

# Supplementary Information

**Table S1** Information on volunteers and time points at which faecal samples were collected. Coding of the disease score is based on Walmsley, R.S. et al. A simple clinical colitis index. *Gut* **1998**, 43, 29-32.

Group and subject	Age	Sex	Disease score	Sample time points (months) and sample codes			
				0 mo.	6 mo.	12 mo.	extra
UC1	56	F	1	1-1	1-2	1-3	-
UC3	51	M	1	3-1	3-2	3-3	3-F <sup>a</sup>
UC4	56	M	1	4-1	4-2	4-3	-
UC5	26	F	4	5-1	5-2	-	-
UC6	58	F	2	6-1	-	-	-
UC7	60	F	3	7-1	7-2	7-1F <sup>a</sup>	7-2F <sup>a</sup>
UC8	55	F	6	8-1	8-2	8-3	8-F <sup>a</sup>
UC10	32	M	2	10-1 <sup>b</sup>	10-2	-	-
UC11	54	F	1	11-1	11-2	11-3	-
UC13	63	M	2	13-1	13-2	-	-
UC14	29	M	2	14-1	-	-	-
UC15	34	F	1	15-1	15-2	-	-
UC16	63	F	1	16-2	-	-	-
IBS1	39	F	NA	1-1	-	-	-
IBS2	36	M	NA	2-1	-	-	-
IBS3	57	F	NA	3-1	3-2	3-3	-
IBS4	55	F	NA	4-1	4-2	4-3	-
IBS5	41	M	NA	5-1	5-2	5-3	-
IBS6	42	F	NA	6-1	6-2	-	-
IBS7	53	F	NA	7-1	7-2	7-3	-
IBS10	35	M	NA	10-1	-	-	-
IBS14	42	M	NA	14-1	14-2	-	-
IBS16	41	F	NA	16-1	16-2	-	-
C1	36	F	NA	1-1	-	-	-
C3	46	F	NA	3-1	3-2	3-3	3-4
C4	54	F	NA	4-1	4-2	4-3	4-4
C5	37	M	NA	5-1	5-2	5-3	5-4
C8	52	F	NA	8-1	8-2	8-3	8-4
C9	37	F	NA	9-1	9-2	9-3	9-4
C10	46	F	NA	10-1	10-2	10-3	10-4
C11	55	F	NA	11-1	11-2	11-3	11-4
C12	40	F	NA	12-1	12-2	12-3	-
C13	60	F	NA	13-1	13-2	13-3	-
C14	39	F	NA	14-1	14-2	14-3	14-4
C15	43	M	NA	15-1	15-2	15-3	15-4
C16	38	F	NA	16-1	16-2	-	-
C17	55	M	NA	17-1	17-2	-	-
C18	56	F	NA	18-1	18-2	18-3	18-4
C19	58	M	NA	19-1	19-2 <sup>b</sup>	-	19-4
C21	40	M	NA	21-1	21-2 <sup>b</sup>	21-3	-
C22	55	F	NA	22-1	22-2	22-3	22-4
C24	60	F	NA	24-1	24-2	24-3	-
C25	37	F	NA	25-1	25-2	25-3	25-4
C26	36	M	NA	26-1	26-2	26-3	-
C27	44	M	NA	27-1	27-2	27-3	-

<sup>a</sup> Flare up sample

<sup>b</sup> Excluded because of poor spectral quality

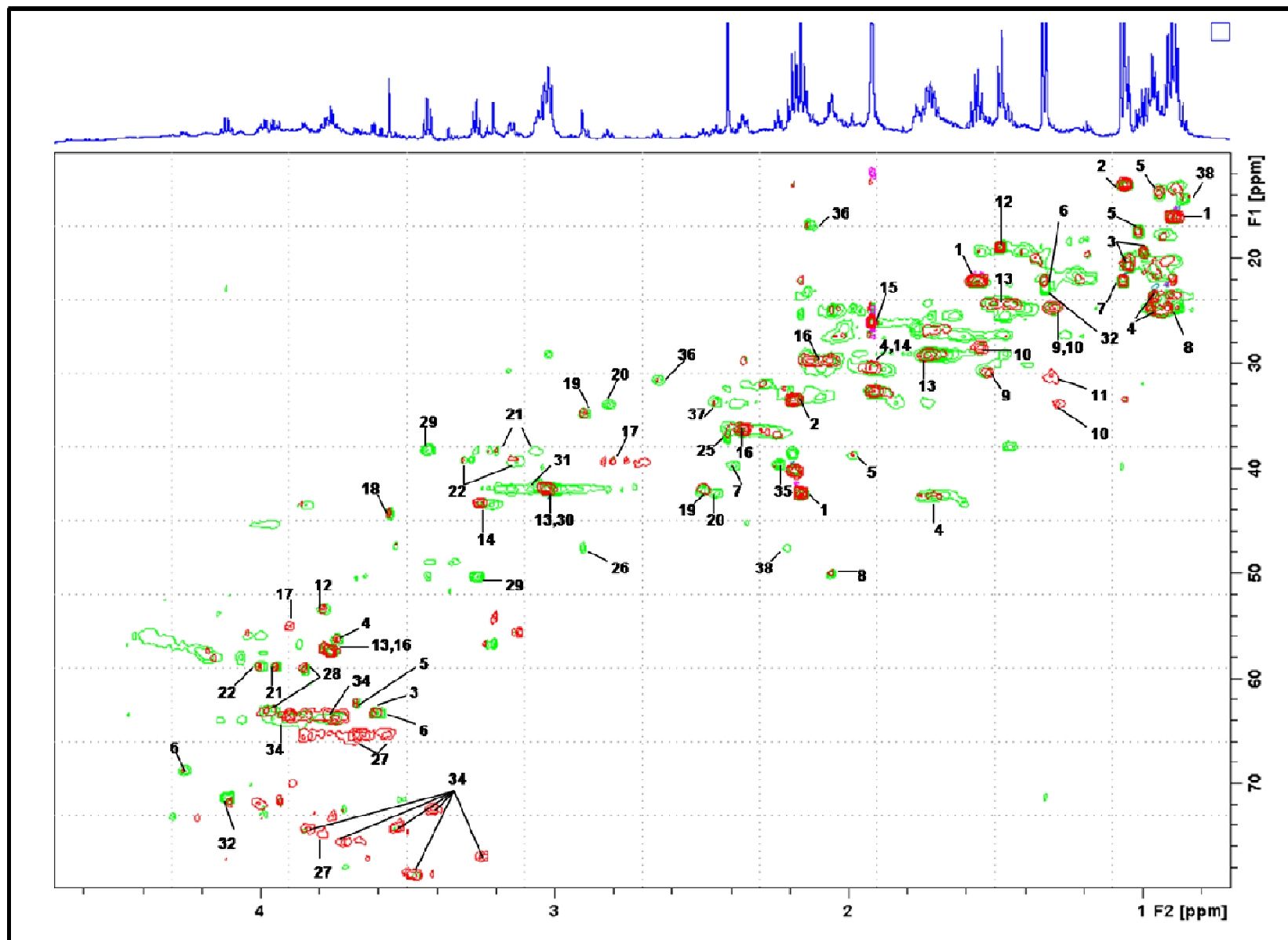
**Table S2** Excel file of NMR buckets used in all data analyses involving UC samples. Each row corresponds to one bucket; first and second columns give upper and lower limits (ppm) of each variable width bucket.

**Table S3** Excel file of NMR buckets used in data analyses *not* involving UC samples. Each row corresponds to one bucket; first and second columns give upper and lower limits (ppm) of each variable width bucket.

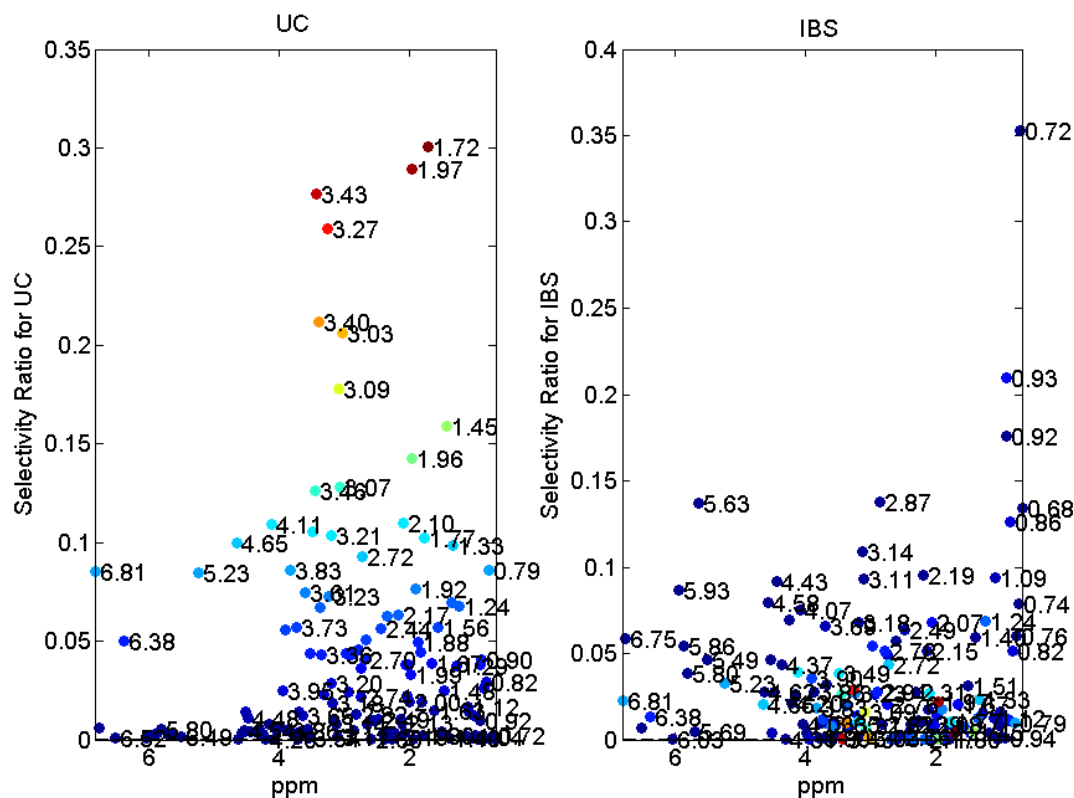
## 2D NMR methods

Standard Bruker pulse sequences (with gradients and water presaturation) were used with long total acquisition times to detect minor signals: COSY *cosygpprqf*, data points 4096x512, spectral width 12 ppm x 12 ppm, 64 scans per  $t_1$  increment; HSQC *hsqcetgpprsisp2.2* 2048 x 256, 12 ppm ( $^1\text{H}$ ) x 165.7 ppm ( $^{13}\text{C}$ ), 192 scans; HMBC *hmbcgplndprqf* 2048 x 256, 12 ppm ( $^1\text{H}$ ) x 250 ppm ( $^{13}\text{C}$ ), 272 scans. Spectra were processed with zero filling in both dimensions and standard windows (unshifted SINE for magnitude mode COSY, JRES and HMBC; 90° phase shifted QSINE for phase sensitive HSQC).

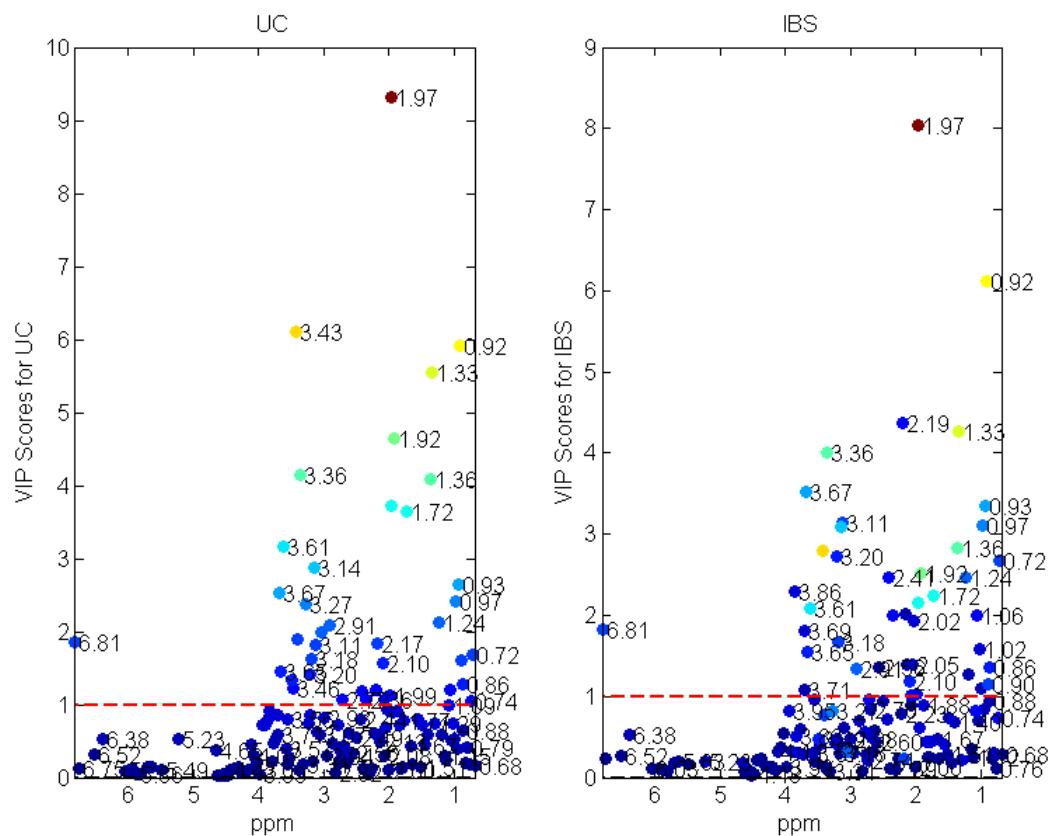
**Figure S1** Superimposed 2D NMR spectra ( $^1\text{H}$ - $^{13}\text{C}$  HSQC) of two typical samples (UC green; C red). Numbering of peaks corresponds to Figure 1 and Table 2. Some low concentration components not shown here were identified by COSY and are also listed in Table 2.



**Figure S2** Selectivity ratios for variables (buckets) in the three group, 11LV PLS-DA model for the UC and IBS groups (see Figures 2c and 2d in main text).



**Figure S3** VIP scores for variables (buckets) in the three group, 11LV PLS-DA model for the UC and IBS groups (see Figures 2c and 2d in main text).



**Figure S4** Faecal calprotectin measurements for C (Control), UC and IBS samples

