

## SUPPLEMENTARY MATERIAL

### **Supplementary Methods**

#### *Bacterial DNA contamination assessment*

To assess the potential bacterial DNA contamination from environment and reagents, several negative controls were included and analysed throughout the 16S rRNA gene sequencing pipeline. Ten negative controls consisting of ultrapure water (NC-EXT-WA) were added at the DNA extraction step to account for reagent and environmental contaminants. NC-EXT-WA samples were processed concomitantly with the DNA extraction from whole blood samples. This study comprised 296 whole blood samples. 16 rounds of DNA extraction were performed adding 2 NC-EXT-WA samples to the extraction round 1 (EXT1), 5 (EXT5), 10 (EXT10), 15 (EXT15) and 16 (EXT16).

To overcome the unbalance between the whole blood sample and the control groups (296 vs 10), we performed 10 randomized statistical analyses by comparing the 10 NC-EXT-WA samples to 10 randomly selected whole blood samples. The beta-diversity was measured through Bray-Curtis index. PERMANOVA and PERMDISP were used to test for homogeneity of multivariate dispersions on Bray-Curtis dissimilarity table for the contamination control.

We used diversity analyses and comparisons of the taxonomic profiles in order to confirm that the differences between DNA extraction negative controls and whole blood samples were large enough. Beta-diversity indices were used to investigate whether blood bacterial DNA differs from that of extraction negative controls. Based on all beta-diversity indices, especially on the Bray-Curtis distance matrix (Supplementary Figure 4), blood bacterial DNA profiles were separated from the negative controls in all the randomized analyses. Moreover, blood samples tended to form a tighter cluster than the negative controls, suggesting that the blood samples shared a specific and more similar DNA bacterial profile. The negative control group was more dispersed, showing more variability in-between samples and suggesting a random profile more evident. PERMANOVA showed statistically significant differences between whole blood samples and negative controls for Bray-Curtis beta-diversity indices (Supplementary Table 2), suggesting that the bacterial

communities differed both in composition and relative abundance between biological and negative control samples. PERMDISP confirmed the results.

We could, therefore, consider that the background noise did not impact the results of the study.

*Supplementary Table 1* Distributions of Observed, Chao1, Shannon and Simpson alpha-diversity indices of control, intestinal adenoma (IA), colon and rectal cancer subjects for bacterial genera and OTUs.

*Supplementary Table 2:* PERMANOVA and PERMDISP on beta-diversity Bray-Curtis index for the 10 randomized comparisons between negative controls (CTRL) and whole blood samples (WB). For all randomized comparisons, the sample size is equal to 20 (10 Neg CTRL vs. 10 WB) at the exception of RandExt16 (10 Neg CTRL vs. 9 WB). The statistical testing used were pseudo-F and F-value for PERMANOVA and PERMDISP, respectively. For both tests, 2000 permutations were performed. P-values <0.05 in yellow, <0.005 in orange and <0.0005 in red.

*Supplementary Figure 1.* Distribution of 16S rRNA gene copies per  $\mu$ l of whole blood among controls, intestinal adenomas and colorectal cancers.

*Supplementary Figure 2.* Mean of the taxonomy composition in terms of relative abundance per group of samples at family taxonomic level. The panel on the left shows the taxonomic composition of controls, intestinal adenoma subjects (IA) and both controls/IA, respectively (as per legend to the right side). The panel on the right shows the taxonomic composition of colorectal cancer (CRC), and rectal and colon cancer cases separately (as per legend to the right side).

*Supplementary Figure 3.* Different taxa between: (A) colorectal cancers (CRC) and controls; (B) CRC and intestinal adenoma (IA); (C) IA and controls by DESeq2 analyses. The taxonomic lineage of each taxon is shown: p, phylum; c, class; o, order; f, family; g, genus; OTU#, Operational Taxonomic Unit. The first two columns show the logarithmic transformation of normalized base mean value for each group. The “padj” column shows the p-value for heterogeneity between groups adjusted for multi-testing analyses. Positive fold changes (shown on a green background) designate taxon overrepresentation in the CRC group. Negative fold changes (shown on a yellow background) designate taxon underrepresentation in the CRC group.

*Supplementary Figure 4.* Multidimensional scaling (MDS) differentiates bacterial patterns of study blood samples and DNA extraction negative controls on Bray-Curtis distance matrices for all 10 randomized comparisons. Study blood samples (in Red) separated from negative controls (in Blue) in all of the randomized comparisons.

**Supplementary Table 1** Distributions of Observed, Chao1, Shannon and Simpson alpha-diversity indices of control, intestinal adenoma (IA), colon and rectal cancer subjects for bacterial genera and OTUs.

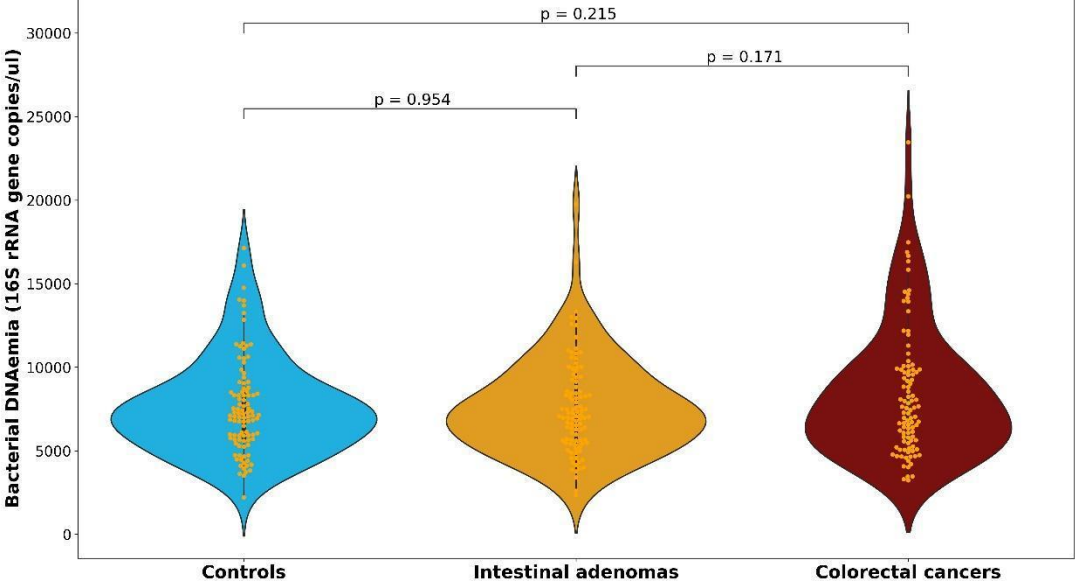
	<i>Alpha-diversity</i>	Median (I-III quartiles)				p* Controls vs IA	p* Colon cancer vs IA	p* Colon cancer vs controls	p* Rectal cancer vs IA	p* Rectal cancer vs controls	p* Colon cancer vs rectal cancer
		Controls	IA	Colon cancer	Rectal cancer						
<b>Genera</b>	<i>Observed</i>	28.5 (25-32)	28 (25-32)	29.5 (25-36)	30 (26-32)	0.917	0.246	0.251	0.364	0.332	0.691
	<i>Chao1</i>	44.9 (33-57.5)	44.5 (34-55.5)	44.5 (36.3-55.3)	41.3 (35-53)	0.686	0.874	0.774	0.617	0.879	0.493
	<i>Shannon</i>	2.34 (2.02-2.53)	2.33 (2.15-2.51)	2.33 (2.00-2.58)	2.32 (2.12-2.51)	0.575	0.824	0.706	0.958	0.540	0.723
	<i>Simpson</i>	0.87 (0.80-0.89)	0.86 (0.83-0.89)	0.86 (0.81-0.90)	0.86 (0.82-0.89)	0.843	0.808	0.884	0.855	0.792	0.981
<b>OTUs</b>	<i>Observed</i>	34 (31-39)	35 (31-42)	37 (32-47)	37 (33-40)	0.542	0.208	0.105	0.440	0.171	0.617
	<i>Chao1</i>	62.1 (48-85)	66 (49.2-90.5)	67.1 (51.0-87.5)	59.9 (49.6-82)	0.297	0.631	0.594	0.299	0.968	0.702
	<i>Shannon</i>	2.58 (2.28-2.77)	2.58 (2.40-2.80)	2.61 (2.27-2.87)	2.51 (2.41-2.74)	0.372	0.874	0.496	0.625	0.667	0.581
	<i>Simpson</i>	0.90 (0.86-0.92)	0.90 (0.87-0.92)	0.90 (0.86-0.92)	0.89 (0.88-0.91)	0.671	0.879	0.749	0.241	0.482	0.378

\* P for heterogeneity estimated from the Wilcoxon rank-sum  
SD: standard deviation; IQ: interquartile

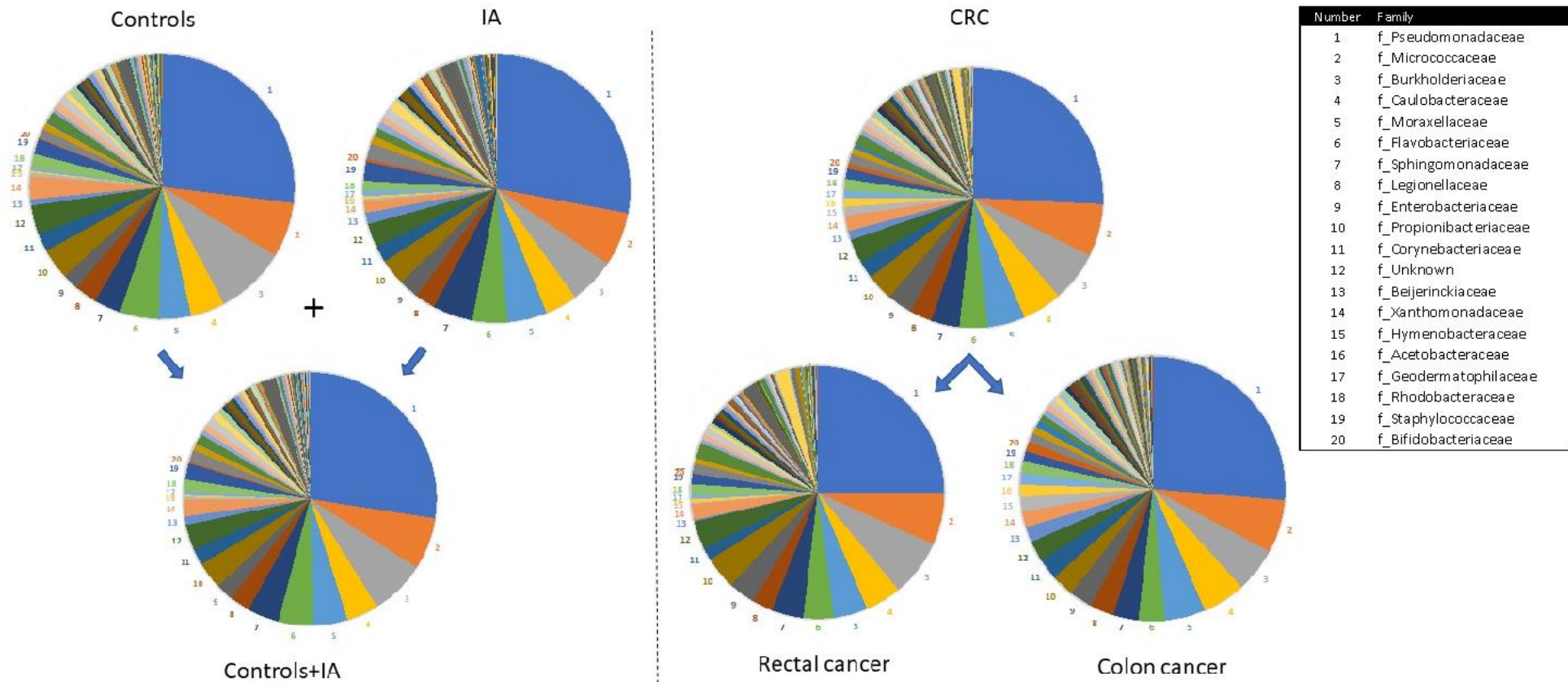
**Supplementary Table 2:** PERMANOVA and PERMDISP on beta-diversity Bray-Curtis index for the 10 randomized comparisons between negative controls (CTRL) and whole blood samples (WB). For all randomized comparisons, the sample size is equal to 20 (10 Neg CTRL vs. 10 WB) at the exception of RandExt16 (10 Neg CTRL vs. 9 WB). The statistical testing used were pseudo-F and F-value for PERMANOVA and PERMDISP, respectively. For both tests, 2000 permutations were performed. P-values <0.05 in yellow, <0.005 in orange and <0.0005 in red.

			RandExt01	RandExt05	RandExt10	RandExt15	RandExt16	RandAllExt	Rand1	Rand2	Rand3	Rand4
Bray-Curtis	PERMANOVA	Test statistic	3,49	4,51	3,51	4,01	4,37	3,03	4,08	3,22	2,94	2,82
		P-value	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005
	PERMDISP	Test statistic	10,67	29,79	15,52	13,87	28,65	9,96	15,13	18,59	8,28	3,60
		P-value	0,0050	0,0005	0,0025	0,0005	0,0005	0,0130	0,0005	0,0015	0,0190	0,0900

**Supplementary Figure 1.** Distribution of 16S rRNA gene copies per  $\mu\text{l}$  of whole blood among controls, intestinal adenomas and colorectal cancers.



**Supplementary Figure 2.** Mean of the taxonomy composition in terms of relative abundance per group of samples at family taxonomic level. The panel on the left shows the taxonomic composition of controls, intestinal adenoma subjects (IA) and both controls/IA, respectively (as per legend to the right side). The panel on the right shows the taxonomic composition of colorectal cancer (CRC), and rectal and colon cancer cases separately (as per legend to the right side).



**Supplementary Figure 3.** Different taxa between: (A) *colorectal cancers (CRC) and controls*; (B) *CRC and intestinal adenoma (IA)*; (C) *IA and controls* by DESeq2 analyses. The taxonomic lineage of each taxon is shown: p, phylum; c, class; o, order; f, family; g, genus; OTU#, Operational Taxonomic Unit. The first two columns show the logarithmic transformation of normalized base mean value for each group. The “padj” column shows the p-value for heterogeneity between groups adjusted for multi-testing analyses. Positive fold changes (shown on a green background) designate taxon overrepresentation in the CRC group. Negative fold changes (shown on a yellow background) designate taxon underrepresentation in the CRC group.





**Supplementary Figure 4.** Multidimensional scaling (MDS) differentiates bacterial patterns of study blood samples and DNA extraction negative controls on Bray-Curtis distance matrices for all 10 randomized comparisons. Study blood samples (in Red) separated from negative controls (in Blue) in all of the randomized comparisons.

