BASIC—ALIMENTARY TRACT

FACS-Assisted Microarray Profiling Implicates Novel Genes and Pathways in Zebrafish Gastrointestinal Tract Development

CARSTEN STUCKENHOLZ,* LILI LU,* PRAKASH THAKUR,* NAFTALI KAMINSKI,‡ and NATHAN BAHARY*,§

*Department of Medicine, Division of Hematology/Oncology, ‡Dorothy P. & Richard P. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, and §Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

BACKGROUND & AIMS: The zebrafish *Danio rerio* is an excellent model system for mammalian gastrointestinal development. To identify differentially regulated genes important in gastrointestinal organogenesis, we profiled the transcriptome of the zebrafish developing gastrointestinal tract. **METHODS:** Embryos from a transgenic zebrafish line expressing green fluorescent protein (GFP) in the developing intestine, liver, and pancreas were dissociated at 4 developmental time points, their cells sorted based on GFP expression with fluorescence-activated cell sorting (FACS), and analyzed with microarrays. To improve our analysis, we annotated the Affymetrix Zebrafish GeneChip with human orthologs. **RESULTS:** Transcriptional profiling showed significant differences between GFP^+ and GFP^- cells. Upregulated genes and pathways were consistent with mammalian gastrointestinal development, such as hepatic nuclear factor gene networks and cancer. We implicate the phosphatidylinositol 3 kinase (PI3K) pathway and show that inhibition with LY294002 causes gastrointestinal defects in zebrafish. We identified novel genes, such as the microRNAs miR-217 and miR-122, the tight junction protein claudin c, the gene fam136a, and a zebrafish tetraspanin. Novel pathways include genes containing a putative transcription factor binding sequence, GGAANCGGAANY, and a nucleolar gene network. The zebrafish microarrays also identify a set of 32 genes that may mediate the effects of gain of chromosome arm 8q in human colon, liver, and pancreatic cancers. **CONCLUSIONS: We successfully combine FACS and microarray profiling to follow organogenesis throughout development. These experiments identify novel genes and pathways that probably play a role in mammalian gastrointestinal development and are potential targets for therapeutic intervention in the management of gastrointestinal disease and cancer.**

Among vertebrates, the zebrafish is particularly well suited to study organogenesis of the gastrointestinal (GI) tract, because this vertebrate can be easily manipulated with both forward and reverse genetics and because the zebrafish organs have been described in physiologic detail and are highly similar to their mammalian counterparts.[1–8] To better understand molecular aspects of zebrafish and mammalian GI development, we transcriptionally profiled the developing zebrafish GI tract, using the gut green fluorescent protein (GFP) transgenic line, expressing GFP in the developing liver, gut, and pancreas.[2–5]

In gut GFP zebrafish, a rod of endodermal cells can be seen in the midsection of the embryo at 1 day post fertilization (dpf).[5] From this rod, the liver and the pancreas begin to successively bud, grow, and differentiate into mature organs,[2,3] but alternative models are possible.[7,8] Specification of hepatocytes requires integration of fibroblast growth factor, bone morphogenetic protein, and Wingless/Int (Wnt)-type signals.[9–11] By 4 dpf, much of liver and pancreatic development appears to be complete.[2,3] Intestinal development begins after 1 dpf, when the endodermal rod itself thickens anteriorly, followed by lumen formation, intestinal cell differentiation, epithelial folding, and onset of gut motility by 5 dpf.[4,6,7,12] These processes are controlled by several transcription factors and the gene networks they control, including *hhex*, *hnf4a*, and *hnf6/onecut3* in liver development, and which are conserved from zebrafish to human beings.[13–17] Because of this conservation, lessons from zebrafish probably also apply to mammalian GI development.

We selected zebrafish embryos at 4 critical developmental time points and separated them by fluorescence-activated cell sorting (FACS) into GFP^+ and GFP^- cell populations, representing GI tissues and the remainder of the embryo. We then profiled the transcriptome by microarray and analyzed our data for known and novel genes expressed in the GI tract during organogenesis and for known and novel pathways and cellular components

Abbreviations used in this paper: ATP, adenosine triphosphate; dpf, days post fertilization; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein; GI, gastrointestinal; GSEA, Gene Set Enrichment Analysis; HCC, hepatocellular carcinoma; HNF, hepatic nuclear factor; IPA, Ingenuity's Pathway Analysis; kb, kilobase; PI3K, phosphatidylinositol 3 kinase.

0016-5085/09/$36.00
doi:10.1053/j.gastro.2009.06.050

that may be important in GI development and function. We found evidence for some aspects of GI organ function as early as 2 dpf, although substantial functionality is probably not attained until 4 dpf. Further evidence suggested a role for phosphatidylinositol 3 kinase (PI3K) signaling in GI development; treatment with a PI3K inhibitor resulted in liver defects. We identified a novel DNA sequence pattern that is significantly enriched in the promoters of GI-specific genes that may be bound by a potentially critical, unknown transcription factor. Our results also support a regulatory role for microRNAs in GI development. Finally, we identified candidate genes located on human chromosome arm 8q that may mediate the effects of early gain of 8q in the development of liver, pancreas, and colon cancers. This study presents proof that combining spatially and temporally restricted GFP expression with flow sorting and microarray analysis is instructive for understanding organogenesis and human disease processes.

Materials and Methods

Fish Husbandry

All fish husbandry was carried out in accordance with local institutional animal care and use committee protocols. We selected a line of homozygous *Tg(XlEef1a1:GFP)*^{s854} transgenic fish (gutGFP) with the strongest, nonvariegating GFP expression throughout the GI tract. We grew embryos in standard embryo media, supplemented with 0.003% 1-phenyl-2-thiourea to reduce autofluorescence, and harvested approximately 700 embryos at 2, 3, 4, and 6 dpf on ice at the same daily time point.

Isolation of GI Cells by FACS and RNA Purification

Embryos were washed in phosphate-buffered saline with Tween 3 times, resuspended in Hank's balanced salt solution with 0.25% trypsin and 0.1% EDTA (CellGro, Manassas, VA) supplemented with 40 μg/mL Proteinase K and 10 μg/mL Collagenase, and dissociated in a Wheaton tapered tissue grinder. Cells were collected by centrifugation and resuspended in tropomyosin I TMI buffer (100 mmol/L NaCl, 5 mmol/L KCI, 5 mmol/L HEPES, 1% polyethelene glycol 20000).[18] After straining the slurry through a 40-μm filter, cell suspensions were separated into GFP$^+$ and GFP$^-$ fractions on a flow cytometer (UPCI Flow Cytometry Facility, Pittsburgh, PA; Supplementary Table 1 and Supplementary Figure 1). RNA was isolated by column purification (Stratagene, La Jolla, CA) and assayed on an Agilent BioAnalyzer (Agilent Technologies, Santa Clara, CA). Only samples with intact, distinct ribosomal peaks were chosen for further analysis.

Hybridization of Affymetrix GeneChips

For each time point, we selected 3 sets of GFP$^+$ and GFP$^-$ RNA from a single flow sort, amplified 50 ng of each sample with Ovation (NuGEN Technologies, San Carlos, CA), and hybridized them to Affymetrix (Santa Clara, CA) Zebrafish GeneChips (UPCI Clinical Genomics Facility, Pittsburgh, PA), as per manufacturer's instructions.

We extracted expression values with the use of robust multiarray averaging (http://rmaexpress.bmbolstad.com/). Expression values were log transformed, centered by median subtraction, and scaled by dividing with the standard deviation. Microarrays were not individually weighted. All primary microarray data and a table of all expression values after normalization are available (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12189). A linear model was fitted, and genes were identified as differentially expressed at a threshold of $P < .05$ after false-positive correction with the use of the Benjamini–Hochberg algorithm.[19]

Annotation of Affymetrix Zebrafish GeneChip With the Use of a BLAST Strategy

To use mammalian gene interaction databases, we improved the available annotation of the Affymetrix Zebrafish GeneChip. We linked as many probe sets as possible with zebrafish RefSeq records with the use of information provided by Affymetrix and Ensembl or through nucleotide BLAST for unassigned probe sets (GenBank nt Database from April 23, 2007; Ensembl Zebrafish Genome Build Zv6 from April 27, 2007). We extracted zebrafish annotation information. To define protein domains and function, we screened corresponding peptide sequences for PFAM protein domains using HMMer (Pfam-ls and Pfam-fs databases v.21 from October 13, 2006; http://pfam.janelia.org and http://hmmer.janelia.org)[20] and collated GO annotation from GenBank, ZFIN, and Pfam. Finally, we screened human protein sequences with zebrafish sequences with the use of TBLASTX (Supplementary Table 2).

Whole Mount In Situ Hybridization of Zebrafish Embryos

See Supplementary Table 3 for a list of genes, accession numbers, ZFIN IDs, primer sequences, and antibodies for Figures 1, 2, and 3 and Supplementary Figures 2 and 8. In situs and immunohistochemistry were carried out as described.[11,21]

PI3K Inhibitor Treatment

Dimethyl sulfoxide was added to 0.05% either alone or with 10 μmol/L LY294002 (Sigma–Aldrich, St Louis, MO) from 2 to 5 dpf. Embryos were analyzed at 6 dpf by bright-field microscopy or embedding in JB4 plastic (Polysciences, Warrington, PA), sectioning, and standard H&E staining.

Results

GFP Transgenic Fish, RNA Processing, and Initial Microarray Analysis

As the basis of our study, we selected the transgenic fish line *Tg(XlEef1a1:GFP)*^{s854}, gutGFP.[2–5] After