Molecular characterization of mouse gastric epithelial progenitor cells

Jason C. Mills*, Niklas Andersson*, Chieu V. Hong*, Thaddeus S. Stappenbeck*, and Jeffrey I. Gordon‡

Departments of *Molecular Biology and Pharmacology and ‡Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110

Contributed by Jeffrey I. Gordon, September 20, 2002

The adult mouse gastric epithelium undergoes continuous renewal in discrete anatomic units. Lineage tracing studies have previously disclosed the morphologic features of gastric epithelial lineage progenitors (GEPs), including those of the presumptive multipotent stem cell. However, their molecular features have not been defined. Here, we present the results of an analysis of genes and pathways expressed in these cells. One hundred forty-seven transcripts enriched in GEPs were identified using an approach that did not require physical disruption of the stem cell niche. Real-time quantitative RT-PCR studies of laser capture microdissected cells retrieved from this niche confirmed enriched expression of a selected set of genes from the GEP list. An algorithm that allows quantitative comparisons of the functional relatedness of automatically annotated expression profiles showed that the GEP profile is similar to a dataset of genes that defines mouse hematopoietic stem cells, and distinct from the profiles of two differentiated GEP descendant lineages (parietal and zymogenic cell). Overall, our analysis revealed that growth factor response pathways are prominent in GEPs, with insulin-like growth factor appearing to play a key role. A substantial fraction of GEP transcripts encode products required for mRNA processing and cytoplasmic localization, including numerous homologs of Drosophila genes (e.g., Y14, staufen, mago nashi) needed for axis formation during oogenesis. mRNA targeting proteins may help these epithelial progenitors establish differential communications with neighboring cells in their niche.

stem cells | growth factor signaling | protein turnover | mRNA localization | bioinformatics

The mammalian gastrointestinal tract is lined by an epithelium that is constantly renewed. Although multipotent stem cells are known to fuel this renewal, the molecular properties of these cells are poorly understood.

Tritiated thymidine/EM autoradiographic lineage tracing studies have delineated the morphological features of the stem cell niche in the adult mouse stomach (1). The glandular epithelium is composed of tubular invaginations termed gastric units. In the corpus (central region) of the stomach, each unit contains an average of ~200 cells, representing three predominant lineages: pit, parietal, and zymogenic. The multipotent stem cell (undifferentiated granule-free progenitor) resides in the unit’s isthmus (ref. 1; Fig. L1). One of its committed daughters, the granule-free prepit cell precursor, produces mucus-secreting pit cells, which differentiate as they climb from the isthmus to the orifice of the unit (2). Another daughter, the granule-free preneck cell precursor, gives rise to pepsinogen-producing neck cells, which differentiate as they climb from the isthmus to the orifice of the unit (2). Unlike the pit and zymogenic lineages, acid-producing parietal cells (PCs) differentiate within the isthmus from granule-free preparietal cell lineages. PC precursors establish differential communications with neighboring cells in their niche.

Materials and Methods

Mice. Conventionally raised FVB/N tox176 transgenic mice (6) were maintained in microisolation cages in a specified pathogen-free state. Nontransgenic, germ-free FVB/N mice were raised in plastic gnotobiotic isolators (8).

GeneChip Comparisons. Whole stomachs were excised and RNA extracted (midi-RNeasy kit; Qiagen, Valencia, CA). Duplicate cRNA targets were prepared from pooled RNAs (see below). Each cRNA was hybridized to a set of Mu11K GeneChips (Affymetrix, Santa Clara, CA). Overall fluorescence across each GeneChip set was scaled to a target intensity of 150 and pairwise comparisons performed using Affymetrix MICROARRAY SUITE software (V.4.0).

SYBR-Green-Based Real-Time Quantitative RT-PCR (qRT-PCR). Assays were performed in triplicate as described (9), using the gene-specific primers listed in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Normalization was to 18S rRNA (ΔΔCT method; ref. 9).

Navigated Laser Capture Microdissection (n-LCM). Stomachs were removed from tox176 mice (n = 4), flushed with PBS, and divided in half along the cephalocaudal axis, and each half-stomach was rinsed with OCT compound. Half-stomachs were placed in a cryomold, overlaid with OCT compound, and frozen (Cytocool II, Richard-Allan Scientific, Kalamazoo, MI).

Abbreviations: AAA, Anguilla anguilla agglutinin; FR, fractional representation; GEP, gastric epithelial lineage progenitor; GO, Gene Ontology; GSII, Griffonia simplicifolia II; HSC, hematopoietic stem cell; IGF, insulin-like growth factor; PC, parietal cell; PCNA, proliferating cell nuclear antigen; qRT-PCR, quantitative RT-PCR.

*C. Mills is now at The Rockefeller University, New York, NY 10021. **N. Andersson is now at the University of Massachusetts Medical School, Worcester, MA 01605. **C. V. Hong is now at Cedars-Sinai Medical Center, Los Angeles, CA 90048. **T. S. Stappenbeck is now at the University of Colorado School of Medicine, Denver, CO 80262. **J. I. Gordon is now at The Broad Institute, Cambridge, MA 02142.

To whom correspondence should be addressed. E-mail: jgordon@molecool.wustl.edu.
age, GEPs account for 10% of the total gastric unit epithelial cell census, and by 20 weeks 25% (~20-fold higher than in age-matched normal littermates; Fig. 1 C and D; ref. 5). Guided by the strategy outlined in the Introduction, we isolated RNA from the intact stomachs of conventionally raised 16-week-old tox176 mice (equal amounts of RNA pooled from five stomachs), age-matched normal germ-free mice (n = 4), and normal embryonic day 18 animals (n = 31). Because tox176 mice lose the acid barrier to microbial colonization, they inevitably develop a mild diffuse chronic gastritis associated with bacterial overgrowth. Because a comparison of gene expression in tox176 versus normal adult germ-free stomachs would likely yield a dataset substantially enriched for GEP-associated mRNAs but “contaminated” by immune response genes, we referenced the “tox176 versus normal” dataset to the “E18 versus normal” dataset (where E18 represents a GEP-enriched, gastritis-free state).

Duplicate cRNA targets, independently generated from each RNA, were used to probe GeneChips representing ~11,000 mouse genes and EST clusters. Genes with enhanced expression in duplicate comparisons between tox176 and normal stomachs and in duplicate comparisons between E18 and normal stomachs were culled, yielding a list of 147 genes and 6 uncharacterized ESTs with quadruplicate “Increased” calls by GeneChip software (see Table 2, which is published as supporting information on the PNAS web site).

**Initial Validation of the GEP Dataset.** The three RNAs used for the GeneChip comparisons were used as templates for qRT-PCR analysis of the expression of seven genes, selected from the GEP dataset based on their known roles in regulating proliferation, differentiation, and polarity in other systems. They encoded proliferating cell nuclear antigen (PCNA; ref. 10), retinoblastoma binding protein 7 (together with RabAP48 increases histone acetyltransferase activity; ref. 11), colony stimulating factor 1 receptor (tyrosine kinase receptor important for hematopoietic and mammary epithelial development; ref. 12), ephrin receptor B4 (receptor tyrosine kinase expressed in undifferentiated hematopoietic and mammary epithelial cells; refs. 13 and 14), mago-m (homolog of Drosophila mago nashi, involved in mRNA localization during fly oogenesis; refs. 15 and 16), lactotransferrin (multiple functions including regulation of IGF bioavailability; ref. 17), and annexin-A1 (participates in endosomal trafficking; ref. 18). The qRT-PCR analysis confirmed that the level of each transcript was 2- to 9-fold higher in intact stomachs with enriched GEP populations (tox176 and/or E18) compared with normal adult germ-free stomachs (see Fig. 5, which is published as supporting information on the PNAS web site).

We next used antibodies specific for PCNA and annexin, as well as three other members of the GEP dataset, to show that all five proteins were present at higher levels in intact tox176 compared with age-matched normal stomachs (see Fig. 5). Light and electron microscopic immunohistochemical studies of normal and tox176 mice confirmed PCNA expression in GEPs (Figs. 1B and 2A).

**qRT-PCR Studies of Navigated Laser Capture Microdissected (n-LCM) GEPs.** To further verify enhanced expression of members of the dataset within the expanded tox176 GEP population, we combined qRT-PCR with a form of LCM we have termed navigated LCM (19). Immunostaining protocols requiring more than a few minutes to complete jeopardize recovery of intact mRNA. Therefore, serial cryosections were prepared from the central third of 16-week-old tox176 stomachs. Odd-numbered sections were stained with AAA to identify pit cell-specific glycans (20), GSII to visualize neck cell-specific glycans, and antitoxin to PCNA to mark GEPs (Figs. 2A and 2B). The multilabeled sections were used as image templates to guide (navigate) dissection of
Functional Comparison of the GEP Expression Profile with Profiles Obtained from Two Descendant Lineages and Mouse Hematopoietic Stem Cells. Because there had been no previous molecular characterization of GEPs, we could not validate the GEP dataset by comparing it to a preexisting list of mRNAs. Therefore, we compared the functional features of the entire GEP dataset with those of gene expression profiles obtained from mouse hematopoietic stem cells (HSCs) and from two differentially GEP descendants, parietal and zymogenic cells. To do so, we needed a tool that could automatically classify an entire dataset, irrespective of species of origin or method of data generation, so that each dataset could be viewed as the sum of its component parts, and its similarity to other lists determined independent of their length.

Terms defined by the Gene Ontology (GO) Consortium (www.geneontology.org/) provided a language for performing this automated annotation and functional comparison. Using a series of algorithms, we assigned GO terms to each gene in each dataset, analyzed the distribution of GO terms across the entire gene list, and determined the fractional representation (FR) of each term. We defined FR as the number of genes with a given GO term in a given list relative to the total number of genes in that list with assigned GO terms.

We had postulated that comparison of tox176 stomachs with normal germ-free mouse stomachs would lead to a progenitor cell expression profile “contaminated” with immune cell-derived transcripts. Thus, we tested our GO-based classification system by comparing the FR of immune/host defense functions in three datasets: the tox176 to normal germ-free stomach comparison (500 genes), the E18 to normal germ-free stomach comparison (580 genes), and the final, triangulated 147-gene GEP dataset. Table 3 (which is published as supporting information on the PNAS web site) shows that our triangulation strategy led to near elimination of immune/defense-related genes.

Fig. 3A compares the FR of the six most frequent GO terms in the GEP database with their FR in a previously published dataset of 767 genes expressed in fetal mouse liver-based HSCs (22), and in two GeneChip-derived datasets [one a list of 231 transcripts enriched in differentiated PCs relative to all other gastric mucosal cell types (23), the other a list of 114 zymogenic cell-enriched mRNAs (N.A., J.C.M., and J.I.G., unpublished work)]. In all six cases, the FR of the GO term in the GEP dataset was closest to that in the HSC dataset; e.g., the term “transcription regulation” was the 5th and 6th most frequently represented term in the GEP and HSC databases (10% and 13%, respectively), but only the 29th most frequent term (3%) in the zymogenic dataset, and 97th in the PC dataset (1%).

Fig. 3B compares the FR of the six most common GO terms obtained in the PC dataset (PC1) with their corresponding representation in the GEP dataset and in another, 259-member GeneChip-derived dataset (PC2) generated from PCs that had been purified by elutriation as opposed to lectin panning (N.A., J.C.M., and J.I.G., unpublished work). The results reveal the functional similarity between the two differentiated PC datasets, and their distinctness from the GEP dataset; e.g., three of the most common PC GO terms (mitochondrion, metabolism, and glycolysis) were not among the top 15 terms in the GEP dataset, or among the top 12 terms in the HSC list.

Together, these results demonstrate that the automated classification scheme can be used to quantify functional similarities and differences between entire gene expression profiles (“profile surfing”). They verify that the “triangulation” strategy identified a gene expression profile in GEPs that is functionally distinct from their differentiated descendants but similar to HSCs.

Further Characterization of the Functional Features of GEPs. Overview. Our GO-based functional analysis prompted a more detailed inspection of the GEP dataset to identify pathways that would...
this notion, we used n-LCM and qRT-PCR to assay for IGF1 and the possibility of autocrine IGF signaling in GEPs. To test factors signal through their receptors to activate PI-3 kinase, members, c-Myb (activated by Akt; ref. 28) and Cdk2, while increasing its promotion of cyclinD1-cdk4 assembly; ref. 24) and the bHLH antagonist Id2 (25) are all in the GEP dataset. Further information about the hierarchical GO-term classification scheme can be obtained from the Gene Ontology web site (http://www.geneontology.org/).

give us further insights about the regulation of progenitor proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche. Thirty-five members of the 147 member GEP list regulate proliferation and differentiation in the isthmal stem cell niche.
encoding proteins that help direct and integrate nuclear mRNA processing, nuclear export, and cytoplasmic localization (Fig. 4). These mRNAs encode polyA binding protein 1, four proteins associated with the spliceosome [DEAD-box helicase 15, SmD1, SmE, and the non-POU domain containing octamer binding protein (nonO)] and four hnRNPs (RBP3, hnRNPA/B, hnRNPC, and hnRNPG), plus three serine/arginine-rich (SR) proteins (9G8, Tra2, and neuro-salient). GEP-enriched transcripts also specify three nuclear pore proteins that participate in RNA export from the nucleus (exportin 1, Ran binding protein 1, and the mouse ortholog of yeast Sec13p). Microtubule-associated proteins are involved in mRNA segregation (44). The GEP dataset contains tubulin β5, the tubulin-bound motor proteins kinesin heavy chain 5 (KIF5B) and Tetex-1 (one of three dynein light chains), plus stathmin (a cell cycle-regulated, tubulin-binding protein). Actin microfilaments have also been implicated in RNA targeting; the GEP member eEF1α binds both mRNA and F-actin to localize translation in specific subcellular domains (45). GEPs also express several components of the translation initiation apparatus [eIF2γ, eIF3 subunit 6, eIF5, and methionylaminopeptidase 2 (removes initiator Met residues and blocks phosphorylation of eIF2α, promoting translation; ref. 46)].

Interestingly, null mutations of Drosophila genes involved in each of the mRNA processing/localization steps described above impede establishment of polarity in developing oocytes. These mutations involve half pint (encodes a spliceosome component; ref. 47), squid (hnRNPA/B-like; ref. 48), tsunagi (Y14 ortholog; ref. 49), and mago nashi. Similar defects are produced by mutations in genes encoding components of the translation apparatus [vasa (eIF4A homolog; ref. 50), aubergine (eIF2C; ref. 51)], and that regulate microtubule-mRNA interactions [staufen, dKhc1 (kinesin heavy chain; ref. 44), and dDlc (dynein light chain; ref. 52)]. In addition, mutations of genes such as cornicklon, which targets proteins to specific cellular surfaces, lead to aberrant axis formation.

To further establish a link between mRNA processing/translation and GEP biology, we used n-LCM to harvest GEPs and pit cells from tox176 gastric units and qRT-PCR to assay for transcripts specifying two recently characterized homologs of fly genes not present in our GEP dataset. The transcript encoding Y14, which, as noted above, binds nascent mRNAs with mago-m, is increased 8-fold in GEPs. Staufen 1, a conserved double-stranded RNA binding protein that coordinates interaction of certain mRNAs with microtubules for proper transcript localization in neurons and oocytes (53), is enriched 6-fold (Fig. 2F).

In the Drosophila oocyte, some mRNA transcripts are targeted anteriorly, and others posteriorly (54). Presumably, the direction of specific mRNA placement reflects the nature of the expressed cellular complement of mRNA-associated targeting proteins. Stem cells in different tissues likely have differing needs for orienting mRNAs relative to other cellular components of their niche. To test the hypothesis that the pattern of expression of RNA localizing genes might be different in progenitor populations distributed at different points along the length of the normal mouse gut, we combined LCM with qRT-PCR to assay for mago-m, Y14, and staufen expression in small intestinal and colonic epithelial progenitor niches located at or near the base of crypts of Lieberkühn (19, 55, 56). Epithelial cells were microdissected from each progenitor niche, and from adjacent compartments containing their differentiated descendants (the villus epithelium in the case of the small intestine, the hexagonal surface epithelial cuff demarcating the orifice of each crypt in the case of the colon; 10,000 cells harvested per compartment per mouse; n = 3 mice). Mago-m mRNA levels were 5-fold higher in the small intestinal crypt base compared with villus epithelium, and 6-fold higher in the colonic crypt base versus surface cuff. Y14 mRNA levels were 3- and 2-fold higher, whereas staufen 1 mRNA, which is present in GEPs, was undetectable in both small intestine and colon.

BLASTn searches against the entire HSC EST database (http://stemcell.princeton.edu/) identified Y14, but not mago-m or staufen, suggesting that HSCs may exhibit a pattern of mRNA localization distinct from gut progenitors.
targeting proteins may help distinguish normal from preneoplastic or fully transformed gastric epithelial progenitors.

We thank Sabrina Wagoner and Jaime Dant for technical assistance, David O’Donnell and Maria Karlsson for maintaining mice, Jung Oh for advice with qRT-PCR assays, and Diane Redmond for computer graphics. This work was supported by National Institutes of Health Grant DK58529. J.C.M. is a Howard Hughes Medical Institute physician-scientist postdoctoral fellow.