nature biotechnology

Conservation of gene expression signatures between zebrafish and human liver tumors and tumor progression

Siew Hong Lam[1], Yi Lian Wu[1], Vinsensius B Vega[2], Lance D Miller[2], Jan Spitsbergen[3], Yan Tong[1], Huiqing Zhan[1], Kunde R Govindarajan[2], Serene Lee[2], Sinnakarupan Mathavan[2], Karuturi R Krishna Murthy[2], Donald R Buhler[3], Edison T Liu[2] & Zhiyuan Gong[1]

The zebrafish (*Danio rerio*) has been long advocated as a model for cancer research, but little is known about the real molecular similarities between zebrafish and human tumors. Comparative analysis of microarray data from zebrafish liver tumors with those from four human tumor types revealed molecular conservation at various levels between fish and human tumors. This approach provides a useful strategy for identifying an expression signature that is strongly associated with a disease phenotype.

Although zebrafish has been long proposed as a cancer model[1,2], one outstanding question is how similar are zebrafish and human tumors at the molecular level. Gene-expression profiles reflecting similar disease phenotypes in phylogenetically related species, for example, human and mouse, are usually conserved as the important regulatory elements that modulate gene expression are largely conserved among these species[3,4]. We hypothesized that regulatory elements that are strongly associated with a disease phenotype may also be conserved in phylogenetically more distant species, such as human and fish. If this were true, comparison of gene-expression profiles of a fish model of a disease and that disease in humans would be a useful strategy to identify expression signatures that are strongly associated with the disease.

To test this hypothesis, we generated liver tumors in zebrafish, by treating them with carcinogens, for comparative gene-expression analysis (**Supplementary Methods** and **Supplementary Fig. 1** online). Experimental procedures were approved by the Oregon State University's Institutional Animal Care and Use Committee. A gene set (ZLTDEGS; Zebrafish Liver Tumor Differentially Expressed Gene Set), consisting of 2,315 gene features (representing 1,861 zebrafish Unigene clusters (Build 85)), was obtained by comparing the expression profiles of zebrafish liver tumors with normal liver tissues using a Wilcoxon rank-sum test adjusted for multiple hypothesis testing at a false discovery rate (FDR) < 5%. We were able to automate the mapping of 1,334 unique zebrafish Unigene clusters to 1,942 unique human Unigene clusters (Build 186) using the NCBI HomoloGene database (Build 43.1) uploaded onto our zebrafish microarray annotation database[5] (http://giscompute.gis.a-star.edu.sg/~govind/zebrafish/index.html) (**Supplementary Methods** and **Supplementary**

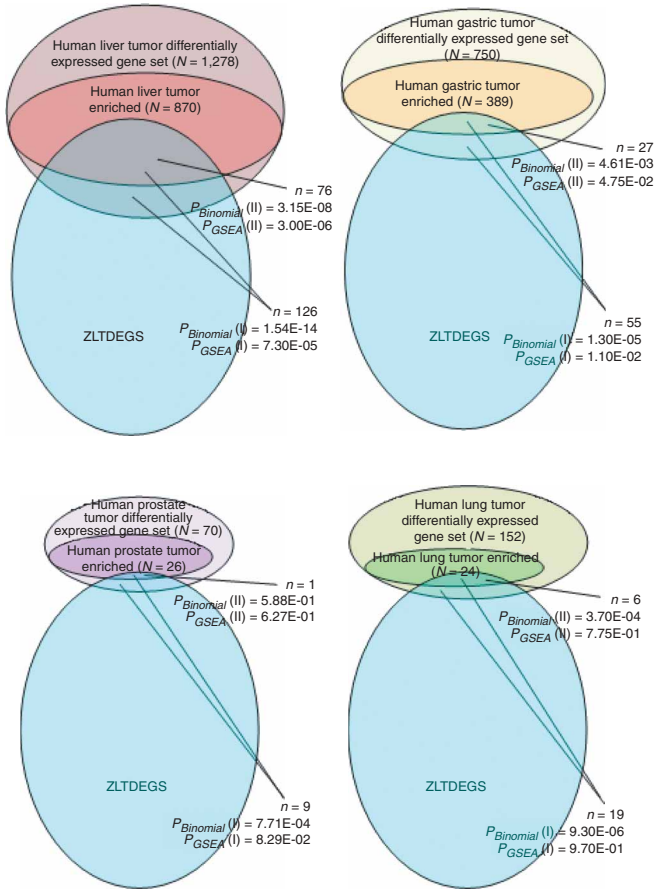

Figure 1 Venn diagrams showing the intersections between zebrafish ZLTDEGS and the human liver, gastric, prostate and lung tumor microarray data sets. N = number of genes in respective human tumor differentially expressed or enriched gene sets. n = number of overlapping genes between ZLTDEGS with respective human tumor differentially expressed or enriched gene set. $P_{Binomial}$ = binomial test probability; P_{GSEA} = gene set enrichment analysis probability. Only intersecting genes with similar expression direction in zebrafish and human data sets are considered.

[1]Department of Biological Sciences, 14 Science Drive 4, National University of Singapore, Singapore 117543. [2]Genome Institute of Singapore, Agency for Science Technology and Research, Genome, 60 Biopolis Street, Singapore 138672. [3]Department of Environmental and Molecular Toxicology and Marine/Freshwater Biomedical Sciences Center, 1007 Agricultural and Life Sciences Building, Oregon State University, Corvallis, Oregon 97331, USA. Correspondence should be addressed to Z.G. (dbsgzy@nus.edu.sg).

Received 3 June; accepted 24 October; published online 4 December 2005; doi:10.1038/nbt1169

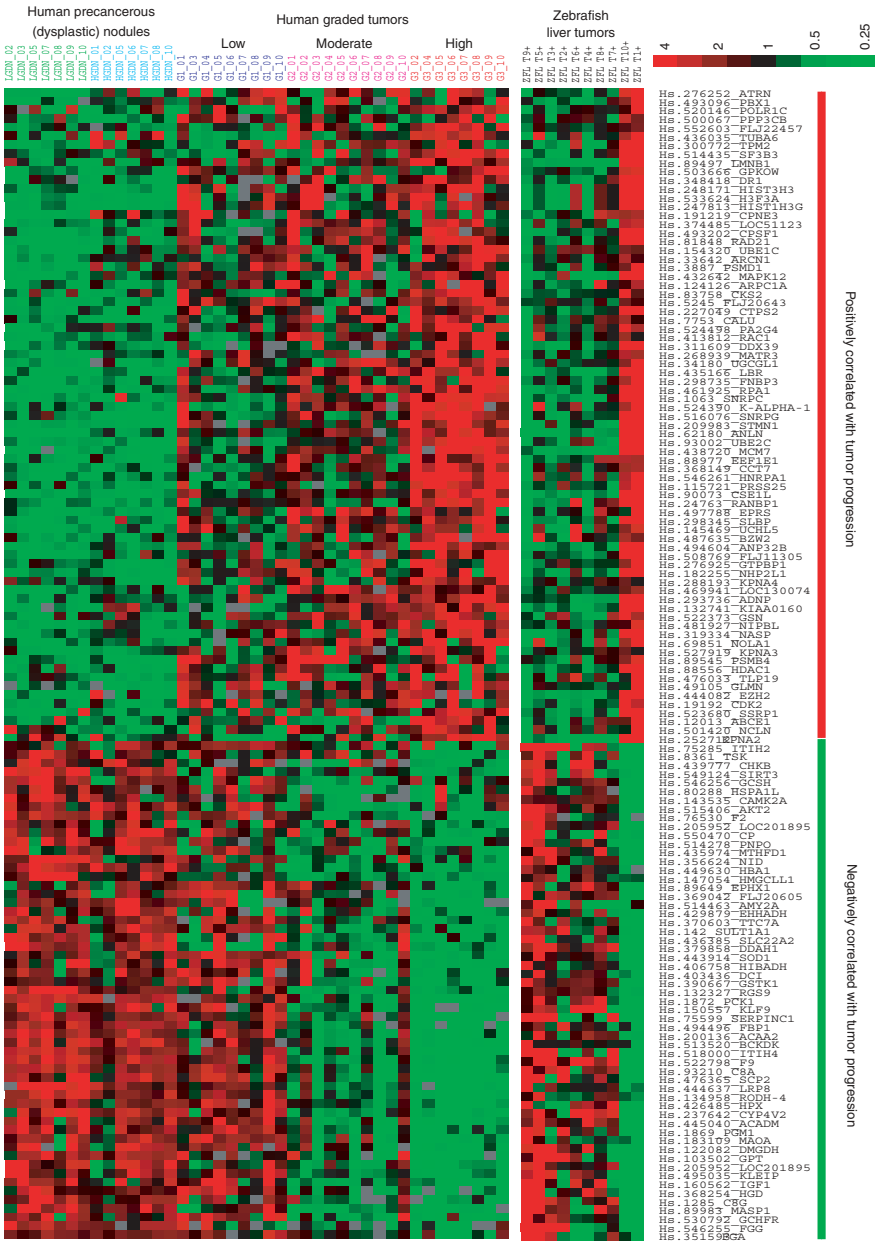

Figure 2 Expression profiles of 132 genes showing similar correlation with tumor progression in both zebrafish and human liver tumors (LGDN, low-grade dysplastic nodule; HGDN, high-grade dysplastic nodule; G1–G3, grade 1–3). The color in each cell reflects the expression level of the corresponding gene in the corresponding tissue sample, relative to its mean expression level across the entire respective set of human and zebrafish tissue samples. Gray cells indicate missing or excluded data.

Data online). Based on gene ontology annotation of the putative human homologs mapped to ZLTDEGS, a large number of genes coding for proteins involved in cell cycle/proliferation, apoptosis, DNA replication and repair, metastasis and cytoskeletal organization, protein synthesis and liver-specific functions were deregulated, indicating that the zebrafish liver tumors possess the general molecular hallmarks of human cancer (**Supplementary Table 1** online).

To further evaluate the significance of ZLTDEGS in relation to human cancers, we explored the intersection of our data with human cancer microarray data from four cancer types, that is, liver[6,7], gastric[8] (unpublished data) (**Supplementary Methods** online), prostate[9,10] and lung[11,12]. Specifically, we asked whether the human homolog set of ZLTDEGS is significantly overrepresented and correlated with differentially expressed genes in human tumors (FDR < 1%) using a Binomial Test and Gene Set Enrichment Analysis (GSEA)[4], respectively (**Supplementary Methods** online). As indicated by $P_{Binomial}$ (I) (**Fig. 1**), the intersections between ZLTDEGS and the differentially expressed gene set from each human tumor type were all significant ($P_{Binomial}$ (I) < 1.00E-03) indicating that the ZLTDEGS contains genes associated generally with human tumors. However, the different levels of significance between the intersecting gene sets suggest that zebrafish liver tumors are closer to human liver tumors ($P_{Binomial}$ (I) = 1.54E-14) than to the other three human tumor types (1.00E-07 < $P_{Binomial}$ (I) < 1.00E-03). GSEA, which considers the relative position of genes in the human data sets, further revealed that ZLTDEGS contains genes that are significantly (P_{GSEA}(I) = 7.30E-05) correlated with the top ranking genes (ranked based on FDR values) in the human liver tumor set and are less or not significantly (P_{GSEA}(I) > 1.00E-02) correlated with the top ranking genes in other human tumor types (**Supplementary Table 2** online). The intersection between ZLTDEGS and the differentially expressed gene set of human liver tumors, which comprised 126 genes, represents a conserved expression signature that is correlated with genes that are highly significant in human liver tumors (**Supplementary Table 3** online).

Next, we asked whether ZLTDEGS possesses an expression signature that is more consistently associated with human liver tumors than with human gastric, prostate and lung tumor types. To do so, we intersected ZLTDEGS with the respective human gene set enriched for one particular tumor type (**Supplementary Methods** online). Notably, the intersection between ZLTDEGS and the human gene set enriched for liver tumors, which comprised 76 genes, was significantly overrepresented ($P_{Binomial}$(II) = 3.15E-08) whereas the intersection between ZLTDEGS and the human gene sets enriched for other nonliver tumor types appeared to be less or not significantly overrepresented ($P_{Binomial}$(II) > 1.00E-04) (**Fig. 1**). When rank-ordered position of the genes was considered, ZLTDEGS correlated significantly (P_{GSEA}(II) = 3.00E-06) with the top ranking genes in the human gene set enriched for liver tumors but the correlation with the top ranking genes for human gene sets enriched for other nonliver tumor types was marginally or not significant at all (P_{GSEA}(II) > 4.00E-02) (**Supplementary Table 2** online). The combination of the 76 genes (**Supplementary Table 4** online) represents an expression signature that is more consistently associated with

human liver tumor than other human tumor types, and the molecular conservation underscores their basic role in liver tumor.

A closer examination of the ZLTDEGS suggests deregulation of Wnt-β-catenin and Ras-MAPK pathways in zebrafish liver tumors, which are frequently associated with human liver cancers[13]. Using real-time PCR, we have validated 63 genes from ZLTDEGS, including several genes associated with the two pathways and some unknown/unannotated genes that are potentially novel cancer genes, and confirmed that they are significantly ($P < 0.05$) differentially expressed (**Supplementary Table 5** online). These data suggest the potential usefulness of zebrafish liver tumors for investigating Wnt-β-catenin and Ras-MAPK pathways in human liver cancer, and for identifying novel cancer genes.

To investigate if the zebrafish liver tumors shared any common molecular features with human liver cancer progression, we compared our data set with another set of human liver cancer microarray data (unpublished data; **Supplementary Methods** online), where the samples had been histopathologically graded according to clinical stages. Among 3,084 unique genes that were significantly associated with human liver tumor grade, there were 132 genes overlapping with ZLTDEGS that shared strikingly similar expression profiles correlating with tumor progression (**Fig. 2**). Remarkably, zebrafish tumors ZFL T1 and T10, which were histologically most anaplastic among all the samples investigated, showed expression profiles for these 132 genes resembling human high-grade liver tumors, whereas ZFL T9, the least anaplastic of all samples investigated, showed expression profiles for the 132 genes resembling human precancerous nodules and low-grade liver tumors. Low-grade tumors exhibit higher expression of genes encoding proteins abundant in liver whereas higher grade tumors show higher expression of genes associated with cell cycle, cytoskeletal organization, metastasis, RNA processing and protein synthesis. The findings suggest that there exist molecular similarities between zebrafish and human liver tumors that extends to tumor progression.

In summary, we have presented evidence demonstrating conservation of expression profiles at different levels between fish and human tumors. More importantly, this study demonstrates the potential of applying comparative transcriptome profile analysis of phylogenetically distant species to the identification of expression signatures that have a strong association with a disease phenotype and whose evolutionary conservation emphasizes their fundamental importance and clinical potential. Finally, this study provides molecular evidence highlighting the potential of zebrafish for modeling human liver cancer.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank S.Y. Neo, S.W. Nam and J.Y. Lee for the human liver cancer data sets. This work is supported by the BioMedical Research Council, Singapore, US Public Health Service Grants ES011587, ES03850, ES00210 and RR12546.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Amatruda, J.F., Shepard, J.L., Stern, H.M. & Zon, L.I. *Cancer Cell* **1**, 229–231 (2002).
2. Stern, H.M. & Zon, L.I. *Nat. Rev. Cancer* **3**, 533–539 (2003).
3. Lee, J. *et al. Nat. Genet.* **36**, 1306–1311 (2004).
4. Sweet-Cordero, A. *et al. Nat. Genet.* **37**, 48–55 (2005).
5. Mathavan, S. *et al. PLoS Genet.* **1**, 260–275 (2005).
6. Chen, X. *et al. Mol. Biol. Cell* **13**, 1929–1939 (2002).
7. Neo, S.Y. *et al. Hepatology* **39**, 944–953 (2004).
8. Chen, X. *et al. Mol. Biol. Cell* **14**, 3208–3215 (2003).
9. Dhanasekaran, S.M. *et al. Nature* **412**, 822–826 (2001).
10. Welsh, J.B. *et al. Cancer Res.* **61**, 5974–5978 (2001).
11. Bhattacharjee, A. *et al. Proc. Natl. Acad. Sci. USA* **98**, 13790–13795 (2001).
12. Garber, M.E. *et al. Proc. Natl. Acad. Sci. USA* **98**, 13784–13789 (2001).
13. Feitelson, M.A. *et al. Oncogene* **21**, 2593–2604 (2002).