Supporting Information

Huttlin et al. 10.1073/pnas.0909282106

SI Text

Supplementary Materials and Methods. Unless otherwise noted, chemicals were purchased from Sigma-Aldrich. All mice were raised in the American Association for Laboratory Animal Care-approved animal facility of the McArdle Laboratory following all appropriate regulations, and all protocols received institutional review board approval.

Metabolic Labeling of Mice. A total of 8 C57BL/6 female mice were used in this experiment: 4 wild-type mice and 4 mice that were heterozygous for the Apc^{Min} mutation (Min mice). Each Min mouse was paired with a wild-type sibling, and 2 pairs were drawn from each of 2 litters. At weaning all mice were fed a diet derived from either natural-abundance or $>$ 98% ^{15}N-labeled *Spirulina* (Spectra Stable Isotopes) and protein-free rodent chow, supplemented with nitrogen-depleted oil and vitamins, based on the method published by Wu et al. (1, 2). This diet was formulated at Harlan Teklad, Inc. Half of the mice (2 Min and 2 wild type) were fed the ^{15}N-labeled diet, and their sibling partners were fed the natural-abundance diet. Mice were separated into 2 cages based on their intended isotopic label and were provided with a specific daily ration of the *Spirulina*-based diet. No obvious differences in health or behavior were observed between mice fed the natural-abundance and ^{15}N-labeled diets.

After 46 days on the *Spirulina*-based diet, all mice were killed at 68 days of age. All major organs and tissues were dissected and individually flash-frozen in liquid nitrogen before storage at −80 °C. Individual colonic tumors and tumors of the small intestine were identified by eye and were individually dissected as well. Tumor samples were compared with normal epithelium taken from the colon or small intestine of a matched sibling with opposite genotype and isotopic label.

Protein Extraction. Pairs of labeled and unlabeled frozen tissues (1 Min and 1 wild-type) were combined and ground to a fine powder under liquid nitrogen. The ground organs then were suspended in boiling buffer containing 50 mM Tris-HCl, pH 8.0/1% (wt/vol) SDS/and 200 mM DTT to a volume of 20 μL per milligram of tissue. These suspensions then were homogenized with a Potter-Elvehjem grinder (30–100 strokes up and down, depending on the tissue). Remaining debris was pelleted via centrifugation (4 min at 13,200 $\times$ *g* in a microcentrifuge), and the protein concentration of the supernatant was determined via the BCA assay (Pierce) or the Warburg-Christian method (3). Samples then were frozen as aliquots at −80 °C until use.

Protein Digestion. After thawing of each protein extract, trichloroacetic acid was added to 10% (vol/vol), and samples were incubated at 4 °C for 10 min to precipitate proteins. Samples then were centrifuged for 4 min at 13,200 $\times$ *g* in a microcentrifuge to pellet protein, and the supernatant was discarded. The protein pellet then was washed 3 times with 80% (vol/vol) acetone in water and allowed to air dry. Once dry, the protein pellet was resuspended in a solution of 8 M urea in water. The solution subsequently was diluted 8-fold with buffer to a final concentration of 40 mM ammonium bicarbonate, pH 8.0/2 mM DTT/and 1 M urea. Sequencing grade, modified porcine trypsin (Promega) then was added at a ratio of 1:20 wt/wt and allowed to digest the protein samples overnight at room temperature. Following digestion, samples were acidified with formic acid to 5% (vol/vol) and were solid-phase extracted using Spec PT C18 solid-phase extraction pipette tips (Varian) according to the manufacturer's protocol. After elution with 90% (vol/vol) acetonitrile in water, samples were evaporated to dryness in a vacuum centrifuge and were resuspended in 0.1% (vol/vol) formic acid in water at an estimated concentration of 1–2 μg/μL for MS analysis.

QTOF Analysis. Samples were analyzed on a Micromass QTOF-II mass spectrometer and an Agilent 1100 HPLC as described previously (4, 5). Briefly, 10-μg peptide samples were loaded onto trapping columns containing 5 μm Zorbax C18 resin (Agilent) and were separated using home-packed fused silica columns (100 μm $\times$ 11 cm) containing 5 μm Matrix C18-AQ resin (Column Engineering Inc.). Peptides were eluted over 2 hours at a constant flow rate of 200 nL/min with a gradient of increasing acetonitrile concentration. The mass spectrometer was operated in data-dependent MS/MS mode, with a single 3.3-second sequencing event following each MS survey scan. Individual precursors were isolated within a $\pm$ 3 *m/z* window with subsequent exclusion within $\pm$ 1.2 *m/z* of the target mass for 2 minutes following each sequencing attempt.

MS and MS/MS data were converted to peaklists using Protein Lynx Global Server 2.1.5 (Micromass). Each MS spectrum was smoothed twice using a 7-point Savitsky-Golay smooth, and background was subtracted assuming a first-order polynomial with a 35% threshold. MS/MS data underwent 2 subsequent smoothing steps via a 7-point Savitsky-Golay algorithm and background subtraction via the adaptive algorithm. Neither MS nor MS/MS spectra were deisotoped.

LC-MALDI TOF/TOF Analysis. LC-MALDI separations were performed using an Agilent 1100 nano-HPLC with a 1100 series micro fraction collector and MALDI target adaptors. These separations used 0.1% (vol/vol) trifluoroacetic acid (TFA) in water as buffer A and 0.1% (vol/vol) TFA, 5% (vol/vol) isopropyl alcohol, and 85% (vol/vol) acetonitrile in water as buffer B. Samples containing 10–15 μg of protein were loaded initially onto a trapping cartridge containing 5 μm Zorbax C18 resin (Agilent). After 20 minutes, the trapping cartridge was aligned with a home-packed 360 $\times$ 75-μm resolving column. The resolving columns were packed with either 5-μm Matrix C18AQ spherical silica beads (100-Å pores), or 3-μm Monitor C18 spherical silica beads (100-Å pores) (Column Engineering, Inc.). Peptides were eluted at a flow rate of 200 nL/min over approximately 4 hours using the following gradient: linear ramp from 1% B to 10% B over 25 min; ramp to 40% B at 185 min; ramp to 60% B at 195 min; ramp to 100% B at 205 min; decrease to 1% B at 215 min; and hold at 1% B to 250 min. Eluent from the column was mixed with MALDI matrix (3.75 mg/mL alpha-cyano-4-hydroxycinnamic acid) in 0.1% (vol/vol) TFA, 75% (vol/vol) acetonitrile, and 30 fmol/μL adrenocorticotropic hormone (ACTH) (18–29 fragment internal standard) at a ratio of 4:1 in a static mixing tee for a combined flow rate of 1.0 μL/min. The combined eluent then was spotted at 0.2-min intervals between 40 min and 250 min, depositing 1050 spots onto a stainless steel MALDI plate (Applied Biosystems, Inc).

MS analysis was performed on an Applied Biosystems 4800 MALDI TOF-TOF analyzer. Positive-ion MS spectra were collected in reflector mode covering 700-4000 *m/z* acquiring 1,200 laser shots per spectrum. Internal calibration was used for MS spectra using the mass of the singly charged ACTH peptide fragment at 2465.199 *m/z*. MS/MS spectra were obtained in 2-kV fragmentation mode using air as the collision-induced dissoci-

ation gas. Precursor selection was automated using a Job-Wide Interpretation method in the 4000 Series Explorer Software (Applied Biosystems). Up to 15 precursors were selected per spot for acquisition in decreasing order of signal strength. Individual precursors were isolated with resolution of 300, and up to 2,000 laser shots were acquired per MS/MS spectrum. External calibration was used for MS/MS data, using peptide standards spotted along the upper and lower edges of the plate and in the first and last spots of the gradient.

Peak lists were generated for database searching using the 4000 Series Explorer Software (Applied Biosystems). Peak lists encompassed the mass range from 60 to 20 *m/z* below the precursor mass-to-charge ratio and required a minimum signal-to-noise ratio of 7 for each peak. The exclusion tolerance was set to 0.1 Da, and a peak density filter was used, limiting the spectrum to no more than 50 peaks per 200 Da. A cluster area filter also was used, with the maximum number of peaks set at 75.

Identification of Peptides and Proteins. The Mascot algorithm was used for identification of all peptides from MS/MS data (6). All searches were performed against a National Center for Biotechnology Information database containing sequences of all known mouse proteins in forward and reverse order as well as the sequences of common protein contaminants such as porcine trypsin and human keratin. All searches assumed trypsin digestion with up to 2 missed cleavages and considered oxidation of methionine and deamidation of glutamine and asparagine as variable modifications. QTOF searches considered 1+, 2+, and 3+ charge states and required precursor and product ion tolerances of ± 0.2 Da. TOF-TOF searches considered only singly charged peptides and used a precursor ion tolerance of ± 20 ppm. Peptide identifications made against the decoy (reversed) protein sequences were used to define FDRs, as discussed previously (7, 8). Minimum Mascot score thresholds were selected to provide a 1% peptide-level FDR for each dataset. For the purposes of assessing FDRs, multiple observations of the same peptide sequence with different posttranslational modifications were considered to be distinct.

Before peptides were assembled into proteins, each peptide sequence was individually compared against the entire mouse proteome, and only peptides whose sequence was unique to a single protein were retained. Although this method is somewhat more restrictive than standards used elsewhere, it helps ensure that the quantitative information provided by each peptide in our dataset reflects only a single protein. These unique peptides then were grouped based on the protein sequences from which they were derived to define a minimal set of proteins that were present in the original sample.

Determination of ^{15}N Incorporation from QTOF Data. Levels of ^{15}N incorporation were automatically determined for each peptide identified by Mascot as described previously (4, 5). Briefly, relative intensities were extracted from MS survey spectra corresponding to the monoisotopic peak and each isotope for all possible isotopomers for each peptide. The peaks corresponding to the natural-abundance form of each peptide were subtracted, and the remaining peaks were matched against a library of predicted isotopic distributions spanning all possible ^{15}N incorporations from 1% to 99% in 1% increments. The ^{15}N incorporation corresponding to the best-fitting isotopic distribution in the library then was assigned for each peptide. A minimum correlation of 0.85 between predicted and observed isotopic distributions was required to remove questionable measurements. Incorporations reported for each mouse tissue represent averages of all individual peptide measurements from each sample. All calculations were performed in Mathematica 5.2 (Wolfram Research).

Peptide Quantitation from TOF-TOF Data. In the most general terms, abundance ratios comparing ^{15}N- and ^{14}N-labeled forms of each peptide were determined from TOF-TOF data by independently scaling theoretical isotopic distributions for both heavy and light forms of each peptide to raw data and comparing the intensities of the scaled distributions on a scan-by-scan basis. Incorporations of ^{15}N were estimated simultaneously for each peptide by comparing isotopic envelopes representing a range of ^{15}N incorporations with the observed raw data. A more detailed step-by-step description of the algorithm follows. All steps were implemented within Mathematica 5.2.

Before peptide quantitation, all MS/MS data were searched via Mascot, and a list of all peptides identified at a 1% FDR was prepared, indicating which MALDI spots contained each peptide. Separately, raw MS survey spectra were exported from the Applied Biosystems Oracle database as individual mzxml files using T2DExtractor, a freely distributed utility developed by Takis Papoulias at the University of Michigan. Individual mzxml files were then read into Mathematica using the freely distributed, Java-based MzXMLParser written by Robert Hubley of the Institute for Systems Biology.

For each peptide identified during an LC-MALDI experiment, the following tasks were performed. First, predicted distributions were calculated for the natural-abundance form of each peptide as well as a user-defined range of ^{15}N incorporations (for these experiments, 85%−95% ^{15}N was considered, with 1% increments). These distributions were calculated via binomial expansion based on the molecular formula of each peptide as well as the isotopic incorporation levels for each of the 6 elements commonly found in peptides (C, H, O, N, P, and S). All elements except nitrogen were assumed to be present at their natural abundances. Each calculated distribution was normalized in the sense that the sum of all peaks in each distribution was 1.0.

After theoretical distributions were calculated, the MS survey scan from the spot in which the peptide was identified was read into memory. The observed intensity of each isotope peak was taken to be the maximum intensity observed within ± 200 ppm of the each isotope's expected mass. Each of the predicted ^{15}N isotopic distributions was aligned with the observed isotope peaks, and the quality of the match was judged by linear regression. The best-matching predicted distribution was used to determine the ^{15}N incorporation observed for the peptide. A minimum R^2 of 0.8 was required to accept an incorporation value for a peptide.

If an acceptable estimate of ^{15}N incorporation was found, the peptide was quantified via a multistep process. First, the predicted natural-abundance distribution was aligned with the isotope peaks observed in the raw spectrum and regressed. The slope provided by the linear regression is a scaling factor that reflects the intensity of the observed distribution compared with the normalized, predicted distribution; the intercept accounts for background noise in the spectrum. The correlation coefficient provides an indication of the quality of the fit. Similarly, the best-matching ^{15}N distribution was regressed against the observed isotope peaks, providing a scaling factor that represents the intensity of the heavy isotopic envelope as well as a correlation coefficient that indicates the quality of the fit. Finally, the scaling factors for both heavy and light forms of each peptide were compared to determine their relative level of abundance. Minimum R^2 values of 0.8 were required for both heavy and light peptide envelopes to accept the resulting ratio for quantitation. This process was repeated 3 times for each peptide: once for the spot in which the peptide was identified via MS/MS and once in each of the spots immediately before and after the peptide-containing spot.

To correct for slight differences in mixing of light and heavy tissue at the outset of each experiment, all peptide ratios were

normalized to the median peptide ratio that was observed for each dataset. All values then were transformed to a $\log_2$ scale to allow symmetric representation of ratios both larger and smaller than 1.0 (or 0 in log space).

Protein Quantification. After all peptides were quantified, normalized, and transformed to log space, the peptides with acceptable ratios were grouped according to their parent proteins. Only unique peptides (i.e., those that occur exactly once in the mouse proteome and are indicative of a single protein) were considered. All measurements from each protein were gathered, and Grubb's test was applied to remove outliers. The mean and standard deviation of all measurements for each protein then were determined, and the mean ratio was taken as an estimate of relative abundance for each protein.

Stable Isotope Dilution Mass Spectrometry. To confirm results observed via metabolic labeling, the proteins coronin, HMG CoA synthase, and SCFP-25 were quantified via SID using synthetic peptides as internal standards. Only unique peptide sequences were considered for use as internal standards. Sequences that were observed via LC-MALDI in our metabolic labeling surveys were favored, and all sequences considered were between 8 and 16 aa long. We avoided peptide sequences that were likely to produce ragged ends during tryptic digestion (i.e., multiple positively charged residues at either end of the tryptic peptide). Peptides that contained amino acids that are reactive or easily oxidized (Trp, Met, and Cys) also were avoided. The peptides selected for monitoring each protein are summarized in Table S3.

The peptide matching coronin was synthesized via standard solid-phase peptide synthesis techniques. In contrast, the peptides matching SCFP-25 and HMG CoA synthase were synthesized at reduced scale via microwave-assisted solid-phase synthesis using 'lantern' peptide synthesis techniques. A single amino acid containing ^{13}C and ^{15}N in place of their natural-abundance counterparts was introduced into each peptide at a specific location (see Table S3). Labeled amino acids were chosen to minimize cost and to ensure a minimum mass difference of 5 Da between labeled and unlabeled forms. Following synthesis, peptides underwent HPLC purification if necessary and were lyophilized. Stock solutions of each were prepared, estimating peptide concentrations based on the weight of the lyophilized powder dissolved. Peptide purity was assessed, and multiple-reaction monitoring transitions were selected and optimized following direct injection MS analysis (see Table S3).

A total of 4 colonic tumor samples were obtained from female *Min* mice ranging in age from 66 to 80 days. Normal colonic tissue was obtained from 8 female wild-type siblings. Proteins were extracted and digested as described for metabolic labeling analyses, except that all protein samples were quantified via the BCA assay (Pierce) before digestion. Internal standard peptides for coronin, HMG CoA synthase, and SCFP-25 were spiked into each sample during tryptic digestion (see Table S3). Digestion conditions were identical to those used for the metabolic labeling experiment, except that trypsin levels were reduced to 1:50 and samples were analyzed via LC-MS directly, without prior solid-phase extraction.

All samples were analyzed via multiple-reaction monitoring on a 3200 QTrap mass spectrometer (Applied Biosystems, Inc.). Peptides were separated on a Vydac polymeric reversed-phase column (2.1 mm × 25 cm) packed with 5-μm beads (# 218TP52). Solvent A was 0.1% formic acid in water; solvent B was 0.1% formic acid in acetonitrile. After loading samples at 5% B and ramping to 10% B over 5 min, peptides were separated via a 30-min gradient from 10% B to 40% B, followed by a quick ramp to 100% and a 15-min re-equilibration at 5% B. Peptides were quantified automatically using Agilent software, although all integrations were inspected manually to ensure proper peak detection and consistent quantification.

Microarray Analysis. A series of 4 colonic tumors were dissected from 3 female C57BL/6 Min mice (aged 68 days) and analyzed via the Agilent microarray platform. A single tumor was obtained from each of 2 mice, and 2 tumors were excised from the third mouse. All tumors were compared with normal colorectal tissue from a single wild-type female mouse. All samples were processed and analyzed on the Agilent standard mouse microarray according to the Two-Color Microarray-Based Gene Expression Analysis Protocol, version 4.0.2 (Agilent Technologies). Briefly, tissue samples were resuspended in RNeasy Lysis Buffer (Buffer RLT, Qiagen), and total RNA was extracted using the AllPrep DNA/RNA Mini kit (Qiagen). The total RNA then was converted to cDNA, which was amplified during conversion to cRNA with T7 RNA polymerase. Each cRNA sample then was labeled with appropriate dyes: tumor samples were labeled with cy3, and normal tissue was labeled with cy5. Labeled cRNA was purified to remove unbound dye, and each sample was quantified using the nanodrop method. Each cy3-labeled tumor sample was combined with an equal amount of cy5-labeled cRNA from normal tissue, and the mixed sample was hybridized to the 60-mer oligonucleotide chip. Then each Agilent microarray chip was washed and scanned. Ratios of abundance for individual probes were normalized using functions in the Agilent Two-Color Microarray-Based Gene Expression Analysis Protocol, version 4.0.2.

1. Wu CC, MacCoss MJ (2007) Quantitative proteomic analysis of mammalian organisms using metabolically labeled tissues. *Methods in Molecular Biology* 359:191–201.
2. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR, III (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* 76(17):4951–4959.
3. Warburg O, Christian W (1942) *Biochem Z* 310:384–421.
4. Huttlin EL, Hegeman AD, Harms AC, Sussman MR (2007) Comparison of full versus partial metabolic labeling for quantitative proteomics analysis in *Arabidopsis thaliana*. *Molecular and Cellular Proteomics* 6(5):860–881.
5. Nelson CJ, Huttlin EL, Hegeman AD, Harms AC, Sussman MR (2007) Implications of 15N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. *Proteomics* 7(8):1279–1292.
6. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567.
7. Huttlin EL, Hegeman AD, Harms AC, Sussman MR (2007) Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. *Journal of Proteome Research* 6(1):392–398.
8. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* 4(3):207–214.

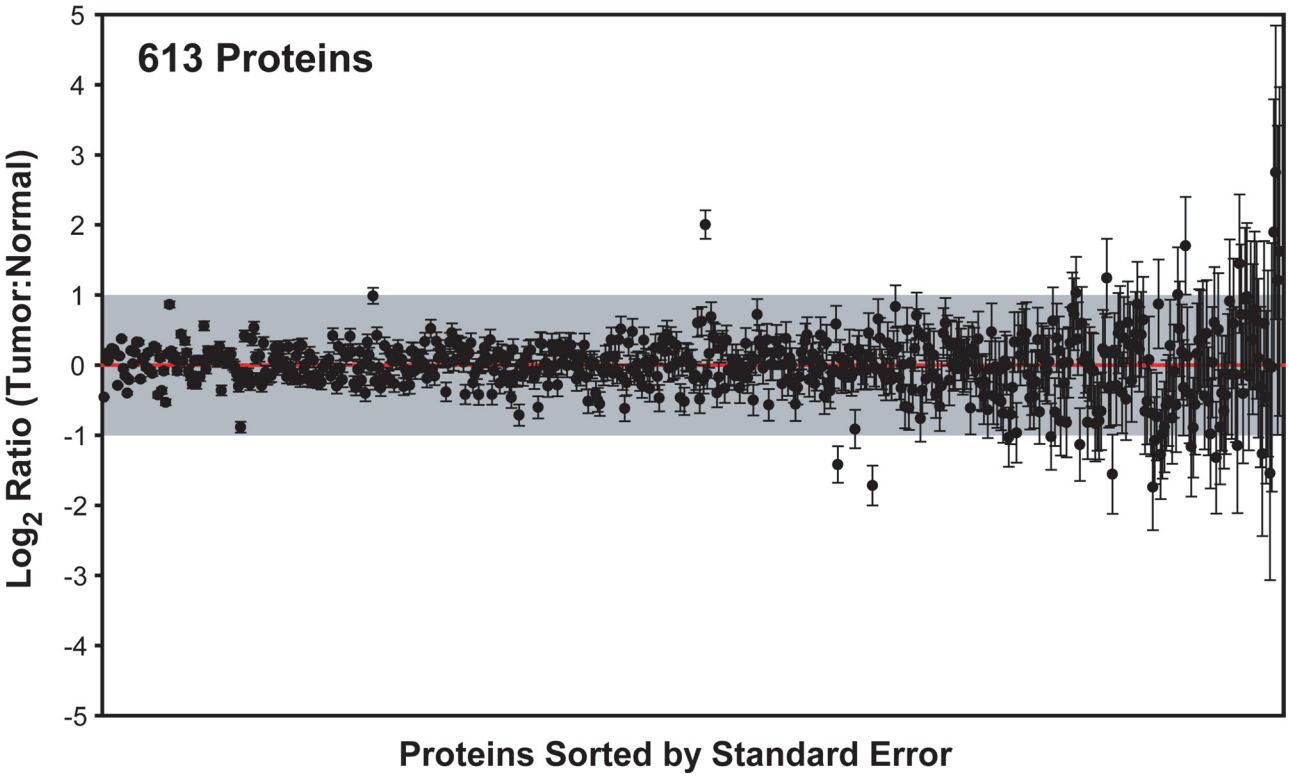

Fig. S1. The distribution of ratios for all proteins identified in our quantitative comparison of tumor versus normal colonic tissue. Proteins are represented by black circles whose location on the *y*-axis reflects that protein's normalized average ratio across all mouse pairs. The error bars associated with each protein represent ±1 standard error. Proteins are arranged along the *x*-axis in order of increasing standard error. The shaded gray area represents ratios that are less than 2-fold. The red line represents a 1:1 ratio (no change).

Other Supporting Information Files

Table S1
Table S2
Table S3
Table S4
Table S5
Table S6
Table S7