Molecular Properties of Adult Mouse Gastric and Intestinal Epithelial Progenitors in Their Niches*§

Received for publication, November 10, 2005, and in revised form, January 25, 2006 Published, JBC Papers in Press, February 7, 2006, DOI 10.1074/jbc.M512118200

Marios Giannakis‡§, Thaddeus S. Stappenbeck‡§†, Jason C. Mills‡§†, Douglas G. Leip‡§†, Michael Lovett‡, Sandra W. Clifton‡, Joseph E. Ippolito‡§, Jarret I. Glasscock‡, Manimozhiyan Arumugam**, Michael R. Brent‡§†, and Jeffrey I. Gordon‡§†*

From the ‡§°Center for Genome Sciences, Departments of ‡§ Molecular Biology and Pharmacology, ‡§ Pathology and Immunology, ‡§¶ Laboratory for Computational Genomics, Washington University, St. Louis, Missouri 63108

We have sequenced 36,641 expressed sequence tags from laser capture microdissected adult mouse gastric and small intestinal epithelial progenitors, obtaining 4031 and 3324 unique transcripts, respectively. Using Gene Ontology (GO) terms, each data set was compared with cDNA libraries from intact adult stomach and small intestine. Genes in GO categories enriched in progenitors were filtered against genes in GO categories represented in hematopoietic, neural, and embryonic stem cell transcriptomes and mapped onto transcription factor networks, plus canonical signal transduction and metabolic pathways. Wnt/β-catenin, phosphoinositide-3/Akt kinase, insulin-like growth factor-1, vascular endothelial growth factor, integrin, and γ-aminobutyric acid receptor signaling cascades, plus glycerolipid, fatty acid, and amino acid metabolic pathways are among those prominently represented in adult gut progenitors. The results reveal shared as well as distinctive features of adult gut stem cells when compared with other stem cell populations.

The adult gastrointestinal tracts of humans and mice contain large populations of multipotent stem cells that produce an impressive number of descendant epithelial cells each day: ~70 billion in humans (~0.25 kg) and 200 million in mice (1, 2). The morphologic features of adult mouse gut epithelial lineage progenitors have been characterized using tritiated thymidine labeling and electron microscopic autoradiography (3, 4). In the stomach, the stem cell niche is positioned in the mid-portion (isthmus) of tubular mucosal invaginations known as gastric units, whereas small intestinal and colonic stem cells reside at or near the base of crypts of Lieberkuhn (4, 5). Analyses of genetic mosaic mice indicate that adult gastric units and intestinal crypts are monoclonal; all epithelial cells in each of these anatomically distinct structures are apparently derived from a single ancestor occupying the highest position in the stem cell hierarchy (6–8).

Knock-out mice have provided direct evidence that the Wnt pathway is important for maintaining a dividing small intestinal stem cell population, and that Notch signaling is involved in specification of its descendant secretary cell lineages (9–11). Nonetheless, we know relatively little about the signaling and metabolic pathways that are active in adult gastric and intestinal stem cells within their niches, the degree to which these progenitors share properties with one another, and how they compare with other characterized adult or embryonic stem cell populations. To address this issue, we have sequenced cDNA libraries prepared from laser capture microdissected adult gastric and small intestinal epithelial progenitors (GEPs); SiEPIPs, taking advantage of two gnotobiotic transgenic mouse models where these cells are represented in increased numbers in discrete regions of gastric units and crypts of Lieberkuhn. The resulting data sets were analyzed using a combination of GO functional annotations and a variety of software tools that allow analysis of the statistical significance of GO term enrichments and of the observed representation of components of canonical signaling and metabolic pathways.

**EXPERIMENTAL PROCEDURES**

Animals—Germ-free FVB/N mice that express an attenuated diphtheria toxin A fragment (tox176) under the control of nucleotides -1,035 to +24 of mouse Atp4b (noncatalytic β-subunit of H+K+-ATPase; Ref. 12) or nucleotides -6500 to +34 of mouse Defcr2 (cyst-2; Ref. 13) were maintained in plastic gnotobiotic isolators (14) under a strict 12-h light cycle. Animals were given an autoclaved chow diet (B & K Universal, East Yorkshire, UK) ad libitum.

Wild-type C57BL/6J mice (The Jackson Laboratory, Bar Harbor, ME) were maintained free of specified pathogens in micro-isolator cages within a barrier facility and fed the same autoclaved chow as their germ-free counterparts. These conventionally raised C57BL/6J mice (10 weeks old) were given a 2.5% aqueous solution of dextran sodium sulfate (DSS; TDB Consultancy AB, Uppsala, Sweden) for 7 days as drinking water to induce a regenerative epithelial response in colonic crypts (15). All experiments involving animals were performed using protocols approved by the Washington University Animal Studies Committee.

**Laser Capture Microdissection (LCM)**—General methods for preparing the mouse gut for cryosectioning and LCM are described in Ref. 16. Specific methods for navigated LCM of GEPs from the expanded isthmus of gastric units in the corpus region of the stomachs of 14-week-old male germ-free Atpb4-tox176 mice are described in Ref. 17. Protocols for LCM of SiEPIPs from the base of crypts located at the

The abbreviations used are: GEP, gastric epithelial progenitor; EST, expressed sequence tag; GO, Gene Ontology; SIPI, small intestinal epithelial progenitor; Atp4b, noncatalytic β-subunit of H+K+-ATPase; Defcr2, cyst-2; DSS, dextran sodium sulfate; LCM, laser capture microdissection; IPA, ingenuity pathways analysis; BrdUrd, bromodeoxyuridine; Dcamk1, double calmodulin-dependent protein kinase-like-1; Mapk14, mitogen-activated protein kinase 14; CgA, chromogranin A; GSII, Griffonia simplicifolia II; HSC, hematopoietic stem cell; ESC, embryonic stem cell; NSC, neural stem cell; GABA<sub>A</sub>, γ-aminobutyric acid, type A; VEGF, vascular endothelial growth factor; PBS, phosphate-buffered saline; Ddc, dicer decarboxylase; AoC3, amine oxidase, copper containing 3; Aldh9A1, aldehyde dehydrogenase 9 family, member A1; Gad1, glutamic acid decarboxylase 1; IPA, Ingenuity Pathways Analysis.

*This work was supported by National Institutes of Health Grant DK63483. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S32 and Figs. S1–S8.

‡These authors contributed equally to this work.

§To whom correspondence should be addressed: Center for Genome Sciences, 4444 Forest Park Blvd., Campus Box 8510, St. Louis, MO 63108. Tel.: 314-362-7243; Fax: 314-362-7047; E-mail: jgordon@wustl.edu.
junction between the middle and distal thirds of the small intestines of 4-week-old Defcr2-tox176 mice are provided in Ref. 13. Epithelial cells were also retrieved by LCM from the lower thirds of crypts located in the descending colons of DSS-treated C57Bl/6 mice (15). Cells were harvested using a PixCell Ile LCM System and CapSure HS LCM caps (Arcturus Bioscience). Total cellular RNA was extracted (10,000 cells/LCM population; PicoPure RNA isolation kit; Arcturus Bioscience), and its quality was evaluated using an Agilent 2100 bioanalyzer.

**Construction of cDNA Libraries**—Detailed protocols and explanatory figures with the sequences of primers used for cDNA synthesis and library construction are available from the authors and in Ref. 17. cDNA was synthesized from 5 ng of RNA on oligo(dT)-linked paramagnetic beads (Dynal Norway) in the presence of a modified “SMART” oligonucleotide (Clontech). Oligonucleotide dimers were removed by digestion with NotI. Bead-linked cDNA was amplified by PCR (four cycles) using the SMART primer and a hybrid primer consisting of oligo(dT) linked to an external and more specific priming site. The original bead-linked cDNAs were captured magnetically and stored. Amplified cDNAs in the supernatant were subsequently amplified for an additional seven cycles of PCR using the more specific external priming sites. cDNA products were passed through a Sephadex G-50 column (Amersham Biosciences). A final round of PCR amplification (five cycles) was performed after adding ligation-independent linkers compatible with the uracil deglycosylase system (Invitrogen). The final cDNA products were size-selected (SizeSep 400 Spun Column; Amersham Biosciences) to isolate fragments >300 bp. 10–20 ng of this cDNA were used in a standard annealing mix with the UDG pMAM1 vector (Invitrogen). Aliquots of the ligation-free annealing mix were transformed directly into Escherichia coli DH10B. The GEP, SiEP, and colon crypt transit amplifying cell libraries each contained >5 × 10^5 independent clones with mean insert lengths of 0.7 kb.

**Alignments of ESTs to the Mouse Genome**—ESTs were aligned to the mm5 assembly of the mouse genome from University of California Santa Cruz (UCSC) with BLAT (gClient version 27 × 1 obtained from UCSC) using minScore = 50 (minimum score, BLAT default scoring scheme) and minIdentity = 95 (minimum percent identity). The alignments were then filtered using minScore = 50 (on a match = 1, mismatch = −1, and gapOpen = −4 scoring scheme), minMatch = 50 (minimum number of matching bases), and minAli = 200 (minimum ratio of query in parts per thousand). The top scoring alignment for each EST was chosen from this filtered set of alignments, along with any alignment of the EST that scored within 5% of the top scoring alignment. These alignments were subsequently filtered to include only those that contained introns. Gene symbols obtained from the EST alignments to mm5 and consensus gene clustering from Unigene can be found online at genome.wustl.edu/GSCGAP.

Similar data sets were obtained when transcripts from the GEP and SiEP libraries were aligned to mm5 using BLASTN (18). A subsequent, more refined alignment with EST_GENOME (19) provided confident intron-exon boundaries. Transcripts that crossed splice boundaries were annotated as likely intron retention events; each group of consistent splice patterns was put into a given cluster, and each cluster, along with its individual members and their library association, was annotated.

**Ingenuity Pathways Analysis (IPA)**—We analyzed our data sets using the following procedures. Once logged onto the IPA system, we clicked on “File” menu, then “New,” and then “Analysis.” We then clicked “Upload” and selected the data set to be analyzed. For the analysis of canonical signaling and metabolic pathways active in GEPs and SiEps, the input was the Entrez Gene identifiers of an EST library (e.g. GEPs) with the genes in a data set of interest (e.g. the uniquely GO-enriched genes expressed in GEPs) designated as focus genes. Because we had EST rather than DNA microarray data, we set the expression value to zero for all genes. We clicked on “Create Analysis” followed by “Run Analysis.” When the analysis was completed by IPA software, we clicked on “Canonical Pathways” and “Customize Chart,” and under “Select Categories to Display” we chose both “Metabolic Pathways” and “Signaling Pathways.” Under “Select Sort Order” we ranked by “Significance” and completed the procedure by clicking “Apply.” For the transcriptional networks in GEPs and SiEps, the input was the transcription-related GO-enriched GEP or SiEP Entrez Gene identifiers. When the analysis was completed by IPA software, we clicked on “Networks,” selected the top two scoring networks, and clicked “View Networks.” The results presented in this study are based on IPA Knowledge Base content statistics from May, 2005.

**Multiplex Immunohistochemistry**—Sections (7 µm thick) were prepared from paraffin-embedded, Bouin’s fixed stomach (corpus region) and zinc-formalin fixed small intestine harvested from normal, 12-week-old conventionally raised FVB/N mice and/or comparably aged germ-free Atpb4-tox176 animals. Mice received an intraperitoneal injection of bromodeoxyuridine (BrdUrd) 1.5 h prior to sacrifice. Sections were sequentially passed through xylene (three washes; 5 min each), 100% isopropyl alcohol (three washes; 5 min each), and deionized water (5 min). Antigen retrieval was performed by boiling slides in 10 mM citrate buffer (pH 6.0) for 10 min, followed by cooling to 23 °C, three washes in PBS (5 min each), and incubation in blocking buffer (1% bovine serum albumin, 0.3% Triton X-100 in PBS; 30 min). Sections were subsequently incubated overnight at 4 °C with the following primary antibodies and/or biotinylated lectins: (i) rabbit anti-mouse double cortin and calcium/calmodulin-dependent protein kinase-like-1 (Dcamk1; final dilution in blocking buffer = 1:200; a gift from Christopher Walsh, Harvard Medical School); (ii) rabbit anti-human mitogen activated protein kinase 14 (Mapk14/p38; 1:200, Santa Cruz Biotechnology); (iii) goat anti-BrdUrd (1:500; see Refs. 12 and 13); (iv) rabbit anti-bovine chromogranin A (CgA; 1:10; ImmunoStar; conjugated to fluorophore with the Zenon tricolor rabbit IgG labeling kit from Molecular Probes); (v) Griffonia simplificacoil II (GSI1) lectin (1:200, specific for neck cells in the FVB/N inbred strain); and (vi) Anguilla anguilla agglutinin (1:200, specific for pit cells). After an overnight incubation at 4 °C, sections were washed with PBS (three cycles; 5 min each), and antigen-antibody complexes and bound biotinylated lectins were detected with Alexa Fluor-, cyanine (Cy)-3-, or fluorescein isothiocyanate-conjugated secondary antibodies or streptavidin (Molecular Probes and Jackson ImmunoResearch). Sections were washed in PBS, incubated with bisbenzamide (Hoechst 33342B; 1 µg/ml) to stain nuclei, and then examined and photographed using a Zeiss Axiolab 200M microscope.

**RESULTS AND DISCUSSION**

**Sequencing cDNA Libraries from Laser Capture Microdissected Gut Epithelial Progenitors**—GEPs, which include cells with the EM morphologic features of the presumptive multipotent stem cell and its descendant oligo-potential pre-parietal, pre-pit, and pre-neck cell lineage progenitors, were retrieved by LCM of cryosections prepared from the middle portion (corpus) of the stomachs of germ-free transgenic mice with an engineered, attenuated diphtheria toxin A fragment (tox176)-mediated ablation of their parietal cells (12, 20) (n = three 16-week-old animals; ~10 cells microdissected per isthmus; total of ~2,000 LCM gastric units). Acid-producing parietal cells terminally differentiate within the isthmal stem cell niche before they migrate to

---

**Mouse Gastric and Intestinal Epithelial Progenitors**

---

**JOURNAL OF BIOLOGICAL CHEMISTRY**

---

APRIL 21, 2006 • VOLUME 281 • NUMBER 16 11293
the upper (pit) and lower (base) domains of the gastric unit (21). tox176-mediated parietal cell ablation results in augmented GEP proliferation, with a progressive increase in their fractional representation to ~10% of all epithelial cells in gastric units by 16 weeks of age, as judged by transmission EM (22). Germ-free Atpb4-tox176 mice were used because loss of the acid-barrier to colonization produces bacterial overgrowth in the stomach and gastritis (20, 22).

SiEPs were recovered from germ-free Defcr2-tox176 transgenic mice with a Paneth cell ablation (13). This epithelial lineage is a key component of the innate immune system of the small intestine and is the only one descended from the multipotent stem cell that completes its differentiation program at the crypt base (23). tox176-directed ablation of Paneth cells results in a consolidation of SiEPs at the crypt base without affecting their proliferative activity, or the differentiation programs of the three other small intestinal epithelial lineages (13). The five most basal crypt epithelial cells in Defcr2-tox176 mice were harvested by LCM of cryosections prepared from the junction between the middle and distal thirds of the small intestine \( n = \text{three Defcr2-tox176 mice; total of} \approx 5,000 \text{ LCM crypts} \).

A generally applicable method was employed to generate GEP and SiEP cDNA libraries starting with 5 ng of total cellular RNA from each LCM progenitor population. This method minimizes skewing of the relative abundance of expressed transcripts. Alignments of 36,641 sequenced ESTs from the two libraries to Unigene and the mm5 build of the mouse genome yielded a data set of 4031 genes expressed in GEPs and 3324 genes in SiEPs (see supplemental Table S1; a complete list is available at genome.wustl.edu/GSGAP/ see supplemental Table S2 for a list of 52 identified alternatively spliced variants).

**Identification of “Biological Process” GO Terms Enriched in GEPs and SiEPs Relative to All Gastric and Small Intestinal Cells** — We used GO terms and GoSurfer (24) to compare each adult gut epithelial progenitor data set with previously sequenced cDNA libraries generated from the intact stomachs and small intestines of conventionally raised, normal adult mice (Fig. 1). The two stomach libraries contained a total of 3303 unique Unigene clusters, whereas the two small intestinal libraries contained 2359 Unigene clusters (supplemental Table S3). GoSurfer takes one or two gene lists as input, finds GO terms associated with the genes, and visualizes them as one of three hierarchical trees corresponding to the broad GO categories of “biological process,” “molecular function,” and “cellular component.” GoSurfer can also compare two gene lists to identify GO terms enriched to a statistically significant degree in one data set versus the other (e.g. supplemental Fig. S1). The following analysis focused on GO terms and pathways that were more prominent in progenitor cells compared with non-stem cell populations.

Sixty five biological process GO terms, representing 1767 genes (transcripts), were enriched in the GEP data base over the intact adult stomach cDNA library data set \( q \text{ value cutoff of} 0.1 \). These 1767 genes, subsumed under enriched GO terms, were termed “GO-enriched.” An analogous GoSurfer comparison of the SiEP data base with the intact adult mouse small intestine cDNA libraries yielded 62 GO terms, rep-
Representing 1824 genes (q value cutoff of 0.1) (see step 1 in Fig. 1 plus supplemental Tables S4 and S5). The 65 GEP-enriched and 62 SiEP-enriched GO terms are listed in supplemental Tables S6 and S7, respectively, and include the biological processes of cell cycle, DNA replication, cell proliferation, morphogenesis, pattern specification, and ubiquitination. A number of transcription-related GO terms were enriched in the two gut epithelial progenitor populations (see supplemental Tables S8 and S9 for these GO terms and their associated genes, including, for example, Notch 1–4. Note that when molecular function rather than biological process GO terms were used for GoSurfer comparisons, 96% of the transcription-related molecular function GO-enriched genes were identical to the transcription-related biological process GO-enriched genes listed in supplemental Tables S8 and S9).

The genes representing these transcription-related GO terms in GEPs and SiEPs were placed onto transcription factor networks using the IPA software tool. This tool utilizes a knowledge base of over one million known functional relationships among proteins. One of the two top scoring networks in GEPs is depicted in Fig. 2A and includes factors involved in Wnt/β-catenin, phosphoinositide 3-kinase/Akt kinase (PI3K/Akt), transforming growth factor-β (Tgf-β), and insulin-like growth factor-1 (Igf-1) signaling. Similarly, one of the two top scoring SiEP-associated networks is shown in Fig. 2B and includes factors that participate in JAK/Stat (Janus kinase-signal transducer and activator of transcription signaling) and cell cycle (G1/S checkpoint) regulation (see supplemental Figs. S2A and S2B for the other top scoring networks in GEPs and SiEPs).

Identification of GO-enriched Genes Whose Expression Is Unique to Gut Epithelial Progenitors or Shared with Other Stem Cell Populations—We performed a second step to distinguish GO-enriched genes unique to GEPs or SiEPs from those shared with other non-gut stem cell populations. This analysis employed all published sequenced cDNA libraries of >1000 ESTs from mouse hematopoietic stem cells (HSCs), embryonic stem cells (ESCs), and neural stem cells (NSCs). For HSCs, we combined seven libraries totaling 5018 Unigene clusters with 4191 genes resulting from our alignment to mm5 of a hematopoietic EST data set, available from the Stem Cell Data Base (see SCDb; a joint project of the labs of I. R. Lemischka, K. A. Moore, and C. Stoeckert). For the ESC data set, we incorporated three cDNA libraries with 5259 unique Unigene clusters. For NSCs we used two libraries with 5093 Unigene clusters (supplemental Table S3). Genes producing transcripts in HSCs, ESCs, and NSCs were compared with the adult whole mouse stomach libraries and with the whole small intestinal libraries using GoSurfer. The results disclosed that 243 of the 1767 GEP-associated GO-enriched genes identified from Fig. 1, step 1, were common to the other stem cell populations, whereas 574 were GO-enriched only in GEPs. Among the 1842 SiEP-associated GO-enriched expressed genes, 268 were common, whereas 742 were uniquely GO-enriched in the transcriptomes of this LCM population (Fig. 1, step 3, and supplemental Tables S10–S13).

The Venn diagrams in Fig. 1 summarize the results of further comparisons of the groups of filtered GO-enriched commonly and uniquely expressed genes identified in GEPs and SiEPs, including a 253-member group shared by GEPs and SiEPs but not with the other stem cell populations, and a group of 134 shared by GEPs, SiEPs, HSCs, NSCs, and ESCs (see supplemental Tables S14–S19 for complete lists of these genes).

To test the validity of this GO enrichment-filtering approach, we analyzed a control population of colonic crypt epithelial cells harvested by LCM from adult conventionally raised C57Bl/6J mice treated for 7 days with DSS in their drinking water. DSS produces a stereotyped pattern of ulcer formation in the distal colonic epithelium and a regenerative proliferative response in crypts surrounding these ulcers (15). EM studies disclosed that the regenerating crypt cells had features of a transit-amplifying population of immature members of the enterocytic and goblet cell lineages (15). A cDNA library was prepared from these LCM cells and sequenced, yielding 12,013 ESTs, representing 2394 unique transcripts (supplemental Table S1). No biological process GO terms were enriched in this set of gene products compared with two adult colonic cDNA libraries with 7470 Unigene clusters (supplemental Table S3). These findings indicate that the GEP and SiEP preparations...
**Mouse Gastric and Intestinal Epithelial Progenitors**

**TABLE 1**

| Canonical signaling/metabolic pathways | Pathway maps | \( p \) value | Enriched genes in the pathway |
|---------------------------------------|--------------|--------------|------------------------------|
| **A. GEPs**                           |              |              |                              |
| G-protein-coupled receptor signaling  | Fig. S3A     | 8.90E-5*     | Drd2, Grm8, Htr2c, P2ry4, Adra1a, Htr7, Oprd1, Adey1, Hrh1, Htr3 |
| Wnt/β-catenin signaling               | Fig. S3B     | 1.53E-3*     | Tie1, Dvl3, Frzb, Wnt16, Wnt9a, Acvr1, Tcf1 |
| Tyrosine metabolism                   | Fig. S3C     | 2.25E-3*     | Aoc3, Fab, Tyr |
| Serotonin receptor signaling          | Fig. S3D     | 5.02E-3*     | Htr2c, Htr7, Htr4 |
| GABA receptor signaling               | Fig. S3E     | 1.52E-2*     | Gabrg2, Gabra2, Gabra3 |
| Amino sugars metabolism               | Fig. S3F     | 2.57E-2*     | Gfpt1, Haxb, Cysb61 |
| **B. GEPs and SiEPS**                 |              |              |                              |
| Oxidative phosphorylation             | Fig. S4A     | 1.58E-2*     | Cc1c, Atp5g3, Atp5e, Uqcr9b, Atp5g2, Atp5h, Atp6v1f |
| GABA receptor signaling               | Fig. S4B     | 6.47E-02     | Gabrg3, Gabrb1 |
| Antigen presentation pathway          | Fig. S4C     | 6.47E-02     | H2-bi, B2m |
| G-protein-coupled receptor signaling  | Fig. S4D     | 7.95E-02     | Camk2b, Agrp1, Gm1, Htr1f |
| NF-κB signaling                       | Fig. S4E     | 7.95E-02     | Prkcz, Tra5, Tnfrsf11a, Tnfrsf1a |
| Glycine, serine and threonine         | Fig. S4F     | 8.82E-02     | Gamt |
| metabolism                           |              |              |                              |
| **C. SiEPS**                          |              |              |                              |
| Glycerolipid metabolism              | Fig. S5A     | 1.68E-9*     | Ppct2, Gpdp2, Pafah1b3, Adh5, Pla2g2a, Lipc, Cd1pt, Oc90, Aldh7a1, Aldh1b1, Aldh2, Adhfe1, Gpd1, Plaqg12a, Grn |
| Fatty acid metabolism                 | Fig. S5B     | 1.39E-8*     | Gdhh, Acads, Acadl, Adh5, Acadvl, Cyp2c8, Aldh7a1, Aldh1b1, Aldh2, Hadhsc, Adhfe1, Hadh2 |
| Oxidative phosphorylation             | Fig. S5C     | 4.92E-8*     | Uqcr6c2, Cox8c, Sdhc, Cyb5, Ndufb1, Atp5d, Atp5j, Ndufb9, Cox5b, Atp6t6b, Sdhb, Qp-c, Cox1l, Dap13, Ndufs4, Cox7s2, Pdp2, Atp5i, Cox7a21, Cox7c |
| Glycolysis/glucogenogenesis            | Fig. S5D     | 9.15E-8*     | Acsb2l, Eno3, Adh5, Aldoa, Aldh7a1, Aldh1b1, Aldh2, Pdha1, Adhfe1, Adh1, Ldhla, Ldc16, Pgm1 |
| Valine, leucine, and isoleucine        | Fig. S5E     | 1.64E-7*     | Gdhh, Pcca, Acads, Bckdh, Acadl, Acadvl, Hmgcs2, Aldh7a1, Aldh1b1, Aldh2, Hadhsc, Hadh2 |
| Propanoate metabolism                 | Fig. S5F     | 1.29E-6*     | Gdhh, Pcca, Acsb2l, Acads, Acadvl, Aldh7a1, Aldh1b1, Aldh2, Ldhla |
| **D. GEPs, HSCs, NSCs, ESCs**         |              |              |                              |
| Wnt/β-catenin signaling               | Fig. S6A     | 2.33E-5*     | Pp2r5c, Lrp1, Mark2, Ctnmb1, Lrp6, Tcf7l2 |
| PI3K/AKT signaling                    | Fig. S6B     | 3.56E-3*     | Tsc2, Ppp2r5c, Ctnmb1, Jak1, Lims1 |
| Interferon signaling                  | Fig. S6C     | 5.01E-3*     | Jak1, Ptpn2 |
| PDGF signaling                        | Fig. S6D     | 5.04E-3*     | Rasa1, Mapk3k1, Jak1 |
| T cell receptor signaling             | Fig. S6E     | 9.84E-3*     | Rasa1, Mapk3k1, Cbl |
| Estrin receptor signaling             | Fig. S6F     | 2.16E-2*     | Pcpp4, Ta9, Pparbp, Ncor1 |
| **E. GEPs, SiEPS, HSCs, NSCs, EBCs**  |              |              |                              |
| Chemokine signaling                   | Fig. S7A     | 7.72E-6*     | Ptk2, Camk2d, Calm2, Mapk1, Rhoa, Nras |
| Integrin signaling                    | Fig. S7B     | 3.01E-5*     | Arf3, Rock1, Dde1f, Ptk2, Cdc42, Mapk1, Rhoa, Nras |
| SAPK/JNK signaling                    | Fig. S7C     | 1.40E-4*     | Hnrrk, Cdc42, Crk, Gna13, Nras |
| VEGF signaling                        | Fig. S7D     | 1.20E-3*     | Ywhae, Ptk2, Mapk1, Arnt, Nras |
| IGF-1 signaling                      | Fig. S7E     | 5.18E-3*     | Ptk2, Mapk1, Nedd4, Nras |
| B cell receptor signaling             | Fig. S7F     | 6.76E-3*     | Camk2d, Calm2, Cdc42, Mapk1, Nras |
| **F. SiEPS, HSCs, NSCs, ESCs**        |              |              |                              |
| Wnt/β-catenin signaling               | Fig. S8A     | 1.36E-3*     | Apc, Ppp2ca, Csnk1d, Mapk3c, Tcf4 |
| IL-6 signaling                        | Fig. S8B     | 2.87E-3*     | Mapk14, Mapk2k4, Mapk3c, I6est |
| Alanine and aspartate metabolism      | Fig. S8C     | 4.93E-3*     | Nars, Dars |
| p38 MAPK signaling                    | Fig. S8D     | 1.24E-2*     | Mapk14, Mapk2k4, Mapk3c |
| Integrin signaling                    | Fig. S8E     | 1.87E-2*     | Mapk2k4, Itga6, Pak2, Pimk3x3, Actb |
| IL-4 signaling                        | Fig. S8F     | 2.13E-2*     | Hmg1a1, Pimk3x3 |

*The representation of a pathway in these progenitor populations is statistically significant \((p < 0.05)\) based on a right-tailed Fisher’s Exact Test.*

obtained by LCM of gnotobiotic Atp4b-tox176 and Defcr2-tox176 mouse guts are enriched for adult gut stem cells.

**Signaling and Metabolic Pathways Active in GEPs and/or SiEPS**—We used the IPA tool to further characterize functional properties of these gut epithelial progenitor populations. IPA annotations take into account GO annotations but are nonetheless distinct and based on a knowledge base of protein-protein interactions, extracted from the published literature. The IPA output includes signaling and metabolic pathways and a statistical assessment of their significance. This assessment utilizes a right-tailed Fisher’s Exact Test to calculate the probability that genes participate in a given pathway relative to their occurrence in all other pathway annotations.

Table 1 lists GEP- and SiEP-associated canonical signaling or metabolic pathways. The table is organized based on the groups defined by the Venn diagrams in Fig. 1 and provides a link to the corresponding pathway maps in the supplemental material. These pathway maps contain filtered GO-enriched gene products as well as all other pathway gene products that are present, albeit not enriched, in the gut progenitor EST libraries. For example, our IPA-based analysis indicated that GEPs have a statistically significant representation of Wnt/β-catenin signaling components (Table 1). Fig. 3 illustrates how GEPs and SiEPS, as well as the other stem cell populations, contain different identified combinations of GO-enriched gene products that map to the Wnt/β-catenin pathway. This provides one measure of the shared and distinctive features of these populations.

Bjerknes and Cheng (25) reported that enteric neurons juxtaposed to the basal surface of SiEPS, and a subset of enteroendocrine cells operate in concert to regulate the properties of multi- and oligo-potential crypt stem cells. An analogous situation may also exist in the isthmal gastric stem cell niche. The IPA tool revealed that a variety of components of
the G-protein-coupled receptor signaling cascades was significantly represented in our LCM GEP preparation, including pyrimidinergic (P2ry4), dopaminergic (Drd2), adrenergic (Adra1a), serotonergic (Htr2c), opioid (Oprd1), and glutamate metabotropic (Grm8) receptors. Interestingly, GEPs also produce a repertoire of subunits sufficient to assemble γ-aminobutyric acid, type A (GABA_A) receptors, i.e. Gabra2 and Gabra3 (α-subunits); Gabrb1 (β-subunit); and Gabrg2 and Gabrg3 (γ-subunits). Moreover, GEPs have the ability to synthesize GABA through a polyamine metabolic pathway that includes dopa decarboxylase (Ddc), an amine oxidase (copper containing 3; Aoc3), and aldehyde dehydrogenase 9 family member A1 (Aldh9a1). GABA_A receptors function as chloride channels; the presence of such channels in GEPs juxtaposed to acid-producing parietal cells may represent a mechanism for communication between these two occupants of the gastric stem cell niche.

We found that components of GABA biosynthetic pathways, including glutamic acid decarboxylase 1 (Gad1), Ddc, and Aoc3 are expressed in amplified GEPs as they undergo SV40 TAg-induced transformation to invasive gastric cancers in a transgenic mouse model (26). SiEPIs also express components of a GABA biosynthetic pathway and GABA_A receptor subunits, as do hematopoietic, neural, and embryonic stem cells (Table S20). These findings indicate that GABA signaling may be a more general feature of stem cells and a means for communicating with members of their respective niches.

Other pathways defined by IPA as having statistically significant representation in GEPs include the following: (i) phosphoinositide 3-kinase/Akt kinase (regulates β-catenin activity in intestinal stem cells; Ref. 27), (ii) integrin (integrins are up-regulated in hair follicle bulge epithelial stem cells compared with their progeny; Ref. 28), (iii) Igf signaling (Igf has a proliferative effect in adult hippocampal progenitors; Ref. 29); (iv) stress-activated protein kinase/c-Jun NH2-terminal kinase; and (v) vascular endothelial growth factor (Vegf) pathways (Table 1).

Among other functions, Vegf pathways affect endothelial cell survival (30). The isthmal stem cell niche of gastric units is surrounded by a complex microvasculature. We have shown previously that two of the three lineages descended from the multipotent gastric stem cell produce regulators of angiogenesis (VegfB in parietal cells and PdgfA and -B in zymogenic cells; see Ref. 31). Parietal cell-deficient gnotobiotic Atpb4-
tox176 mice have a 2-fold reduction in the density of the capillary network that surrounds their gastric units (31). Together, these findings suggest that angiogenesis is modulated by a cross-talk involving GEPs, their descendant lineages, and mesenchymal elements interposed between gastric units.

Many of the SiEP-specific filtered GO-enriched genes (Fig. 1, group C) are components of metabolic pathways. The glycerolipid pathway has the greatest statistical representation (p value, 1.68E-9) and includes three aldehyde dehydrogenases (Aldh1b1, Aldh2, and Aldh7a1) (Table 1 and supplemental Fig. S5A). The multipotent crypt stem cell is long lived.4 In other cell systems, generation of reactive oxygen species has been implicated in the pathway of lipid-induced cell death (e.g. Ref. 32).

Most intriguingly, these enzymes, which are involved in detoxification of aldehydes, may contribute to SiEP survival by limiting accumulation of byproducts of lipid metabolism. 

4Small intestinal stem cells in adult tri-transgenic FVB/N mice were marked using a doxycycline-inducible Cre-recombinase and an R26R Cre reporter, as described in Ref. 8. Analysis of mice sacrificed 0, 90, 100, 140, 170, and 250 days after a 4-day exposure to doxycycline in their drinking water revealed that the life span of the multipotent stem cell is 170 days but less than 250 days (n = 2–3 mice, 3805–4660 crypts from the middle third of the small intestine scored per time point).
of lipid peroxidation products. Other statistically significant pathways include fatty acid metabolism, oxidative phosphorylation, glycolysis/gluconeogenesis, propanoate, valine, leucine, and isoleucine metabolism (Table I; supplemental Fig. S5B–S5F). The metabolic properties expressed by stem cells in their niche have not been well characterized. These findings should help direct biochemical assessment of the principal substrates that SiEPs consume for energy and the amino acids required to fuel their well represented protein turnover pathways.

**Immunohistochemical Studies of Gut Epithelial Progenitors**—One potential weakness of our approach is that the progenitor populations we harvested by LCM came from gnotobiotic transgenic mice with a genetically engineered ablation of their gastric parietal cells or small intestinal Paneth cells. Although this allowed us to harvest sufficient quantities of progenitors from a manageable number of cryosections for subsequent cDNA library construction and sequencing, it nonetheless raises the question of whether some of the identified properties of these cells may reflect the effects of the ablutions we employed to promote their expansion (in the case of GEPs) or consolidation (in the case of SiEPs). Our GEP and SiEP data sets provide a starting point for determining whether the products of identified GO-enriched genes can be used to mark gut progenitors in normal conventionally raised animals. For example, we discovered that antibodies to Dcamkl1, the product of a GO-enriched transcript identified in the comparison between GEP and whole stomach EST libraries, mark single cells in the isthmal stem cell niche of gastric units present in normal, conventionally raised adult FVB/N mice (Fig. 4A). These solitary Dcamkl1-positive cells do not express biomarkers associated with differentiating members of the enteroendocrine (Fig. 4A, inset), parietal, or pit cell lineages (data not shown). Light microscopic, multilabel immunohistochemical studies disclosed that a subset of these solitary Dcamkl1 cells co-express glycans recognized by the neck cell-specific lectin GSII (Fig. 4A, arrowhead). Follow-up EM immunohistochemical studies confirmed that Dcamkl1 was produced by pre-neck cell progenitors (data not shown). Moreover, the fractional representation of Dcamkl1-positive cells is increased in parietal cell-deficient Atpb4-tox176 mice where the isthmal stem cell niche contains an expanded population of GEPs. Dcamkl1-positive cells are juxtaposed to rapidly cycling BrdUrd-positive progenitors in both normal (Fig. 4B, inset) and Atpb4-tox176 mice but are not themselves labeled after a 1.5-h exposure to BrdUrd (Fig. 4B). In addition, small intestinal crypts in normal mice contain solitary Dcamkl1-positive cells positioned just below the transit-amplifying cell population, in the region where the multipotential stem cell is thought to reside (33) (Fig. 4C). Dcamkl1 is a microtubule-associated kinase that was known to be expressed in neurons (34). Our findings now suggest that it is also a marker of adult gut stem cells.

**Mapk14** (p38) is a GO-enriched gene product in gut epithelial progenitors as well as HSCs, NSCs, and ESCs. We found that Mapk14 is expressed in isthmal epithelial cells in the normal mouse stomach (Fig. 4D); the census of Mapk14-positive isthmal cells increases in Atpb4-tox176 mice, just as their census of GEPs expands (Fig. 4D). We also noted that Mapk14 is also expressed near the base of normal small intestinal crypts (Fig. 4F). The latter finding supports our bioinformatics approach, which had shown that two Mapk14-containing pathways, p38 MAPK signaling and interleukin-6 signaling, were both represented to a statistically significant degree in SiEPs (see Table 1, group F; and supplemental Figs. S8D and S8B).

**Prospectus**—Our studies establish the feasibility and utility of harvesting consolidated populations of adult gut epithelial progenitors directly from their niches using laser capture microdissection, defining their transcriptomes by sequencing cDNA libraries, and then characterizing their biological properties using bioinformatics approaches. We have predicated on enrichment of GO-based functional terms rather than on levels of gene expression.5 The GEP and SiEP data sets reported here should help direct further efforts to mark adult gut stem cells in normal mice and humans using antibodies directed toward their intracellular proteins (e.g., see Fig. 4) or their integral membrane proteins (see supplemental Tables S21 and S22). The latter could provide a means for retrieving adult gut stem cells from their niches for *ex vivo* manipulation and further analyses.

**Acknowledgments**—We thank David O’Donnell, Maria Karlsson, Sabrina Waggoner, Sarah Pull, Chris Chen, Jaime Dant, and Yulia Korshunova for invaluable technical assistance. We also thank Melissa Wong, Jung Oh, Magnus Bjursell, and Sherif Karam for their help with various phases of this study.

**REFERENCES**

1. Croft, D. N., and Cotton, P. B. (1973) *Digestion* 8, 144–160
2. Potten, C. S. (1992) *Cancer Metastasis Rev.* 11, 179–195
3. Cheng, H., and Leblond, C. P. (1974) *Am. J. Anat.* 141, 461–479
4. Karam, S. M., and Leblond, C. P. (1993) *Anat. Rec.* 236, 259–279
5. Bjerknes, M., and Cheng, H. (1981) *Am. J. Anat.* 160, 77–91
6. Bjerknes, M., and Cheng, H. (1999) *Gastroenterology* 116, 7–14
7. Bjerknes, M., and Cheng, H. (2002) *Am. J. Physiol.* 283, G767–G777
8. Wong, M. H., Sazani, J. R., Stappenbeck, T. S., Rexer, C. H., and Gordon, J. I. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 12601–12606
9. Radtke, F., and Clevers, H. (2005) *Science* 307, 1904–1909
10. van Es, J. H., van Gijn, M. E., Riccio, O., van den Born, M., Vooijs, M., Begthel, H., Cojzner, M., Robine, S., Winton, D. J., Radtke, F., and Clevers, H. (2005) *Nature* 435, 959–963
11. Fre, S., Huyghe, M., Mourikis, P., Robine, S., Louvard, D., and Artavanis-Tsakonas, S. (2005) *Nature* 435, 964–968
12. Mills, J. C., Andersonson, N., Hong, C. V., Stappenbeck, T. S., and Gordon, J. I. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 14819–14824
13. Stappenbeck, T. S., Mills, J. C., and Gordon, J. I. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 1004–1009
14. Hooper, L. V., Mills, J. C., Roth, K. A., Stappenbeck, T. S., Wong, M. H., and Gordon, J. I. (2002) *Methods Microbiol.* 31, 559–589
15. Pulli, S. L., Doherty, J. M., Mills, J. C., Gordon, J. I., and Stappenbeck, T. S. (2005) *Proc. Natl. Acad. Sci. U. S. A.* 102, 99–104
16. Stappenbeck, T. S., Hooper, L. V., Manchester, J. K., Wong, M. H., and Gordon, J. I. (2002) *Methods Enzymol.* 356, 168–196
17. Hawkins, R. D., Bashiardecs, S., Helms, C. A., Hu, L., Saccone, N. L., Warchol, M. E., and Lovett, M. (2003) *Hum. Mol. Genet.* 12, 1261–1272
18. Altshul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410
19. Mott, R. (1997) *Comput. Appl. Biosci.* 13, 477–478
20. Syder, A. J., Gurge, J. L., Li, Q., Oleksiewicz, C., Lorenz, R. G., Karam, S. M., Falk, P. G., and Gordon, J. I. (1999) *Mol. Cell* 3, 263–274
21. Karam, S. M. (1993) *Anat. Rec.* 236, 314–332
22. Syder, A. J., Oh, J. D., Gurge, J. L., O’Donnell, D., Karlsson, M., Mills, J. C., Björkhoml, B. M., and Gordon, J. I. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 3467–3472
23. Cheng, H. (1978) *Am. J. Anat.* 141, 521–535
24. Zhong, Z., Storch, K. F., Lipan, O., Kao, M. C., Weitz, C. J., and Wong, W. H. (2004) *Appl. Bioinformatics* 3, 261–264
25. Bjerknes, M., and Cheng, H. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 98, 12497–12502
26. Syder, A. J., Karam, S. M., Mills, J. C., Ippolito, J. E., Ansari, H. R., Farook, V., and Gordon, J. I. (2004) *Proc. Natl. Acad. Sci. U. S. A.* 101, 4471–4476
27. He, X. C., Zhang, J., Tong, W. G., Tawfik, O., Ross, J., Scoville, D. H., Tian, Q., Zeng, X., He, X., Wiedemann, L. M., Mishina, Y., and Li, L. (2004) *Nature* 436, 1117–1121

5 For this reason, we did not incorporate a GeneChip data set representing label-retaining cells that populate the bulge stem cell niche of mouse hair follicles (28) in the analysis pipeline described in Fig. 1. However, 32% (1296/4031) (32%) of GEP transcripts and 33% (1105/3324) of SiEP transcripts are present in bulge label-retaining cells (supplemental Table S3). In contrast, SiEPs and GEPs share only 17 and 23% of their transcripts with whole adult small intestinal and stomach cDNA libraries, respectively.
28. Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W. E., Rendl, M., and Fuchs, E. (2004) Science 303, 359–363
29. Aberg, M. A., Aberg, N. D., Palmer, T. D., Alborn, A. M., Carlsson-Skwirut, C., Bang, P., Rosengren, L. E., Olsson, T., Gage, F. H., and Eriksson, P. S. (2003) Mol. Cell. Neurosci. 24, 25–40
30. Gupta, K., Kshirsagar, S., Li, W., Gui, L., Ramakrishnan, S., Gupta, P., Law, P. Y., and Hebbel, R. P. (1999) Exp. Cell Res. 247, 495–504
31. Mills, J. C., Andersson, N., Stappenbeck, T. S., Chen, C. M., and Gordon, J. I. (2003) J. Biol. Chem. 278, 46138–46145
32. Listenberger, L. L., Ory, D. S., and Shaffer, J. E. (2001) J. Biol. Chem. 276, 14890–14895
33. Li, L., and Xie, T. (2005) Annu. Rev. Cell Dev. Biol. 21, 605–631
34. Lin, P. T., Gleeson, J. G., Corbo, J. C., Flanagan, L., and Walsh, C. A. (2000) J. Neurosci. 20, 9152–9161