

**Captions for Supplementary Tables for**  
**“Patients with mesenchymal tumours and high**  
***Fusobacteriales* prevalence have worse prognosis in**  
**colorectal cancer (CRC)”**

Manuela Salvucci<sup>1</sup>, Nyree Crawford<sup>2</sup>, Katie Stott<sup>2</sup>, Susan Bullman<sup>3,4</sup>, Daniel B. Longley<sup>2</sup>, and  
Jochen H.M. Prehn<sup>1\*</sup>

<sup>1</sup>Centre for Systems Medicine, Department of Physiology and Medical Physics, Royal College of Surgeons in Ireland, Dublin, Ireland;

<sup>2</sup>Patrick G. Johnston Centre for Cancer Research, School of Medicine, Dentistry and Biomedical Science, Queen’s University Belfast, Northern Ireland, UK;

<sup>3</sup>Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA;

<sup>4</sup>Fred Hutchinson Cancer Research Center, Human Biology Division, Seattle, USA.

**Corresponding author:** Prof. Jochen H. M. Prehn, Department of Physiology and Medical Physics, Royal College of Surgeons in Ireland, 123 St. Stephen’s Green, Dublin 2, Ireland. Tel.: +353-1-402-2255; Fax: +353-1-402-2447; E-mail: [jprehn@rcsi.ie](mailto:jprehn@rcsi.ie).

**Data and code availability:** Datasets and source code will be publicly available and archived upon publication at Zenodo (<https://10.5281/zenodo.4019142>).

Transcriptomic-dependent *Fn/Fusobacteriales* impact.**Supplementary Table 1.**

Clinico-pathological and demographic characteristics of the CRC patients included in this study (“Overall”) and grouped by cohort, namely “in house Taxonomy” and “TCGA-COAD-READ”. For continuous variables, median, interquartile range, and statistical significance (P-value) determined by Kruskal-Wallis tests are reported. For categorical values, number, and percentage of cases by level and statistical significance (P-value) determined by  $\chi^2$  tests are reported.

**Supplementary Table 2.**

Patient-level bacterium data for cases of the Taxonomy and TCGA-COAD-READ cohort. *Fn* load measured by qPCR for patients of the Taxonomy cohort is available in the sheet “Taxonomy cohort (n=140)”. Relative abundance of *Fusobacteriales* and higher resolution taxonomic ranks (family, genus and species) including the *Fn* species for the patients in the TCGA-COAD-READ cohort is available in the sheet “TCGA-COAD-READ cohort (n=605)”. For the TCGA-COAD-READ cohort, genres/species with an average relative abundance lower than 0.05 were aggregated as “Other”.

**Supplementary Table 3.**

Association between mutational status and *Fusobacteriales* relative abundance in the TCGA-COAD-READ patients. Statistical significance was assessed by  $\chi^2$  independence tests and  $\chi^2$  statistics, unadjusted- and FDR-corrected mod-likelihood P-values are reported for each mutation that was either selected *a priori* or was found to be statistically significant altered when comparing *Fusobacteriales*-low vs. -high patients (75<sup>th</sup> percentile cut-off) of the TCGA-COAD-READ cohort.

Salvucci et al.

Transcriptomic-dependent *Fn/Fusobacteriales* impact.**Supplementary Table 4.**

Association between recurrent copy number aberrations identified by GISTIC analysis when comparing *Fusobacteriales*-low vs. -high patients (75<sup>th</sup> percentile cut-off) of the TCGA-COAD-READ cohort.

**Supplementary Table 5.**

Association between gene expression profiles and *Fusobacteriales* relative abundance in the TCGA-COAD-READ patients was assessed by Spearman correlation. Correlation coefficient R, corresponding 95% confidence intervals, unadjusted- and FDR-corrected P-values are reported for each protein that was found to be statistically significant altered in the TCGA-COAD-READ cohort.

**Supplementary Table 6.**

Association between protein expression profiles and *Fusobacteriales* relative abundance in the TCGA-COAD-READ patients was assessed by Spearman correlation. Correlation coefficient R, corresponding 95% confidence intervals, unadjusted- and FDR-corrected P-values are reported for each protein that was found to be statistically significant altered in the TCGA-COAD-READ cohort.

**Supplementary Table 7.**

Un-adjusted and adjusted Cox regression models for patients of the TCGA-COAD-READ cohort. Cox regression models were fitted with an interaction term between *Fusobacteriales* (high vs. low, using the 75<sup>th</sup> percentile relative abundance as cut-off) and mesenchymal status

Salvucci et al.

### Transcriptomic-dependent *Fn/Fusobacteriales* impact.

(mesenchymal vs. non-mesenchymal). Adjusted model 1 and 2 were fitted including precision variables. Model 1 used as adjustment covariates key clinical-pathological characteristics, namely age (continuous), stage (categorical, I to IV), tumour location (categorical, colon vs. rectum) and sex (categorical, male vs. female). Model 2 used as adjustment covariates a more extensive set (i. e. super-set) of clinico-pathological characteristics additionally including history of colon polyps (categorical, yes vs. no) and history of other malignancy as comorbidities.

### Supplementary Table 8.

Detailed statistical output (coefficients and P-values) of logistic models 1 (*Fusobacteriales~gene/signature*) and 2 (*Fusobacteriales~gene/signature:molecular subtype*) fitted for a set of hypothesis-driven gene/signature profiles in patients of the TCGA-COAD-READ cohort presented in **Fig. 6A**. Table include all genes/signatures tested (regardless of statistical significance) reported in ascending order of interaction P-values from model 2 determined in the TCGA-COAD-READ cohort (discovery cohort). For completeness, detailed statistical output is also reported for the Taxonomy cohort.

Salvucci et al.