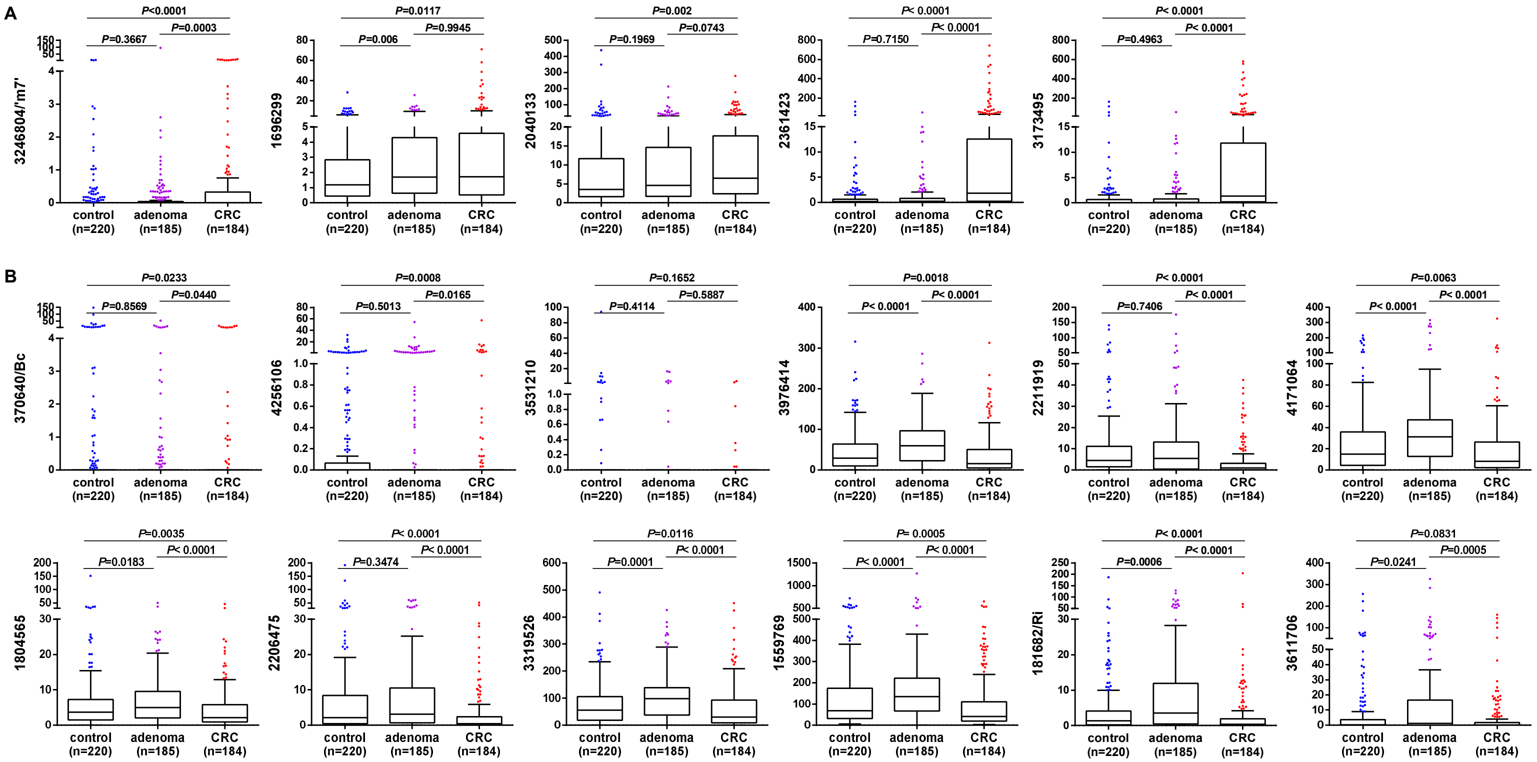
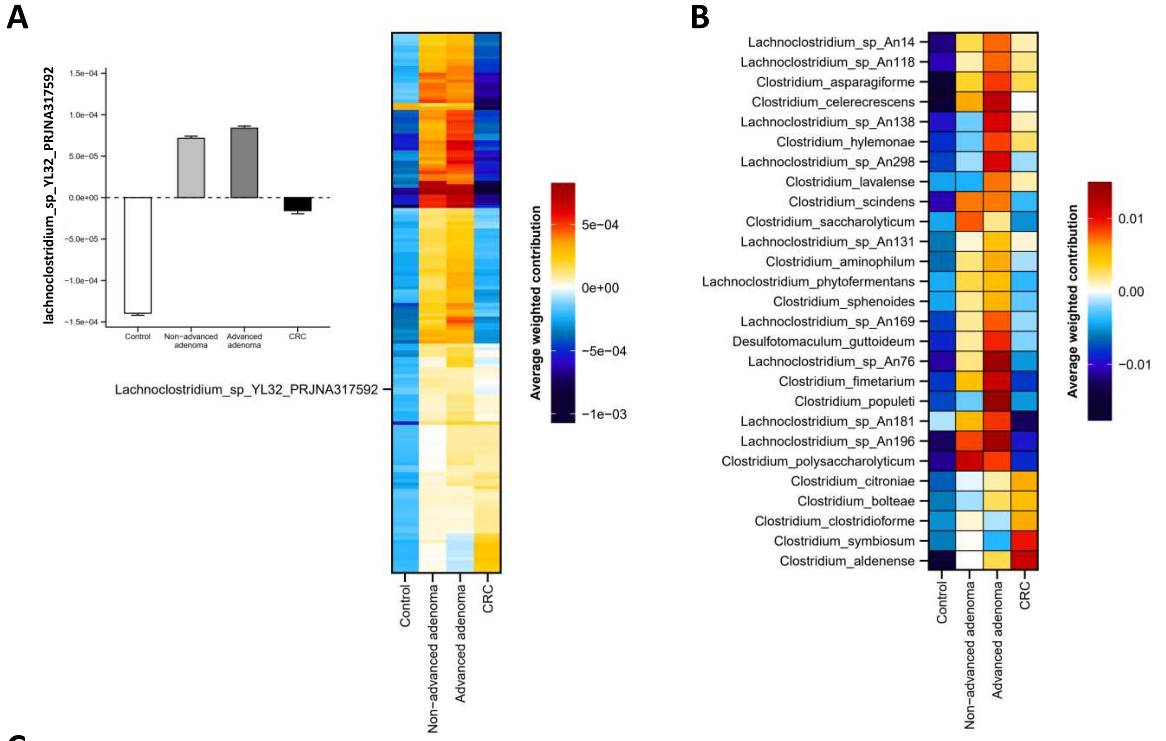
Supplementary Figure 1

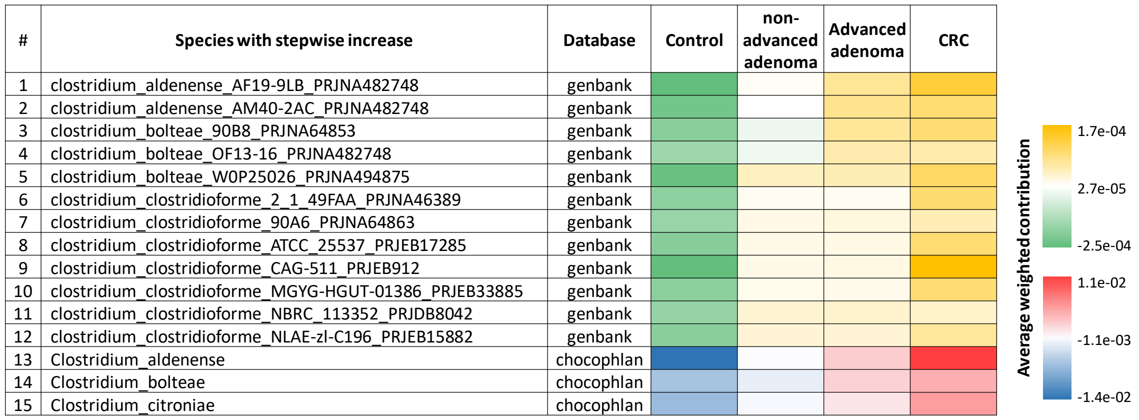


Supplementary Figure 1 Fecal abundances of 17 previously identified bacterial gene markers in the new cohort of CRC patients, adenoma patients and control subjects by metagenomics analysis. (**A**) Abundances of five bacterial gene markers that were previously identified to be enriched in stool samples of CRC patients as compared to control subjects. Marker 1696299 was significantly increased in adenoma patients as compared to control subjects, but it was not good enough to serve as a CRC marker (AUROC= 0.5604 for distinguishing CRC patients from control subjects) and therefore was not included in this study. (**B**) Abundances of 12 bacterial gene markers that were previously identified to be enriched in stool samples of control subjects as compared to CRC patients. Comparisons were conducted by Mann Whitney tests.

Supplementary Figure 2



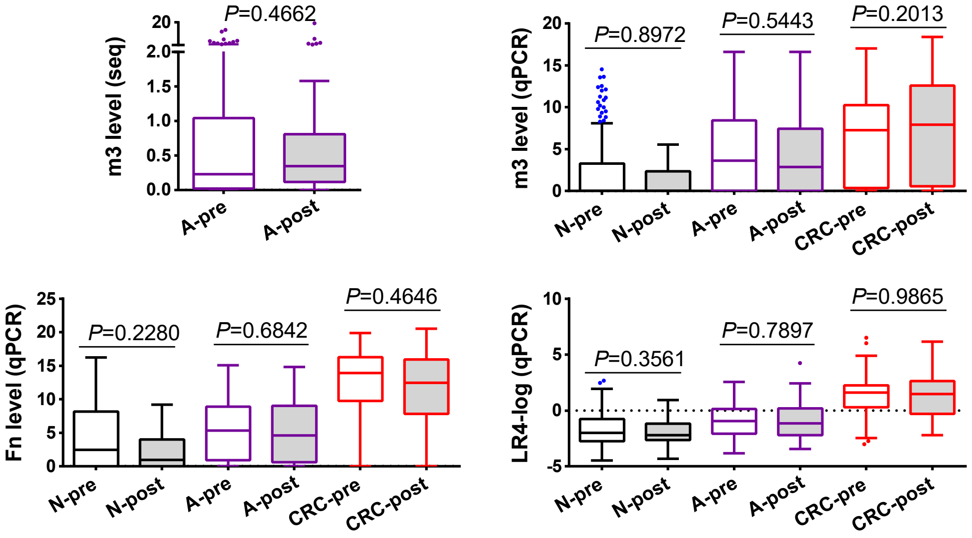
C



#	Species with stepwise increase	Database	Control	non-advanced adenoma	Advanced adenoma	CRC
1	clostridium_aldenense_AF19-9LB_PRJNA482748	genbank	🟩	⬜	🟨	🟨
2	clostridium_aldenense_AM40-2AC_PRJNA482748	genbank	🟩	⬜	🟨	🟨
3	clostridium_bolteae_90B8_PRJNA64853	genbank	🟩	🟩	🟨	🟨
4	clostridium_bolteae_OF13-16_PRJNA482748	genbank	🟩	🟩	🟨	🟨
5	clostridium_bolteae_W0P25026_PRJNA494875	genbank	🟩	🟨	🟨	🟨
6	clostridium_clostridioforme_2_1_49FAA_PRJNA46389	genbank	🟩	🟨	⬜	🟨
7	clostridium_clostridioforme_90A6_PRJNA64863	genbank	🟩	🟨	🟨	🟨
8	clostridium_clostridioforme_ATCC_25537_PRJEB17285	genbank	🟩	🟨	🟨	🟨
9	clostridium_clostridioforme_CAG-511_PRJEB912	genbank	🟩	🟨	🟨	🟨
10	clostridium_clostridioforme_MGYG-HGUT-01386_PRJEB33885	genbank	🟩	🟨	🟨	🟨
11	clostridium_clostridioforme_NBRC_113352_PRJDB8042	genbank	🟩	🟨	🟨	🟨
12	clostridium_clostridioforme_NLAE-zl-C196_PRJEB15882	genbank	🟩	🟨	🟨	🟨
13	Clostridium_aldenense	chocophlan	🟦	⬜	🟥	🟥
14	Clostridium_bolteae	chocophlan	🟦	🟦	🟥	🟥
15	Clostridium_citroniae	chocophlan	🟦	⬜	🟥	🟥

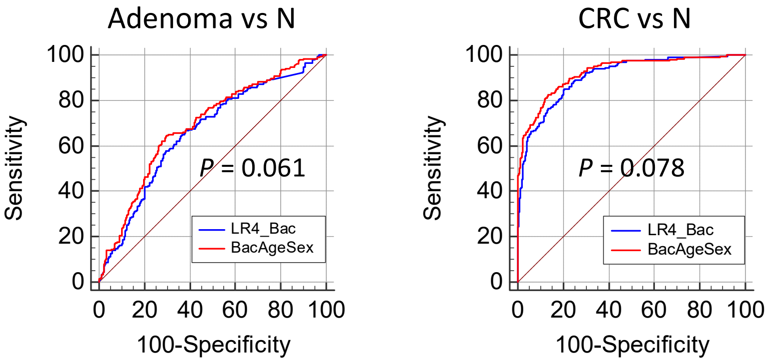
Supplementary Figure 2 Differental abundance analysis of *Lachnoclostridium* genomes in our in-house metagenomics data. (**A**) Heatmap summarizing the average weighted contributions (AWCs) of 160 *Lachnoclostridium* genomes from the NCBI GenBank (release 234.0; accessed October 16, 2019), and an individual bar-chart showing AWCs of *Lachnoclostridium sp. YL32*. (**B**) Heatmap summarizing the average weighted contributions of 27 *Lachnoclostridium* species from the ChocoPhlAn pangenome database (v293). (**C**) Fifteen *Lachnoclostridium* strains with stepwise increase in AWCs from control to adenoma to cancer.

Supplementary Figure 3



Supplementary Figure 3 There was no significant difference in m3 level between pre- and post-colonoscopy adenoma samples by metagenome sequencing (top left) or qPCR (top right). Comparison of the qPCR results showed that there were no difference in Fn, m3 or the combined score (4Bac) between pre-colonoscopy and post-colonoscopy samples of the control, adenoma or CRC groups (top right and bottom panels).

Supplementary Figure 4



Supplementary Figure 4 Pairwise comparison of the ROC curves by the logistic regression models based on four bacterial markers (LR4_4Bac) and on four bacterial markers plus age and gender (BacAgeSex) respectively.