Planell N et al. Supplementary Methods

MATERIALS AND METHODS

Microarray data analysis

Microarray raw data was analyzed using Bioconductor tools in R (www.r-project.org). The CG content-adjusted robust multi-array (GC-RMA) algorithm, which computes expression values based on probe intensity values that incorporate probe sequence information, was applied to Affymetrix raw data files in order to normalize them, resulting in a log2 expression value for each probe set. We then employed a conservative probe-filtering step, excluding those probe sets that failed to reach a log2 expression value of 5 in at least 1 sample, which resulted in the selection of a total of 22,206 probe sets out of the original 54,675. Principal component analysis of the given log2 microarray expression data matrix was carried out using basic tools in R (prcomp). Differential expression analysis was assessed by using linear models for microarray data (LIMMA), based on empirical Bayes moderated t-statistics for all filtered probe sets (22,206 probe sets). In addition, F-statistics were used to test for any change in expression over comparisons. To correct for multiple testing, the false discovery rate (FDR) was estimated from p-values derived from the moderated *t*-statistics using the method of Benjamini and Hochberg. Expression values were Z-transformed and empirical bayesian statistics and fold change were stored in an AFM-macro excel file. Genes were considered significantly up- or down-regulated if they had a p-value < 0.02 (FDR-adjusted empirical Bayes (LIMMA), FDR < 0.05) and at least a 1.5-fold variation in mean expression (log2 fc > |0.5849|).

Pathway analysis

Pathway analysis was performed for those genes significantly regulated ($p<0.02$; $FDR<0.05$; log2 fc > |0.5849|) in the affected colonic biopsies from active UC patients compared to non-IBD controls or patients in remission, using Ingenuity Pathways Analysis (Ingenuity™ Systems, www.ingenuity.com). Functional Analysis identified the biological functions and/or diseases that were most significant to the data set. Genes from the dataset that met the cut-off and that were associated with biological functions and/or diseases in the Ingenuity Pathways Knowledge Base were considered for analysis. Fischer's exact test was used to calculate a p-value for determining the probability that each biological function and/or disease assigned to that data set was due to chance alone. Genes whose expression was differentially regulated to a significant degree were overlaid onto a global molecular network developed from information contained in the Ingenuity Pathways Knowledge Base. Networks of these focus genes were then algorithmically generated based on their connectivity.

Quantitative real-time RT-PCR

Total RNA (1 μg) was transcribed to cDNA using reverse transcriptase (High Capacity cDNA Archive RT kit, Applied Biosystems). PCR was performed in TaqMan Universal PCR Master Mix (Applied Biosystems) according to the manufacturer's instructions. Fluorescence was detected in an ABI PRISM 7900 HT Fast Real-Time PCR System (Applied Biosystems). RealTime StatMiner software (Integromics) was used to calculate the cycle threshold (Ct).

Quantitative Real-Time RT-PCR Data Analysis

Real-time RT-PCR raw data was analyzed using Bioconductor tools in R statistical environment. In order to normalize Ct values the DeltaCts (ΔCt = Ct mean of reference gene – Ct target gene) were calculated using two endogenous control genes, UBA3 (ubiquitin-like modifier activating enzyme 3) and RPS3A (ribosomal protein S3A), selected by the Vandesompele method[1]. Missing expression data was imputed using the k-nearest neighbors method on Delta Ct values[2]. A non-parametric Kruskal-Wallis test and a Mann-Whitney-Wilcoxon test were done to examine statistically different expression patterns between groups. To correct for multiple testing, the method of Benjamini and Hochberg was used. An adjusted p-value of <0.05 was considered statistically significant.

Immunostaining of intestinal samples

Paraffin-embedded sections from mucosa colonic biopsies were pre-treated for deparaffinization, rehydration, and epitope retrieval using Dako EnVision Flex Target Retrieval Solution low pH (50x) in conjunction with PT Link (Dako, Glostrup, Denmark), with a warm step of 20 min at 95ºC, for immunohistochemical and dual immunofluorescent staining.

For immunohistochemical staining, sections were blocked with serum (Vectastain ABC kit; Vector Laboratories, USA) for 30 minutes and incubated overnight at 4ºC with commercially available antibodies: REG-IV polyclonal goat antibody (R&D Systems, USA; dilution 1:100), REG-1A polyclonal rabbit antibody (USBiological, USA; dilution 1:600), S100P polyclonal goat antibody (R&D Systems; dilution 1:200) and SERPINB5 monoclonal mouse antibody (BD Biosciences, USA; dilution 1:100). Sections were incubated with 3% H_2O_2 for 10 min in order to block peroxidase activity. Immunohistochemical staining was carried out using DAB chromogen (DAB, Sigma-Aldrich, St. Louis Missouri, USA) in the presence of a peroxidase enzyme (avidin/biotinylated enzyme complex, ABC). The sections were mounted with DPX Mountant for histology (Fluka) and examined with a Nikon Eclipse Ti microscope.

Dual immunofluorescent staining was performed following a simultaneous staining for EpCAM (monoclonal mouse anti-human EpCAM; 1:100, DakoCytomation, M0804) and REG4, REG1A or S100P, or a serial staining for EpCAM and SERPINB5.

For simultaneous staining, sections were blocked with 1% BSA for 30 minutes and incubated different mixtures of two primary antibodies: 1) anti-EpCAM and anti-REG-IV (2 hours at RT), 2) anti-EpCAM and anti-REG-1A (2 hours at RT), or 3) anti-EpCAM and anti-S100P (overnight at 4ºC). Donkey anti-mouse CY5 (Jackson Immunoresearch, Suffolk, UK), donkey anti-rabbit Alexa 488 (Invitrogen), and donkey anti-goat Alexa 488 (Invitrogen) were used as secondary antibodies. For serial staining, sections were blocked with 1% BSA for 30 minutes and incubated first with the primary (anti-SERPINB5, overnight at 4ºC) and secondary antibodies(Goat anti-mouse Alexa 488 (Invitrogen)), followed by a blocking step with 1% BSA for 30 minutes. Sections were then incubated with an anti-EpCAM mAb (2 hours at RT) followed by a secondary antibody (Donkey anti-mouse CY5).

Following immunostaining, all sections were mounted with Vectashield Mounting Medium with DAPI (Vector Laboratories) and examined with an Olympus DP72 microscope.

Regarding negative controls, specimens were processed under the same conditions in the absence of the corresponding primary antibodies.

REFERENCES

1. Vandesompele J, De Preter K, Pattyn K et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002; 3: 1-12.
2. Troyanskaya O, Cantor M, Sherlock G et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.