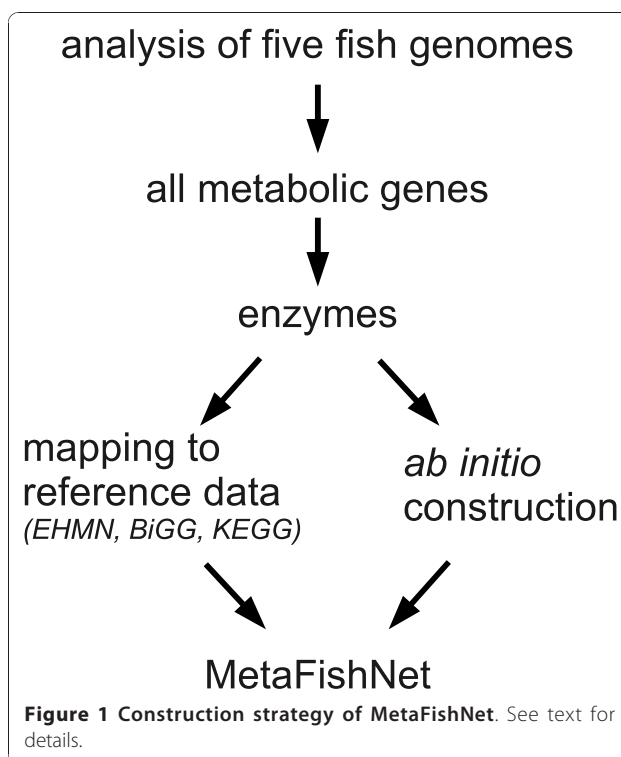

Figure 1 Construction strategy of MetaFishNet. See text for details.

integrated from Edinburgh Human Metabolic Network (EHMN) [31], the human metabolic network from Palsson's group at UCSD (BiGG) [20] and the zebrafish metabolic network from KEGG. Finally, the whole network is formed by linking all reactions.

To illustrate the construction process, let us consider two pieces of sequences from the medaka genome.

Sequence ENSORLG00000001750 is mapped to a human homolog PIK3CG, which is a phosphoinositide-3-kinase (enzyme commission number 2.7.1.153). This enzyme is associated to a reaction in the EHMN model that converts 1-Phosphatidyl-D-myo-inositol 4,5-bisphosphate to Phosphatidylinositol-3,4,5-trisphosphate. Thus, this same reaction is carried over to the MetaFishNet model. Another sequence ENSORLG00000018911 also has a human homolog, PIP4K2B, which is a phosphatidylinositol-5-phosphate 4-kinase with enzyme commission number 2.7.1.149. Although no reaction for this enzyme is found for any of the reference models, we learn from the KEGG LIGAND database that this enzyme converts 1-Phosphatidyl-

Table 1 Metabolic Enzymes found in five fish genomes

Species	Number of metabolic genes	Number of ECs
Zebrafish	3,853	654
Medaka	3,998	765
Takifugu	4,103	771
Tetraodon	4,424	782
Stickleback	4,324	791

1D-myo-inositol 5-phosphate to 1-Phosphatidyl-D-myo-inositol 4,5-bisphosphate. This reaction is added to MetaFishNet as an inferred reaction. Furthermore, because the second reaction produces the substrate for the first reaction, the two reactions are linked together in the 'Phosphatidylinositol phosphate metabolism' pathway.

We carefully reconciled the pathway organization during integration of the three reference models by comparing the reactions in each pathway. Thus, the pathway organization in MetaFishNet follows biochemical conventions wherever possible. Yet, over 600 reactions still do not map directly to these reference pathways. Since pathways can be viewed as modules within a metabolic network [33], we extracted network modules from these reactions using a modularity algorithm [34]. The resulting modules were manually inspected to either become a new pathway, to merge with an existing pathway, or to be invalidated. Meanwhile, individual reactions were attached to a pathway when they connect metabolites in that pathway. This combined procedure of module finding and manual curation was repeated iteratively until no further change could be made.

Even though this model contains data specific to each of the five fish species, we choose to present a combined fish metabolic network model because a) a combined model will be more useful for other under-represented fish species; b) genome annotations are far from perfect - combining five genome sequences will reduce the chance of missing true metabolic genes. For example in the TCA cycle, we did not find ATP citrate synthase in the zebrafish genome, nor succinate-CoA ligase in the *Tetraodon* genome (Ensembl 51). Since these are critical enzymes in a central pathway, these missing enzymes reflect annotation errors. The combined model is thus more comprehensive than using any single species alone (Additional file 2). In total, 911 enzymes, 3,342 reactions and 115 pathways are included in MetaFishNet version 1.9.6. Data integration at the reaction level is shown in Figure 2. All MetaFishNet pathways are given in Additional file 3, reaction data in Additional file 4 and SBML (Systems Biology Markup Language) distribution in Additional file 5.

A MySQL database was set up to host MetaFishNet data. As we elected to use Google App Engine to host the project website [35], a port to Google BigTable database is actually behind the website. The website supports browsing and queries of data at various levels, with graphic display of all pathways. Utility programs in MetaFishNet include 'SeaSpider' for sequence analysis, 'FishEye' for pathway visualization, and 'FisherExpress' for pathway enrichment analysis. SeaSpider is used for both the initial construction and for mapping new sequences to MetaFishNet. FishEye was developed because 1) KEGG graphs can no longer support the much expanded network, and 2) an automatic pathway visualization tool is of great general interest by itself. Our project website provides links to download these programs and model data.

Metabolic genes show less evolutionary diversity

It is now widely accepted that teleost fish underwent an extra round of genome duplication after their evolutionary separation from the mammalian line [36,37]. Genome duplication is an important mechanism for generating gene diversity, as the extra copy can evolve more freely than the single copy before duplication. Only a small portion of these duplicated genes would gain new functionality and remain, while most duplicated genes got lost over time.

When comparing the fish metabolic genes in MetaFishNet to their human orthologs, we have noticed that the level of ortholog mapping differs between metabolic genes and other genes. As seen in Table 2, for the identifiable orthologs, most of the fish species have over 10% more genes than humans, yet the percentages of extra duplicated metabolic genes are significantly less. The final numbers may vary when the genomes are more accurately annotated. Still, these data suggest that metabolic genes are better conserved between human and fish than other genes. This suggests that a core metabolic network was established early in evolution: by the time of the genome duplication in fish, the central metabolic machinery was already well tuned and left little room for changes. By implication, research on some fish metabolic pathways may be easily extrapolated to human.

Comparison between human and fish metabolic pathways

Multiple genes may have the same catalytic activity (isozymes), differing only in their sequences or regulatory contexts. We do not distinguish isozymes in this study, but leave them for future refinement. At the enzyme level, we have identified 911 enzymes from fish genomes. They overlap with the human data by 772 enzymes (Figure 3; Additional file 6 gives a complete list of these enzymes). The true overlap may be greater because the EC numbers in fish were computationally inferred, and are not as well curated as human ECs. We can nonetheless start making some comparisons between human and fish at the pathway level.

Over 50% of the enzymes are in common between human and fish for the majority of the pathways. Table 3 shows the most and least conserved pathways between humans and fish, in terms of the numbers of overlapping enzymes. Since most biomedical research in fish aims to extend the results to human, this pathway comparison reveals important information on how well fish may
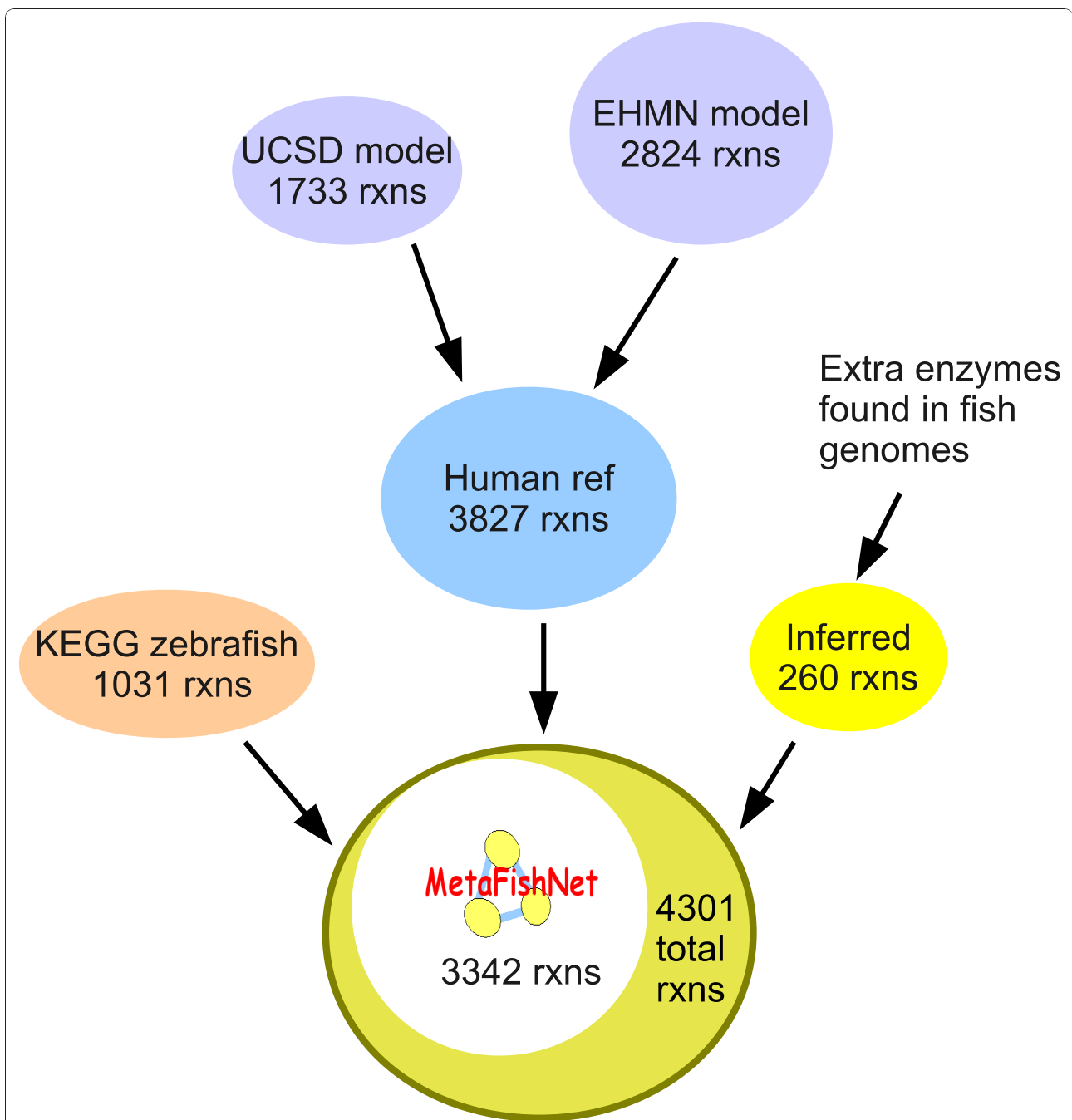
Figure 2 Data integration at reaction level for MetaFishNet. The UCSD and EHMN models were merged into a human reference network, which was then merged with the KEGG zebrafish model and newly inferred reactions based on genome sequences. The total reference model has 4,301 reactions, while 3,342 reactions are included in the fish metabolic network.

model human on a specific subject. For instance, fish may be a good model for studying vitamin B9, but probably a poor model for studying vitamin C.

In the sizable pathway, 'proteoglycan biosynthesis', all 16 enzymes are common between human and fish. This suggests that the whole pathway may be identical between human and fish. Impairment of the proteoglycan biosynthesis pathway is responsible for a major class of enzyme deficiency diseases, mucopolysaccharidosis. Seven clinical types, including Hurler syndrome and Hunter syndrome, have been identified in this class, depending on defects of different enzymes in the pathway (Online Mendelian Inheritance in Man [38]). Given the great similarity between human and fish in this pathway, small fish, with

Table 2 Comparisons between fish and human orthologs

Species	Extra duplicated genes (%)	Extra duplicated metabolic genes (%)
Zebrafish	15.4	0.6
Medaka	8.9	1.5
Takifugu	12.2	3.8
Tetraodon	14.4	5.8
Stickleback	11.9	4.5

An extra round of genome duplication produced more genes in fish than human. The number of total human orthologs found in a fish species is typically around 12,000, as analyzed from Ensembl data.

their high throughput capacity, may be a good model for studying mucopolysaccharidosis.

Omega-3 fatty acids are deemed essential nutrients, boosting a popular dietary preference for fish and fish oil consumption. But fish, just like humans, do not produce omega-3 fatty acids *per se* - they accumulate them from their diet, algae [39]. However, the molecular mechanism of this omega-3 fatty acid accumulation is still unidentified. A theoretical explanation is now provided by our MetaFishNet model. As shown in Figure 4, compared to the human omega-3 fatty acid metabolism, fish lack enzymes such as linoleoyl-CoA desaturase in the pathway. As a result, fish can easily process the metabolites in the top and bottom parts of the pathway, but not the intermediate metabolites, which will then accumulate to a high level. In fact, these intermediate compounds include variants of most of the common omega-3 fatty acids, such as alpha-Linolenic acid, Stearidonic acid, Eicosatetraenoic acid, Eicosapentaenoic acid, Docosapentaenoic acid and Tetracosapentaenoic acid.
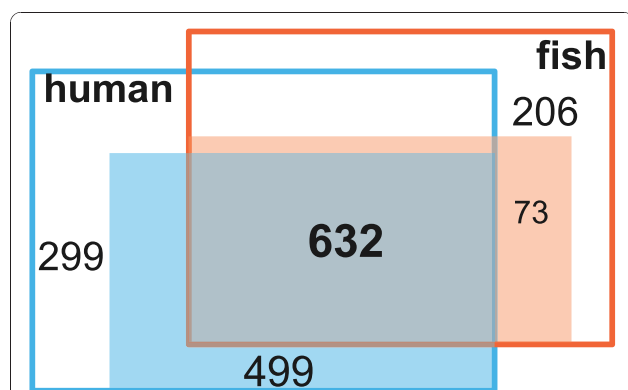

Figure 3 Metabolic enzymes in common between human and fish. Among the 1,430 human enzymes compiled from ExPASy and BRENDA [91] databases, 1,131 are included in the human metabolic models (shaded in light blue). Among the 911 enzymes found in fish genomes, 705 are included in MetaFishNet reactions (shaded in salmon). In the models, 632 enzymes are shared between human and fish. The disparity of numbers reflects that human enzymes are better annotated than fish. Please note that isozymes are not distinguished here.

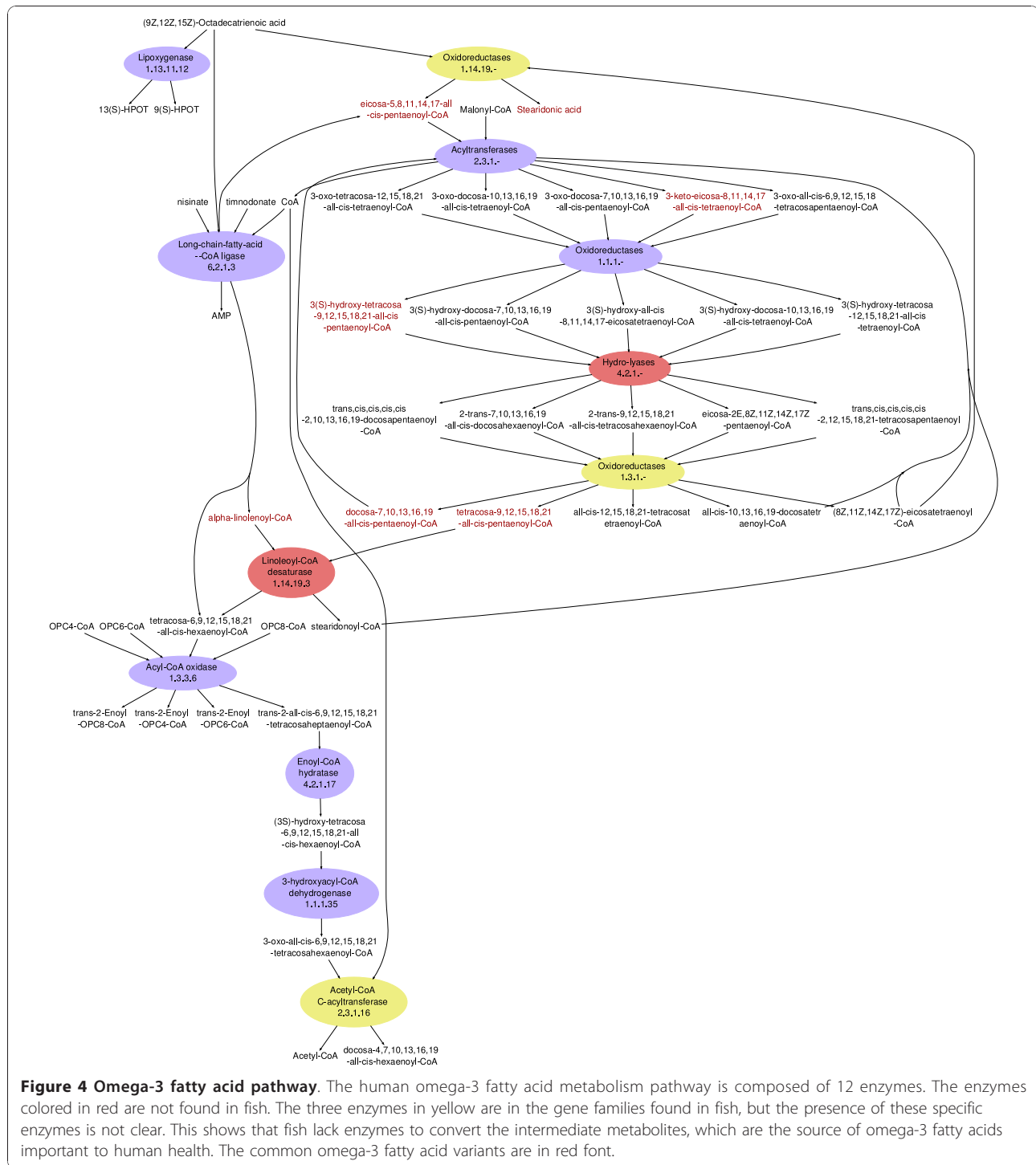**Table 3 Comparisons between fish and human metabolic pathways**

Most conserved pathways			
Pathway	Human ECs	Fish ECs	Overlap	Ratio
---	---	---	---	---
1- and 2-Methylnaphthalene degradation	2	3	2	1
Hyaluronan metabolism	3	3	3	1
Sialic acid metabolism	18	18	18	1
Hexose phosphorylation	5	5	5	1
Electron transport chain	4	5	4	1
Limonene and pinene degradation	3	4	3	1
Proteoglycan biosynthesis	16	16	16	1
Glycosphingolipid biosynthesis - ganglioseries	18	17	17	0.94
N-Glycan degradation	8	7	7	0.87
Di-unsaturated fatty acid beta-oxidation	7	6	6	0.85
Vitamin B1 (thiamin) metabolism	7	6	6	0.85
Glycosphingolipid metabolism	28	24	24	0.85
Glutamate metabolism	14	12	12	0.85
TCA cycle	18	15	15	0.83
Vitamin B9 (folate) metabolism	17	14	14	0.82
Linoleate metabolism	11	9	9	0.81
Least conserved pathways			
Pathway	Human ECs	Fish ECs	Overlap	Ratio
---	---	---	---	---
Phytanic acid peroxisomal oxidation	13	5	5	0.38
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	3	1	1	0.33
Vitamin H (biotin) metabolism	6	2	2	0.33
Vitamin B12 (cyanocobalamin) metabolism	3	2	1	0.33
Glyoxylate and Dicarboxylate metabolism	7	2	2	0.28
Pentose and Glucuronate interconversions	9	2	2	0.22
Ascorbate (vitamin C) and aldarate metabolism	8	1	1	0.12

The ratio is the number of shared ECs over the number of human ECs. Only pathways with three or more enzymes were considered. The complete comparison is given in Additional file 9. Please see Discussion section on the bias towards human data. The sizes of fish pathways may grow with improved annotation, but this is unlikely to change the ratios because all overlapping enzymes are already included here.

It will be interesting to see if this computationally generated hypothesis will be supported by experimental data.

Several metabolic pathways are misregulated in zebrafish liver cancer

We next demonstrate the application of MetaFishNet model to the analysis of gene expression data in a case of zebrafish as a cancer model. Gong and coworkers conducted microarray experiments to examine the similarity between zebrafish and human liver tumors at the

Figure 4 Omega-3 fatty acid pathway. The human omega-3 fatty acid metabolism pathway is composed of 12 enzymes. The enzymes colored in red are not found in fish. The three enzymes in yellow are in the gene families found in fish, but the presence of these specific enzymes is not clear. This shows that fish lack enzymes to convert the intermediate metabolites, which are the source of omega-3 fatty acids important to human health. The common omega-3 fatty acid variants are in red font.

level of gene expression [40]. Although they found the overlapping of gene expression was statistically significant, in-depth data analysis was limited to Gene Set Enrichment Analysis (GSEA) and to two signaling pathways (Wnt-beta-catenin and Ras-MAPK). We shall demonstrate here that MetaFishNet is a valuable addition to the arsenal of microarray data analysis.

The microarray data from [40] were retrieved from Gene Expression Omnibus (GEO [41]) via accession number [GEO:GSE3519]. The arrays contained 16,512 features, with 10 tumor samples and 10 control samples. Significance Analysis of Microarrays (SAM [42]) was used to select 1,888 differentially expressed clones between tumor samples and controls with a False

Discovery Rate under 0.01. (These selected clones are comparable to the 2,315 clones selected by a less mainstream method in the original paper.) The pathway analysis component in MetaFishNet is FisherExpress, which maps the selected genes to enzymes and then to corresponding pathways via queries to the MetaFishNet database. Fisher's Exact Test is used to compute the significance of enrichment of metabolic pathways.

The result, shown in Table 4, suggests that several metabolic pathways are misregulated in zebrafish liver cancer. The identification of the glycolysis and gluconeogenesis pathway reflects the adaptation of tumor cells to aerobic glycolysis, known as the hallmark 'Warburg effect', which also alters pathways closely related to gluconeogenesis, such as butanoate metabolism [43,44]. The reprogramming of metabolism in tumor cells is also believed to generate toxic byproducts [43], in particular elevated levels of reactive oxygen species [45]. The downregulation of xenobiotics metabolism and ROS detoxification reflects these impaired cellular functions in tumor tissues. The involvement of tyrosine metabolism in tumor cells is not clear, but may possibly be related to their excessive tyrosine kinase activities [46,47]. Tryptophan metabolism is known to be part of the immune suppression mechanism by tumor cells [48]. The significance of leukotriene metabolism could come either from tumor cells that use leukotrienes in their strategies for survival, proliferation and migration, or from the inflammation of surrounding tissues [49].

Fatty acid metabolism is also well known to be involved in cancer biology [43,50]. However, the selection of the fatty acid metabolism pathway in our analysis came from three enzymes it shares with the leukotriene

metabolism pathway. Pathway overlap is an inherent limit of this type of analysis, that can only be clarified by further investigation. Several Glycosylphosphatidylinositol(GPI)-anchor proteins are already used as markers for liver cancer [51-53], making (GPI)-anchor biosynthesis an interesting pathway to investigate. The MetaFishNet model thus has been shown to be a valuable tool to identify significantly regulated pathways in expression data. In addition, the regulations can be visualized in the context of each pathway, as exemplified in Figure 5, to facilitate mechanistic studies.
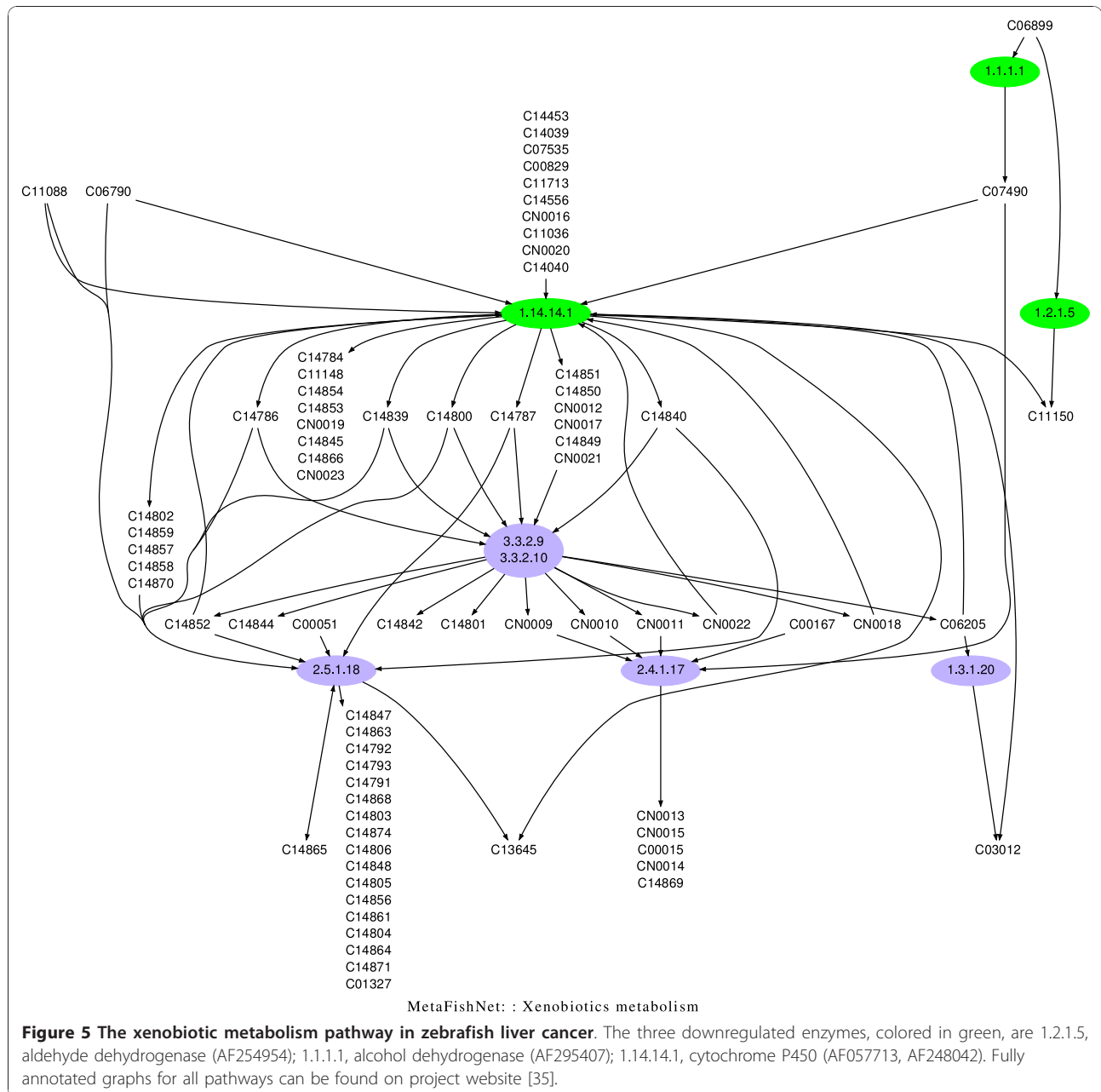
Comparison to KegArray and KEGG pathways

KEGG also offers an expression analysis tool, KegArray [21], which may be used to map differentially expressed genes to zebrafish pathways. For example, the 1,888 selected clones in zebrafish liver cancer in Section 2.4 can be converted to UniGene identifiers and input to KegArray (version 1.2.3). The result is a list of 49 metabolic pathways that match from one to five differentially expressed enzymes (Additional file 7). This is a rather long list, containing about half of all pathways, which raises the question of false positive rate. The problem is caused by the fact that KegArray does not include any pathway statistical analysis, which is important for ranking the significances and reducing false positives at the individual gene level. Pathway enrichment analysis usually takes one of two forms: 1) feature selection followed by set enrichment statistics, such as presented in this paper and 2) competitive statistics without prior feature selection. The best known example of the latter is GSEA [22], which uses Kolmogorov-Smirnov statistics to rank pathways according the positional distribution of member genes. As the MetaFishNet model itself is not tied to any statistical method, we also offer a gene matrix file to be used with GSEA, downloadable at our project website.

Ultimately, the quality of pathway data determines the quality of analysis. MetaFishNet, with 3,342 reactions over the 1,031 reactions in KEGG zebrafish model, not only allows applications to other fish species, but also improve the data for zebrafish. A better comparison between the KEGG zebrafish model and MetaFishNet is to use the same enrichment statistics. That is, we use the KEGG pathways in our software instead of MetaFishNet pathways to reanalyze the zebrafish liver cancer data in Section 2.4. The result is shown in Additional file 8. In comparison to Table 4, leukotriene metabolism and ROS detoxification pathways are missing in the KEGG result as they are absent in the KEGG model.

Xenobiotics metabolism is a pathway that is improved from five enzymes in KEGG to eight enzymes in MetaFishNet. Accordingly, the MetaFishNet pathway has three hits while the KEGG pathway has two hits. The Methane

Table 4 Metabolic pathways that are affected in zebrafish liver cancer with *P*-value < 0.05

MetaFishNet pathway	Selected enzymes	Enzymes in pathway	*P*-value
ROS detoxification	2	2	0.002
3-Chloroacrylic acid degradation	2	2	0.002
Tyrosine metabolism	8	55	0.002
Xenobiotics metabolism	3	8	0.004
Glycolysis and Gluconeogenesis	6	44	0.013
Fatty acid metabolism	3	13	0.019
Butanoate metabolism	3	14	0.023
Leukotriene metabolism	3	17	0.040
Tryptophan metabolism	4	29	0.040
Ascorbate (vitamin C) and aldarate metabolism	1	1	0.046
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1	1	0.046

Figure 5 The xenobiotic metabolism pathway in zebrafish liver cancer. The three downregulated enzymes, colored in green, are 1.2.1.5, aldehyde dehydrogenase (AF254954); 1.1.1.1, alcohol dehydrogenase (AF295407); 1.14.14.1, cytochrome P450 (AF057713, AF248042). Fully annotated graphs for all pathways can be found on project website [35].

metabolism pathway, nonexistent in MetaFishNet, was also identified in KEGG. The KEGG Methane metabolism pathway is rather a bacterial pathway that is mapped to zebrafish with only three reactions. Reaction R06983 is catalyzed by an enzyme (1.1.1.284) that is yet to be confirmed in any fish genome. Reaction R00945 converts 5,10-Methylenetetrahydrofolate to Tetrahydrofolate, thus is assigned to vitamin B9 (folate) metabolism pathway in MetaFishNet. This leaves only one reaction, which does not justify a pathway in MetaFishNet. We think the improved data and pathways in MetaFishNet will benefit downstream studies.

MetaFishNet analysis of cadmium exposure in sheepshead minnows

Finally, we apply MetaFishNet to a fish species with little functional data. Sheepshead minnow (*C. variegatus*) is a common, small estuarine fish that is found along the Atlantic and Gulf coasts of the United States. The US Environmental Protection Agency has adopted *C. variegatus* as a model organism for studying pollution levels in estuarine waters [54]. We have designed a custom DNA microarray with 4,101 clones for sheepshead minnows. Sheepshead minnow larvae were exposed to cadmium, a heavy metal pollutant, for seven days in a

controlled laboratory experiment. DNA microarrays were used to measure their RNA expression. Even though each biological replicate was a pool of 80 individuals, only three biological replicates per group were included in this microarray experiment. The analytical power at the gene level was also weakened because the samples were extracted from whole bodies instead of specific tissues. Indeed, with FDR < 0.05 in SAM, only four clones were selected as significant, including metallothionein, which has been extensively reported to be upregulated by cadmium exposure [55,56].

Another problem is the poor annotation of these microarrays. Less than 40% of our sheepshead minnow clones carry sequence homology to known genes, a situation typical for many fish species that limits the functional information from gene expression.

To analyze the data in MetaFishNet, we first selected 325 differentially expressed clones between the treated group and control group by Wilcoxon's rank sum test (*P* < 0.05). This is a less stringent selection, but additional statistical strength is gained at the pathway level by incorporating collective pathway information. Sheepshead minnow clones were then mapped to MetaFishNet by sequence comparison via SeaSpider. MetaFishNet pathway enrichment was computed again by Fisher's Exact Test and the result is shown in Table 5. The pathways in Table 5 again have overlaps, among which are CYP1A and glutathione S-transferase (GST). The induction of CYP1A and GST by cadmium is in concordance with previous reports [57-61]. Both CYP1A and GST are pivotal detoxification enzymes, and central players in xenobiotics metabolism. The fact that these genes are picked up by pathway analysis and not by SAM demonstrates the improved strength of pathway analysis. The upregulation of four enzymes, CYP1A, GST, acyltransferase and long-chain-fatty-acid-CoA ligase, is indicative of the activation of leukotriene metabolism pathway by the commonly observed inflammation induced by cadmium exposure (Figure 6).

Table 5 Metabolic pathways that are affected by cadmium exposure in sheepshead minnows with *P*-value < 0.05

MetaFishNet pathway	Selected enzymes	Enzymes in pathway	*P*-value
Leukotriene metabolism	4	17	0.001
Fatty acid metabolism	3	13	0.005
Omega-3 fatty acid metabolism	2	7	0.016
Squalene and cholesterol biosynthesis	3	20	0.018
Xenobiotics metabolism	2	8	0.021
Omega-6 fatty acid metabolism	2	10	0.032
Tryptophan metabolism	3	29	0.049

In conclusion, MetaFishNet adds extra functional insight into the otherwise very limited data analysis available for non-model species.

Discussion

We have presented the first genome-wide fish metabolic network model. The first and primary role of our MetaFishNet model is a bioinformatic tool for analyzing high throughput expression data. Two case applications of pathway enrichment analysis are included in this report. Pathway analysis offers two advantages: it is less susceptible to noise than analysis at the level of individual genes, and gives contextual insights to biological mechanisms [62,63]. MetaFishNet has demonstrated good promise to bring these advantages into fish studies. By combining data from five fish genomes, our model overcomes some of the coverage problems in individual genome annotations. However, this also masks the difference between these fish species. While this combined model is recommended for gene expression analysis, species specific data should be consulted for more specific genetic and biochemical studies (available at the project website).

A new visualization tool (FishEye) was developed in this project to draw pathway maps automatically.

Even though visualization tools are abundant, there is a particular challenge to balance automation with the kind of clarity desired in a metabolic map. KEGG, and many other pathway databases, creates graphs manually. Hence, all downstream automatic programs in fact depends on the original manual versions.

CellDesigner [64] is an excellent tool, but essentially is for manual editing. On the other hand, CytoScape [65] and VisANT [66] can do automatic drawing, but their results tend to be cluttered and difficult for detailed studies of metabolic pathways. FishEye is a light-weight and flexible Python program based on the widely used Graphviz package from AT&T Research Labs [67]. Rgraphviz [68] is a similar package that offers R binding of Graphviz. The unique strength of FishEye is its optimization for rendering biological pathways via analyzing network structure and labels. FishEye has worked successfully for this project. Its limit seems to be only challenged by two pathways that exceed 400 edges. For these cases, a 'zoom' feature was introduced to reduce the cluttering of edges. We hope that FishEye will find uses in other similar contexts.

We should emphasize that the knowledge of vertebrate metabolism is still very incomplete. This is already evident when considering the obvious differences between the two human models [20,31]. With the assistance of modularity analysis, we constructed several new pathways that were not present in the reference models. For instance, our analysis showed that all 18 enzymes in a newly identified

Figure 6 The leukotriene metabolism pathway as modulated by cadmium exposure in sheepshead minnow. Four upregulated enzymes are colored in red. Only a partial pathway is shown. Some metabolites are connected by reaction IDs when the enzymes are not known.

'sialic acid metabolism' pathway are in fact present in both fish and humans. This shows both the strength of our construction approach and the incompleteness of current models. In general, when one compares the fish pathways versus human pathways (Table 3), the latter seem to contain more enzymes. Because the UCSD and EHMN projects were intensively curated and contained many more data than previous models, a combined human dataset in this project is unlikely to be surpassed by any computational model. Due to the bias in annotations, fish enzymes that have human homologs are also more likely to be incorporated into MetaFishNet. On the other hand, as discussed above, we actually further augmented the human data through constructing MetaFishNet (demonstrated in Additional file 9).

As a first generation model, MetaFishNet will need much refinement to fully realize the power of a genome-wide metabolic model. Traditionally, metabolism was studied piecemeal by dissecting enzyme activities and tracking metabolites. Powerful new tools have now been introduced to genome-wide models [69,70]. For example, mass balance of metabolites can be achieved by a combination of the stoichiometrics of reactions and physiologically plausible kinetics and thermodynamics of pertinent enzymatic reactions. Even with incomplete

information, system constraints such as metabolite flux can be deduced. Missing reactions in the model can be inferred in a similar fashion. While improvements can be expected from accumulating data and annotations, with this MetaFishNet framework now in place, it is possible to design systematic experiments to define and refine fish metabolome. That is, metabolic constraints can be inferred from MetaFishNet model; experimental data can then be gathered, utilizing mutants or knockouts, to verify and update the model iteratively [71-73]. Such works will lead the way for species specific models.

Recent studies have shown that gene expression data, combined with metabolic network models, can successfully predict metabolic flux regulation in specific biological contexts [74-76]. This opens up an exciting opportunity to advance fish metabolic modeling. Finally, metabolic networks are a natural platform to integrate multiple high throughput data types. For example, Yizhak *et al*. used a *E. coli* metabolic network [30] to combine proteomic data with metabolomics to predict knockout phenotypes [77].

Connor *et al*. combined transcriptomics and metabolomics on Ingenuity's human metabolic pathways http://www.ingenuity.com to identify type two diabetes markers [78]. With the advancing of fish omics, in particular metabolomics [79-81], MetaFishNet is in a good position

to fulfill a similar important role for fish studies. The rate of discovery can be greatly accelerated when MetaFish-Net is combined with these high throughput technologies.

Methods
Identification of fish metabolic enzymes and sequence analysis
All cDNA sequences of the five fish species were retrieved from the Ensembl database [82]. Identification of metabolic genes was accomplished by Gene Ontology (GO) computation [83]. Among the five fish species, only zebrafish had good GO annotations. Sequences from the other four species were analyzed by SeaSpider, our sequence analysis tool. The queries to SeaSpider are first directed against zebrafish sequences, then against reference sequences in the GO database. When homology is found (BLAST E-value under 1E-5 and a minimum 33 of identical bases in local alignment), GO terms are assigned to the sequence in query. All genes with a GO term under the tree of metabolism are considered to be metabolic genes. Even though this initial selection is overly inclusive - for example, transport proteins can also get a GO term under metabolism - only genes that can match to EC numbers are used in Meta-FishNet construction. We inferred EC numbers in two ways. The first approach was to carry over EC numbers from human orthologs. The orthologous relationships between fish and human genes were adopted from Ensembl, which has thoroughly computed ortholog/paralog relationships based on the phylogenetic tree of the gene family. Human EC to gene associations were parsed from the ExPASy database [84] and the EHMN data [31]. The second approach of EC inference was through annotations in the GO database by similarity to the enzyme consensus sequences, which have been constructed across species. It should be pointed out that the EC numbers in MetaFishNet are tentative - the Nomenclature Committee of IUBMB actually requires strict experimental evidence for assigning an official EC number.

Integration of reference reaction data
We first integrated the two high-quality human metabolic models [20,31]. The zebrafish metabolic model was then extracted from KEGG, and combined into the reference data. The UCSD model contained 1,496 genes and 3,311 reactions, counting transport reactions and compartmentalization. A highlight of this work was the manual curation of literature supports, which was labor intensive but improved the data quality.

The EHMN model has 2,322 genes and 2,824 reactions (excluding transport reactions). The EHMN model included previous metabolic data from all major databases,

and streamlined the identities of compounds. Automatic extraction of metabolic models from KEGG has been a challenge. Even though KEGG offers an XML (Extensible Markup Language) distribution (called KGML) of its pathways, molecular interactions were mixed with visual elements in these KGML les. KEGG API (Application Programming Interface) was also limited by not distinguishing reactants from products. We developed a practical solution by combining KGML files and KEGG API, where KGML defines the scope of reactions and API confirms relationships. Our Python script, leveraging on SBML libraries, successfully parsed out the 101 zebrafish metabolic pathways from KEGG (retrieved March 24, 2008), with 517 ECs and 1,031 reactions.

The integration of three models was at both the reaction and pathway levels. Two reactions were considered identical when they have the same enzymes and major compounds. To gain the most compatibility, EC numbers and KEGG compound IDs were used wherever possible. The conventional pathways in MetaFishNet primarily followed the pathway organization in EHMN. Pathways were merged if they shared a significant number of common reactions. Different naming styles were reconciled. For example, the 'Cholesterol Metabolism' pathway in the UCSD model overlaps with the 'Squalene and cholesterol biosynthesis' pathway in the EHMN model by 14 enzymes and 16 reactions. The two pathways were merged during the integration of the two human models. All three reactions in the KEGG zebrafish pathway 'Terpenoid biosynthesis' are included in the human 'Squalene and cholesterol biosynthesis' pathway and were therefore merged with the latter. Nine out of 11 enzymes in the zebrafish 'Biosynthesis of steroids' pathway are included in the human Squalene and cholesterol biosynthesis pathway, and were therefore merged as well. Complete lists of pathway reorganization are given in the Additional file 1. The current model does not take into account cellular compartmentalization.

Ab initio construction, modularity analysis and manual curation
Among the 911fish enzymes identified in this project, 561 could be matched to the reference data. For the remaining 350 enzymes, their associated compounds were retrieved from the KEGG LIGAND database wherever available. These enzyme-compounds interactions formed 260 newly inferred reactions. Since there was no way to distinguish reactants from products in these inferred metabolic data, the directions of these reactions were treated as unknown. These newly inferred reactions, plus the isolated reactions from the reference data, were subjected to a combined approach of module-finding and manual curation. We adopted an algorithm by Mark

Newman, which partitions network modules according to the eigenvectors of a characteristic matrix for the network [34]. The modularity program produced a number of candidate modules, which were then manually inspected for pathway organization. This process iterated until no further change could be made. Isolated reactions were also inspected to determine if they could be attached to existing pathways. At this stage, a number of redundant reactions from UCSD were removed from the model, and pathways with too few reactions were dismantled to isolated reactions. Through this approach, the 'sialic acid metabolism', 'dynorphin metabolism', 'electron transport chain', 'parathion degradation' and 'hexose phosphorylation' pathways were created from *ab initio* construction, while a number of modules were organized into existing pathways (Additional file 1).

Pathway visualization

FishEye, our pathway visualization tool, is built on Networkx and PyGraphviz [85]. It extended a development version of Networkx to support bipartite networks. Many details of styling are manipulated through mid-level markups. In order to keep pathway graphs less cluttered, we did a number of optimizations. Two versions of pathway graphs are offered, one with EC numbers and compound IDs (for example Figure 5) and one with enzyme names and compound names (for example Figure 4 and 6). Both versions for all pathways are available at the project website. Similar edges in a pathway can be merged in the visualized graph, and long names are wrapped. A common practice in the field is to omit all currency metabolites, as they bring on an excessive number of edges. We adopted the list of currency metabolites in [86], as it conforms identically to the most connected nodes in MetaFishNet. However, we leave the inclusion of currency metabolites optional, depending on their degrees in specific pathways.

Expression profiling of sheepshead minnows exposed to cadmium

We have previously generated Suppressive Subtractive Hybridization libraries for sheepshead minnows, and sequenced over 10,000 clones [87]. Based on these sequences, we designed a DNA microarray of 14,494 probes for 4,101 clones. All probes were synthesized on microarray chips by Nimblegen Inc. with four replicates.

Exposures and animal sampling were performed as previously described [88,89]. Cadmium (0.3 mg/L) was administered to sheepshead minnow larvae at 24 hours post hatch via precision syringe pumps in an intermittent flow-through system [90]. The study included three biological replicates, each containing 80 larvae in four cups. After seven days of exposures, whole larvae were

sacrificed and stored in RNAlater (Ambion Inc., Austin, TX). Total RNAs were then extracted using the phenol/chloroform method, and treated with DNase. The purified RNAs were checked by NanoDrop and BioAnalyzer for quality assurance. The labeling of RNAs was carried out according to recommendation by Nimblegen Inc. In short, mRNAs were converted to double-strand cDNA. Cy3-labeled random nonamers were used as primers for DNA polymerase reaction, which produced labeled DNA targets off the double-strand cDNA. These labeled targets were purified and hybridized to microarrays. The resulted fluorescent intensities were corrected by quantile normalization. Data at the probe level were averaged over on-slide replicates, with outliers removed. The expression values at the gene level were summarized as the geometric mean of its probe intensities.

Additional material

Additional file 1: Supplemental method [92-101].
Additional file 2: Species specific statistics of pathways.
Additional file 3: List of MetaFishNet pathways.
Additional file 4: MetaFishNet reaction data.
Additional file 5: SBML distribution of MetaFishNet pathways.
Additional file 6: Fish and human enzymes.
Additional file 7: Analysis of zebrafish liver cancer data by KegArray.
Additional file 8: Analysis of zebrafish liver cancer data by KEGG pathways and Fisher's exact test.
Additional file 9: Complete comparison between fish and human metabolic pathways.

Abbreviations
API: application programming interface; EC: enzyme commission; EHMN: Edinburgh human metabolic network; FDR: false discovery rate; GEO: gene expression omnibus; GO: gene ontology; GSEA: gene set enrichment analysis; IUBMB: international union of biochemistry and molecular biology; KEGG: Kyoto encyclopedia of genes and genomes; KGML: KEGG markup language; SAM: significance analysis of microarrays; SBML: systems biology markup language; UCSD: University of California at San Diego; XML: extensible markup language.

Author details
[1]Gulf Coast Research Laboratory, Department of Coastal Sciences, University of Southern Mississippi, 703 East Beach Drive, Ocean Springs, MS 39564, USA. [2]Current address: Emory Vaccine Center, 954 Gatewood Rd, Atlanta, GA 30329, USA. [3]Current address: Max Planck Institute, August-Thienemann-Str. 2, Ploen 24306, Germany.

Authors' contributions
SL designed and performed most of the computational work. MB designed and supervised the experimental study. AP and MB provided critical guidance of the project and valuable discussions. CSM performed the cadmium exposure of sheepshead minnows. NBP and RR dissected the fish,

extracted and labeled RNA. AP coordinated the sheepshead minnow microarray design and experiments. SL and MB wrote the manuscript.

References

1. Meierjohann S, Schartl M: **From Mendelian to molecular genetics: the Xiphophorus melanoma model.** *Trends in Genetics* 2006, **22**:654-661.
2. Walter R, Kazianis S: **Xiphophorus interspecies hybrids as genetic models of induced neoplasia.** *ILAR Journal/National Research Council, Institute of Laboratory Animal Resources* 2001, **42**:299.
3. Cheek A, Brouwer T, Carroll S, Manning S, McLachlan J, Brouwer M: **Experimental evaluation of vitellogenin as a predictive biomarker for reproductive disruption.** *Environmental Health Perspectives* 2001, **109**:681.
4. Zon L, Peterson R: **In vivo drug discovery in the zebrafish.** *Nature Reviews Drug Discovery* 2005, **4**:35-44.
5. Megason S, Fraser S: **Imaging in systems biology.** *Cell* 2007, **130**:784-795.
6. Sabaliauskas N, Foutz C, Mest J, Budgeon L, Sidor A, Gershenson J, Joshi S, Cheng K: **High-throughput zebrafish histology.** *Methods* 2006, **39**:246-254.
7. Goessling W, North T, Zon L: **Ultrasound biomicroscopy permits in vivo characterization of zebrafish liver tumors.** *Nature Methods* 2007, **4**:551-553.
8. Keller P, Schmidt A, Wittbrodt J, Stelzer E: **Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy.** *Science* 2008, **322**:1065.
9. Area S, Index A: **Animal models of human disease: zebrafish swim into view.** *Nature Reviews Genetics* 2007, **8**:353-367.
10. Guyon J, Steffen L, Howell M, Pusack T, Lawrence C, Kunkel L: **Modeling human muscle disease in zebrafish.** *BBA-Molecular Basis of Disease* 2007, **1772**:205-215.
11. Feitsma H, Cuppen E: **Zebrafish as a cancer model.** *Molecular Cancer Research* 2008, **6**:685.
12. Kokel D, Bryan J, Laggner C, White R, Cheung C, Mateus R, Healey D, Kim S, Werdich A, Haggarty S, MacRae CA, Shoichet B, Peterson RT: **Rapid behavior-based identification of neuroactive small molecules in the zebrafish.** *Nature Chemical Biology* 2010, **6**:231-237.
13. Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, Jang S, Haggarty S, Kokel D, Rubin LL, Peterson RT, Schier AF: **Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation.** *Science* 2010, **327**:348.
14. Snape J, Maund S, Pickford D, Hutchinson T: **Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology.** *Aquatic Toxicology* 2004, **67**:143-154.
15. Ju Z, Wells M, Walter R: **DNA microarray technology in toxicogenomics of aquatic models: Methods and applications.** *Comp Biochem Physiol C Toxicol Pharmacol* 2007, **145**:5-14.
16. Denslow N, Garcia-Reyero N, Barber D: **Fish 'n'chips: the use of microarrays for aquatic toxicology.** *Molecular Biosystems* 2007, **3**:172.
17. Waters M, Fostel J: **Toxicogenomics and systems toxicology: aims and prospects.** *Nature Reviews Genetics* 2004, **5**:936-948.
18. Heijne W, Kienhuis A, van Ommen B, Stierum R, Groten J: **Systems toxicology: applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology.** *Expert Review of Proteomics* 2005, **2**:767-780.
19. Oberhardt M, Palsson B, Papin J: **Applications of genome-scale metabolic reconstructions.** *Molecular Systems Biology* 2009, **5**:320.
20. Duarte N, Becker S, Jamshidi N, Thiele I, Mo M, Vo T, Srivas R, Palsson B: **Global reconstruction of the human metabolic network based on genomic and bibliomic data.** *Proc Natl Acad Sci U S A* 2007, **104**:1777-1782.
21. Wheelock C, Wheelock Å, Kawashima S, Diez D, Kanehisa M, Erk M, Kleemann R, Haeggström J, Goto S: **Systems biology approaches and pathway tools for investigating cardiovascular disease.** *Molecular BioSystems* 2009, **5**:588-602.
22. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub T, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.
23. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44-57.
24. Cox B, Kotlyar M, Evangelou A, Ignatchenko V, Ignatchenko A, Whiteley K, Jurisica I, Adamson S, Rossant J, Kislinger T: **Comparative systems biology of human and mouse as a tool to guide the modeling of human placental pathology.** *Molecular Systems Biology* 2009, **5**:279.
25. Schilling C, Covert M, Famili I, Church G, Edwards J, Palsson B: **Genome-scale metabolic model of Helicobacter pylori 26695.** *Journal of Bacteriology* 2002, **184**:4582-4593.
26. Ma H, Zeng A: **Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms.** *Bioinformatics* 2003, **19**:270.
27. Becker S, Palsson B: **Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation.** *BMC Microbiology* 2005, **5**:8.
28. Heinemann M, Kummel A, Ruinatscha R, Panke S: **In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network.** *Biotechnol Bioeng* 2005, **92**:850-864.
29. Förster J, Famili I, Fu P, Palsson B, Nielsen J: **Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network.** *Genome Research* 2003, **13**:244.
30. Feist A, Henry C, Reed J, Krummenacker M, Joyce A, Karp P, Broadbelt L, Hatzimanikatis V, Palsson B: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Molecular Systems Biology* 2007, **3**:121.
31. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I: **The Edinburgh human metabolic network reconstruction and its functional analysis.** *Molecular Systems Biology* 2007, **3**:135.
32. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Research* 2006, **34**:D354.
33. Ma H, Zhao X, Yuan Y, Zeng A: **Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph.** *Bioinformatics* 2004, **20**:1870-1876.
34. Newman M: **Modularity and community structure in networks.** *Proc Natl Acad Sci U S A* 2006, **103**:8577-8582.
35. **MetaFishNet website.** [http://metafishnet.appspot.com].
36. Jaillon O, Aury J, Brunet F, Petit J, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Cattolico L, Poulain J, et al: **Genome duplication in the teleost sh Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431**:946-957.
37. Vandepoele K, De Vos W, Taylor J, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci U S A* 2004, **101**:1638-1643.
38. **Online Mendelian Inheritance in Man.** [http://www.ncbi.nlm.nih.gov/omim].
39. Surette M: **The science behind dietary omega-3 fatty acids.** *Canadian Medical Association Journal* 2008, **178**:177.
40. Lam SH, Wu YL, Vega VB, Miller LD, Spitsbergen J, Tong Y, Zhan H, Govindarajan KR, Lee S, Mathavan S, Murthy KR, Buhler DR, Liu ET, Gong Z: **Conservation of gene expression signatures between zebrafish and human liver tumors and tumor progression.** *Nature Biotechnology* 2005, **24**:73-75.
41. **Gene Expression Omnibus.** [http://www.ncbi.nlm.nih.gov/geo/].
42. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98**:5116-5121.
43. Hsu P, Sabatini D: **Cancer cell metabolism: Warburg and beyond.** *Cell* 2008, **134**:703-707.
44. Perroud B, Lee J, Valkova N, Dhirapong A, Lin P, Fiehn O, Kültz D, Weiss R: **Pathway analysis of kidney cancer using proteomics and metabolic profiling.** *Molecular Cancer* 2006, **5**:64.
45. Pelicano H, Carney D, Huang P: **ROS stress in cancer cells and therapeutic implications.** *Drug Resistance Updates* 2004, **7**:97-110.
46. Kroemer G, Pouyssegur J: **Tumor cell metabolism: cancer's Achilles' heel.** *Cancer Cell* 2008, **13**:472-482.
47. Hitosugi T, Kang S, Vander Heiden MG, Chung TW, Elf S, Lythgoe K, Dong S, Lonial S, Wang X, Chen GZ, Xie J, Gu TL, Polakiewicz RD, Roesel JL, Boggon TJ, Khuri FR, Gilliland DG, Cantley LC, Kaufman J, Chen J: **Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth.** *Science Signaling* 2009, **2**:ra73.

48. Uyttenhove C, Pilotte L, Théate I, Stroobant V, Colau D, Parmentier N, Boon T, Van den Eynde B: **Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2, 3-dioxygenase.** *Nature Medicine* 2003, **9**:1269-1274.

49. Wang D, DuBois R: **Eicosanoids and cancer.** *Nature Reviews Cancer* 2010, **10**:181-93.

50. Zhou W, Tu Y, Simpson P, Kuhajda F: **Malonyl-CoA decarboxylase inhibition is selectively cytotoxic to human breast cancer cells.** *Oncogene* 2009, **28**:2979-2987.

51. Wang L, Vuolo M, Suhrland M, Schlesinger K: **HepPar1, MOC-31, pCEA, mCEA and CD10 for distinguishing hepatocellular carcinoma vs. metastatic adenocarcinoma in liver fine needle aspirates.** *Acta Cytologica* 2006, **50**:257.

52. Kondo K, Chijiiwa K, Funagayama M, Kai M, Otani K, Ohuchida J: **Differences in long-term outcome and prognostic factors according to viral status in patients with hepatocellular carcinoma treated by surgery.** *Journal of Gastrointestinal Surgery* 2008, **12**:468-476.

53. Kakar S, Gown A, Goodman Z, Ferrell L: **Best practices in diagnostic immunohistochemistry: hepatocellular carcinoma versus metastatic neoplasms.** *Archives of Pathology & Laboratory Medicine* 2007, **131**:1648.

54. EPA: *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Marine and Estuarine Organisms.* third edition. United States Environmental Protection Agency; 2002.

55. Hawse J, Cumming J, Oppermann B, Sheets N, Reddy V, Kantorow M: **Activation of metallothioneins and -crystallin/sHSPs in Human lens epithelial cells by specific metals and the metal content of aging clear human lenses.** *Investigative Ophthalmology & Visual Science* 2003, **44**:672-679.

56. Loumbourdis N, Kostaropoulos I, Theodoropoulou B, Kalmanti D: **Heavy metal accumulation and metallothionein concentration in the frog Rana ridibunda after exposure to chromium or a mixture of chromium and cadmium.** *Environmental Pollution* 2007, **145**:787-792.

57. Yang L, Kemadjou J, Zinsmeister C, Bauer M, Legradi J, Müller F, Pankratz M, Jäkel J, Strähle U: **Transcriptional profiling reveals barcode-like toxicogenomic responses in the zebrafish embryo.** *Genome Biology* 2007, **8**:R227.

58. Koskinen H, Pehkonen P, Vehniäinen E, Krasnov A, Rexroad C, Afanasyev S, Mölsa H, Oikari A: **Response of rainbow trout transcriptome to model chemical contaminants.** *Biochem Biophys Res Commun* 2004, **320**:745-753.

59. Williams T, Diab A, Ortega F, Sabine V, Godfrey R, Falciani F, Chipman J, George S: **Transcriptomic responses of European flounder (Platichthys flesus) to model toxicants.** *Aquatic Toxicology* 2008, **90**:83-91.

60. Anwar-Mohamed A, Elbekai R, El-Kadi A: **Regulation of CYP1A1 by heavy metals and consequences for drug metabolism.** *Expert Opin Drug Metab Toxicol* 2009, **5**:501-21.

61. Casalino E, Sblano C, Calzaretti G, Landriscina C: **Acute cadmium intoxication induces alpha-class glutathione S-transferase protein synthesis and enzyme activity in rat liver.** *Toxicology* 2006, **217**:240-245.

62. Segal E, Friedman N, Kaminski N, Regev A, Koller D: **From signatures to models: understanding cancer using microarrays.** *Nature Genetics* 2005, **37**:S38-S45.

63. Nam D, Kim S: **Gene-set approach for expression pattern analysis.** *Briefings in Bioinformatics* 2008, **9**:189.

64. Funahashi A, Morohashi M, Kitano H, Tanimura N: **CellDesigner: a process diagram editor for gene-regulatory and biochemical networks.** *Biosilico* 2003, **1**:159-162.

65. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Research* 2003, **13**:2498.

66. Hu Z, Mellor J, Wu J, DeLisi C: **VisANT: an online visualization and analysis tool for biological interaction data.** *BMC Bioinformatics* 2004, **5**:17.

67. **Graphviz.** [http://graphviz.org/].

68. Gentry J, Carey V, Gansner E, Gentleman R: **Laying out pathways with Rgraphviz.** *R News* 2004, **4**:14-18[http://CRAN.R-project.org/doc/Rnews/].

69. Terzer M, Maynard N, Covert M, Stelling J: **Genome-scale metabolic networks.** *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 2009, **1**:285-297.

70. Breitling R, Vitkup D, Barrett M: **New surveyor tools for charting microbial metabolic maps.** *Nature Reviews Microbiology* 2008, **6**:156-161.

71. Ideker T, Thorsson V, Ranish J, Christmas R, Buhler J, Eng J, Bumgarner R, Goodlett D, Aebersold R, Hood L: **Integrated genomic and proteomic**

analyses of a systematically perturbed metabolic network. *Science* 2001, **292**:929.

72. Covert M, Knight E, Reed J, Herrgard M, Palsson B: **Integrating high-throughput and computational data elucidates bacterial networks.** *Nature* 2004, **429**:92-96.

73. Shlomi1 T, Cabili M, Ruppin E: **Predicting metabolic biomarkers of human inborn errors of metabolism.** *Molecular Systems Biology* 2009, **5**:263.

74. Becker S, Palsson B: **Context-specific metabolic networks are consistent with experiments.** *PLoS Computational Biology* 2008, **4**:e1000082.

75. Shlomi T, Cabili M, Herrgård M, Palsson B, Ruppin E: **Network-based prediction of human tissue-specific metabolism.** *Nature Biotechnology* 2008, **26**:1003-1010.

76. Colijn C, Brandes A, Zucker J, Lun D, Weiner B, Farhat M, Cheng T, Moody D, Murray M, Galagan J: **Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production.** *PLoS Computational Biology* 2009, **5**:e1000489.

77. Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T: **Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model.** *Bioinformatics* 2010, **26**:i255.

78. Connor S, Hansen M, Corner A, Smith R, Ryan T: **Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes.** *Molecular BioSystems* 2010, **6**:909-921.

79. Samuelsson L, Larsson D: **Contributions from metabolomics to fish research.** *Molecular BioSystems* 2008, **4**:974-979.

80. Bundy J, Davey M, Viant M: **Environmental metabolomics: a critical review and future perspectives.** *Metabolomics* 2009, **5**:3-21.

81. Williams T, Wu H, Santos E, Ball J, Katsiadaki I, Brown M, Baker P, Ortega F, Falciani F, Craft J, Tyler CR, Chipman JK, Viant MR: **Hepatic transcriptomic and metabolomic responses in the stickleback (Gasterosteus aculeatus) exposed to environmentally relevant concentrations of dibenzanthracene.** *Environmental Science & Technology* 2009, **43**:6341-6348.

82. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, et al: **Ensembl 2007.** *Nucleic Acids Research* 2007, **35**:D610-D617.

83. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nature Genetics* 2000, **25**:25-9.

84. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel R, Bairoch A: **ExPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic Acids Research* 2003, **31**:3784.

85. **Networkx and PyGraphviz.** [http://networkx.lanl.gov].

86. Holme P, Huss M: *Currency metabolites and network representations of metabolism* 2008, Arxiv preprint arXiv:0806.2763.

87. Pozhitkov A, Pirooznia M, Ryan R, Zhang C, Gong P, Perkins E, Deng Y, Brouwer M: **Generation and analysis of expressed sequence tags from the Sheepshead minnow (Cyprinodon variegatus).** *BMC Genomics* 2010, **11**:S4.

88. Hendon L, Carlson E, Manning S, Brouwer M: **Molecular and developmental effects of exposure to pyrene in the early life-stages of Cyprinodon variegatus.** *Comp Biochem Physiol C Toxicol Pharmacol* 2008, **147**:205-215.

89. Brouwer M, Brown-Peterson N, Hoexum-Brouwer T, Manning S, Denslow N: **Changes in mitochondrial gene and protein expression in grass shrimp, Palaemonetes pugio, exposed to chronic hypoxia.** *Marine Environmental Research* 2008, **6**:143.

90. Manning C, Schesny A, Hawkins W, Barnes D, Barnes C, Walker W: **Exposure methodologies and systems for long-term chemical carcinogenicity studies with small fish species.** *Toxicology Mechanisms and Methods* 1999, **9**:201-217.

91. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D: **BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009.** *Nucleic Acids Research* 2009, **37**:D588.

92. Albert R, Barabási A: **Statistical mechanics of complex networks.** *Rev Mod Phys* 2002, **74**:47-97.

93. Barabasi A, Oltvai Z: **Network biology: understanding the cell's functional organization.** *Nature Reviews Genetics* 2004, **5**:101-113.

94. Jeong H, Tombor B, Albert R, Oltvai Z, Barabasi A: **The large-scale organization of metabolic networks.** *Nature* 2000, **407**:651-653.
95. Newman ME, Girvan M: **Finding and evaluating community structure in networks.** *Phys Rev E Stat Nonlin Soft Matter Phys* 2004, **69**:026113.
96. Wagner A, Fell DA: **The small world inside large metabolic networks.** *Proc Biol Sci* 2001, **268**:1803-1810.
97. Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T: **Exploring the pathway structure of metabolism: decomposition into subnetworks and application to Mycoplasma pneumoniae.** *Bioinformatics* 2002, **18**:351-61.
98. Huss M, Holme P: **Currency and commodity metabolites: their identification and relation to the modularity of metabolic networks.** *IET Syst Biol* 2007, **1**:280-285.
99. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research* 1997, **25**:3389-3402.
100. Sprague J, Doerry E, Douglas S, Westerfield M: **The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research.** *Nucleic Acids Research* 2001, **29**:87.
101. Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe D, Mani P, Ramachandran S, Schaper K, Segerdell E, Song P, Sprunger B, Taylor S, Van Slyke E, Westerfield M: **The Zebrafish Information Network: the zebrafish model organism database.** *Nucleic Acids Research* 2006, **34**:D581.