Supplementary Methods

Study Subjects

Patients with FAP were recruited from the familial colon cancer program and registry at The University of Texas MD Anderson Cancer Center. To be eligible for enrollment in this phase II celecoxib study, patients had to have a measurable polyp burden in either the native colon or, in patients who had undergone colectomy and ileorectal anastomosis, the residual rectal mucosa. Several additional inclusion and exclusion criteria were employed, typical of those used for clinical chemoprevention trials. The study was approved by MD Anderson's Institutional Review Board, and patients gave informed consent before participating.

Inclusion and Exclusion Criteria

Inclusion criteria for the study were (a) adequate bone marrow function (indicated by an absolute neutrophil count of > 1,500/ml and a platelet count of > 100,000/ml) and serum creatinine, total bilirubin, and alanine aminotransferase levels < 1.5 × the upper limit of normal, (b) age older than 18 years, (c) ability to provide informed consent, and (d) for women of childbearing potential (defined as all women except those 2 or more years postmenopausal or surgically sterile), absence of pregnancy or lactation; use of adequate contraceptive measures (i.e., abstinence, IUD, oral contraceptives, or diaphragm or condom with spermicidal gel) starting with the most recent menses and throughout the study duration; and a negative serum pregnancy test within 14 days before starting celecoxib.

Exclusion criteria for the study included (a) total colon and rectum resection, (b) inflammatory bowel disease, (c) anti-inflammatory medications (e.g., NSAIDs, aspirin, or sulfasalazine) that could not be discontinued 3 days before enrollment, (d) chemotherapy or radiotherapy less than 6 months before the time of enrollment, (e) warfarin anticoagulant

therapy, (f) any investigational chemopreventive agent during the month prior to the biopsies, and (g) a history of bleeding diathesis. Following the release of celecoxib study results showing increased risk of thrombosis with COX-2 inhibitors in 2005, the exclusion criteria were amended to exclude subjects with any of the following:

(h) history of cardiovascular diseases (myocardial infraction, angina, coronary angioplasty, congestive heart failure, stroke, or coronary bypass surgery).

(i) uncontrolled hypertension (>135/>85 mm Hg) on three repeated measurements during the 6 weeks prior to enrollment on the study.

(j) diabetes.

(k) cigarette smoking within the 6 months prior to the date of enrollment.

(l) uncontrolled hypercholesteremia (low-density lipoprotein cholesterol > 130 mg/dl). Hypercholesteremia needed to be controlled following the updated National Cholesterol Education Program Adult Treatment Panel III Guidelines for at least 3 months prior to enrollment on the study. Hypercholesteremia treatment needed to be continued during enrollment on the protocol.

(m) family history of premature coronary disease (i.e. onset <55 years of age).

(n) metabolic syndrome diagnosis in patients 30 years or older. The diagnosis of metabolic syndrome is made when three or more of these risk factors are present:

1. Waist circumference: men >102 cm (>40 in); women >88 cm (>35 in).
2. Triglycerides $\geq$150 mg/dl ($\leq$1.69 mmol/L).
3. High-density lipoprotein cholesterol: men <40 mg/dl (<1.03 mmol/L); women <50 mg/dl (<1.29 mmol/L).
4. Blood pressure $\geq$130/=85 mm Hg

5. Fasting glucose ≥110 mg/dl (=6.1 mmol/L).

(o) history of deep venous thrombosis, pulmonary embolism, systemic lupus erythematosus, family history of protein S or C deficiencies, or prior heparin-induced thrombocytopenia.

Study Design

This single-arm celecoxib study enrolled 47 patients between November 2004 and May 2010. Before celecoxib treatment, patients underwent a history and physical examination and laboratory evaluations including fasting blood glucose and lipid profile (cholesterol, LDL, HDL, and triglycerides), a complete blood cell count, creatinine, bilirubin, and ALT measurement, and a serum pregnancy test for women of childbearing age. Additionally, study patients completed questionnaires regarding medications, vitamins, dietary food frequency, and nutritional supplements, similar to what was previously described (1). A baseline colonoscopy (or sigmoidoscopy in patients who had undergone colectomy) was performed before initiation of celecoxib, and a follow-up colonoscopy/sigmoidoscopy was performed after 6 months of celecoxib treatment. The celecoxib dose was 400 mg by mouth twice daily for 6 months. Blood specimens were collected for analysis of celecoxib serum levels before initiation of celecoxib and at the end of months 2, 4, and 6. Patients were contacted by phone 72 hours after the first celecoxib dose and every 2 weeks thereafter for toxicity assessment for the entire period of the study (6 months). Patients were asked to maintain a diary of celecoxib intake.

Endoscopic Evaluation

In patients with an intact colon, the entire colon was visualized in a continuous clockwise spiral fashion, and images were simultaneously recorded to DVD using proprietary software (Endoworks). The recordings were maintained on separated tracks: cecum-ascending, transverse, descending-sigmoid, and rectum. Following recording of the cecum-ascending segment, indigo

carmine dye spray was applied using a spray catheter. This was performed in the interest of biomarker sampling. In patients with dense polyposis or attenuated FAP, application of dye spray was necessary to identify areas of normal mucosa free of microadenomas or "aberrant crypt foci." It also facilitated sampling of small, flat adenomas that would not likely be counted on white light examination so as to minimize the extent to which mechanical removal of adenomas might introduce systematic bias into the counting of polyps. Care was taken to apply indigo carmine to only a small area of mucosa to minimize downstream staining of the mucosa.

In patients with prior colectomy and ileorectal anastomosis, the rectum was videotaped in a spiral fashion down to the anal verge. To obtain tissue for later studies designed to identify potential molecular targets of celecoxib, 20 forceps biopsy specimens were obtained from representative polyps. Care was taken to avoid removing polyps that would otherwise be counted on post-treatment examination.

Statistical Assessment of Polyp Burden Measures

Given the absence of a gold standard measure for total polyp burden, we focused our analysis on inter-reviewer reliability to see whether different reviewers obtained similar measures of polyp burden for a given patient using our methods. For assessment of inter-reviewer reliability, for each polyp burden measure, we summed the polyp burden across all regions within each patient, which yielded 5 different total polyp burden estimates (1 from each type of measure) for each patient at each time point (before and after celecoxib treatment) from each reviewer. For each polyp burden measure, we then computed the correlations between each pair of reviewers. Higher correlations (closer to 1) indicated better agreement among reviewers, while lower correlations (closer to 0) indicated greater heterogeneity among reviewers. Pearson and Spearman correlations were computed. For the Pearson correlations, we calculated the square

roots of the counts before computing the correlations to stabilize the variance, as is typical for count data. We also computed correlations separately for each colon region—cecum-ascending, transverse, descending-sigmoid, and rectum—to assess whether inter-reviewer reliability varied across regions.

We also assessed inter-reviewer reliability by comparing (1) variability in counts across videos and patients, (2) systematic reviewer variability, and (3) non-systematic reviewer variability. The systematic reviewer variability represents the magnitude of consistent reviewer effects (e.g., if one reviewer consistently tended to give higher counts than the others), while the non-systematic reviewer variability represents variability across reviewers in video counts that varies across videos (e.g., if a reviewer gives higher counts for one video than the others, but lower counts on another video than the others). If inter-reviewer reliability is high, we would expect the reviewer variabilities (systematic and non-systematic) to be smaller in magnitude than the natural variability in counts across videos and patients. To assess these sources of variability, we fit linear mixed models for each polyp burden measure (after a square-root transform to engender normality assumptions), with random effect variance components for the subject effect, the *systematic reviewer effect*, and the residual variance (which can be interpreted as a subject-by-reviewer interaction or *non-systematic reviewer effect*). We can view the subject effect as biological variability and the reviewer-related effects as technical or measurement variability. At each level (subject, reviewer, and residual), we then computed the relative variability at that level by taking the ratio of the respective variance component with their sums. If the relative variability between subjects dominated the relative variability of the systematic and non-systematic reviewer effects, this would suggest high inter-reviewer reliability.

To test whether there was a significant reduction in polyp burden across the patients in the study, we performed a sign test on the change in polyp burden from baseline to 6 months. We computed this test separately for each reviewer and polyp burden measure, and for each polyp burden measure we also computed it in aggregate by taking the mean or median across reviewers.

Reference

1. Shureiqi I, Chen D, Day RS, Zuo X, Hochman FL, Ross WA, et al. Profiling lipoxygenase metabolism in specific steps of colorectal tumorigenesis. Cancer Prev Res (Phila). 2010 Jul;3(7):829-38. PubMed PMID: 20570882. Pubmed Central PMCID: 2900425. Epub 2010/06/24. eng.