## Supplemental Methods

### *Experimental Diets.*

After a 1-wk acclimation on standard pelleted diet, rats were assigned to one of four diet groups, which differed in the type of fat and fiber as previously described (4). Diets contained (g/100 g diet): dextrose, 51.00; casein, 22.40; D,L-methionine, 0.34; AIN-76 salt mix, 3.91; AIN-76 vitamin mix, 1.12; choline chloride, 0.13; pectin or cellulose, 6.00. The total fat content of each diet was 15% by weight with the n-6 PUFA diet containing 15.0 g corn oil/100 g diet and the n-3 PUFA diet containing 11.5 g fish oil/100g diet plus 3.5 g corn oil/100 g diet to prevent essential fatty acid deficiency.

### *mRNA Analysis.*

Following incubation in PBS containing 100 µg/ml cycloheximide, colonic epithelial cells were allowed to swell in LSB (20 mM Tris pH 7.5, 10 mM NaCl and 3 mM $MgCl_2$) containing 1 mM dithiothreitol and 50 U RNase inhibitor for 2 min followed by lysis in LSB containing 0.2 M sucrose and 1.2% Triton X-100. After removal of nuclei by centrifugation, the supernatant was layered over a 15-50% linear sucrose gradient (in LSB) and centrifuged at 247,000 x g for 2 h at 4°C in a swinging bucket rotor. Gradients were fractionated, aliquots were taken for absorbance at 254 nm, and 3 vol of denaturation solution (Ambion Totally RNA kit) was immediately added to the remainder of each fraction. Samples were frozen at -80°C until RNA was isolated using the Totally RNA kit (Ambion) as per manufacturer's instructions followed by DNase treatment. Both total RNA and polysome RNA were analyzed on an Agilent Bioanalyzer to assess RNA integrity. CodeLink rat whole genome bioarrays (Applied Microarray) were used to assess gene expression (3).

### *Gene set enrichment analysis (GSEA).*

GSEA was used to determine whether a "gene set" of interest is uniformly randomly distributed (or not) in a larger list of genes (gene list) sorted according to a t-test for differential expression (8, 9). In experiments described herein, the target list corresponds to genes in the total or polysomal

expression data sets which are ranked by the respective t-test for differential expression. The gene sets are made up of the putative targets of microRNAs obtained from Target Scan. GSEA calculates an enrichment score (ES) that reflects the degree to which the targets within the gene set are over-represented in the respective gene list. A p-value is assigned to the ES score by a permutation test, which states whether the enrichment is significant or not. GSEA then calculates a normalized enrichment score (NES), which takes into account the number of genes within the independent gene set.

***Cumulative distribution function analysis.***

For each comparison group (i.e., Tumor/Saline, CA/FA, and CCA/FPA), both polysomal and total mRNA expression data sets were created using appropriate samples (rows=probes, columns=rats). Each data set was individually median and quantile normalized using only rows containing all "good" probe readings. Average probe values for each of the two treatments in the data set were computed, and then used as expression data (e.g., probe fold change between treatments would be the ratio of these averages). Based on a previous analysis, differentially expressed microRNAs were identified for each comparison group, and for each microRNA, mRNAs were classified according to Target Scan by conserved 8mer, conserved 7mer types + non-conserved 8mer and 7mer types, and "other". These groupings roughly correspond to "strong", "weak", and "non" targets (7). Distributions of fold change between treatments for each of these groups were estimated using those mRNA that are available in the data set. A comparison of the empirical fold change distributions of conserved 8mer targets and "non" targets was made for selected microRNA's of known interest using the Kolmogorov-Smirnov test, and p-values and their associated q-values (false discovery rate levels) were computed.

***Classification of microRNAs as biomarkers of colon tumor development.***

For this analysis, the main goal was to determine if microRNA expression can be used to discriminate between the different experimental treatments: Tumor vs saline (T vs S), corn oil+AOM

vs fish oil+AOM (CA vs FA) and corn oil+cellulose+AOM vs fish oil+pectin+AOM (CCA vs FPA). We have previously used a linear classifier algorithm for feature set identification (2, 10). For the purpose of identifying feature sets, we designed classifiers that categorize samples based on the expression values of the microRNAs from the intersection of the microRNAs altered in tumor versus (not altered) saline. Classifiers for microRNAs feature sets of sizes 1, 2, and 3 were identified. Generally, there are two reasons why it is desirable to design classifiers involving small numbers of features: (a) the limited number of samples often available in clinical studies makes classifier design and error estimation problematic for large feature sets (5) and (b) small microRNAs sets facilitate design of practical diagnostic panels. For similar reasons, simple classifiers are preferable for small samples; indeed, for small samples, if good classification is possible, then a simple classifier such as linear discriminant analysis (LDA) using a small number of microRNAs will typically outperform a complex classifier (6)

Given a set of features on which to base a classifier, one has to address not only the classifier design from sample data but also the estimation of its error, i.e., the precision with which the error of the designed classifier estimates the error of the optimal classifier. When data are limited, an error estimator may have a large variance and therefore may often be low even if it is approximately unbiased. This can produce many feature sets and classifiers with low error estimates. The algorithm used in this study mitigates this problem by applying the bolstered error estimation (1). Basically, this approach “bolsters” the original empirical distribution of the available data by means of suitable bolstering kernels placed at each data point location. The error can be computed analytically in some cases, such as in the case of LDA. The result of the overall approach is a list of “best” feature sets from among all possible feature sets. Hence, the best feature set is the one possessing minimum classification error. Because we only have data and not the underlying feature-label distributions, the errors have been estimated from the data. This approach takes into account that, in small-sample settings, we do not have much confidence in any single feature set and that it is much more likely that,

if there is an adequate sized collection of good performing feature sets, then there are likely to be some that perform well on the overall population (6).

## REFERENCES:


1. **Braga-Neto UM, and Dougherty ER**. Is cross-validation valid for small-sample microarray classification? *Bioinformatics* 20: 374-380, 2004.
2. **Chapkin RS, Zhao C, Ivanov I, Davidson LA, Goldsby JS, Lupton JR, Mathai RA, Monaco MH, Rai D, Russell WM, Donovan SM, and Dougherty ER**. Noninvasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *Am J Physiol Gastrointest Liver Physiol* 298: G582-589, 2010.
3. **Davidson LA, Wang N, Ivanov I, Goldsby J, Lupton JR, and Chapkin RS**. Identification of actively translated mRNA transcripts in a rat model of early-stage colon carcinogenesis. *Cancer Prev Res (Phila Pa)* 2: 984-994, 2009.
4. **Davidson LA, Wang N, Shah MS, Lupton JR, Ivanov I, and Chapkin RS**. n-3 Polyunsaturated fatty acids modulate carcinogen-directed non-coding microRNA signatures in rat colon. *Carcinogenesis* 30: 2077-2084, 2009.
5. **Davis CA, Gerick F, Hintermair V, Friedel CC, Fundel K, Kuffner R, and Zimmer R**. Reliable gene signatures for microarray classification: assessment of stability and performance. *Bioinformatics* 22: 2356-2363, 2006.
6. **Dougherty ER, Brun M, Trent JM, and Bittner ML**. Conditioning-based modeling of contextual genomic regulation. *IEEE/ACM Trans Comput Biol Bioinform* 6: 310-320, 2009.
7. **Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, and Bartel DP**. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27: 91-105, 2007.
8. **Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, and Groop**

**LC**. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267-273, 2003.

9. **Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP**. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545-15550, 2005.
10. **Zhao C, Ivanov I, Dougherty ER, Hartman TJ, Lanza E, Bobe G, Colburn NH, Lupton JR, Davidson LA, and Chapkin RS**. Noninvasive detection of candidate molecular biomarkers in subjects with a history of insulin resistance and colorectal adenomas. *Cancer Prev Res (Phila Pa)* 2: 590-597, 2009.