## 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 µl 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.

	Alpha-diversity	Median (I-III quartiles)				$\mathbf{p}^*$	$p^*$	$p^*$	$\mathbf{p}^*$	$\mathbf{p}^*$	$p^*$
		Controls	IA	Colon cancer	Rectal cancer	Controls vs IA	Colon cancer vs IA	Colon cancer vs controls	Rectal cancer vs IA	Rectal cancer vs controls	Colon cancer vs rectal cancer
Genera	Observed	28.5 (25-32)	28 (25-32)	29.5 (25-36)	30 (26-32)	0.917	0.246	0.251	0.364	0.332	0691
	Chao1	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
	Shannon	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
	Simpson	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
OTUs	Observed	34 (31-39)	35 (31-42)	37 (32-47)	37 (33-40)	0.542	0.208	0.105	0.440	0.171	0.617
	Chao1	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
	Shannon	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
	Simpson	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  $\mu$ 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.

CRC Co	ontrols	#OTU	Tai, ono my	ondi	los2 Fold Change
			p_Actino tacte ris x_Acitimic to bila	31-02	4,92
			p_Actinobacte ris x_Actinobacte ris xo_Bifid obacte ris les;†_Bifidobacte riscese	61-09	10,21
			p_Actinobacteria x_Actinobacteria xo_Bifd obacteria tes, T_Bifd obacteria cea e x_Bifd obacterium	61-09	10,21
		0701.00	p, Actino backerne ∠, Actino backerne jo _Conyne backerne is z _Conyne backerne backerz _Conyne backernum n. Actino backerne z _Actino backeria z _Conyne backeria is z _Conyne backeria _Conyne backerium z _Conyne backerium z r	91-05	-7,33
		0.0_04	a Actino tacteria y Actino tacteria yo Frankia testi Geodermeto a hibizeae	21-04	4.32
			p_Actino tacte ris x_Actino bacte ris x_Micrococcas les;f_Micrococcascese;g_Rothis	SE-06	4,45
			p_Actino teacterie x_Actino bacterie xo_o_Bifido bacterie les	61-09	10,21
			p_Actinobacte risy_Actinobacte risyo_Propionibacte ristes;*_Nocas raiobia aceae	21-02	-4,17
			p_actino decte reix_ intermo teo politite n. Rectero identes x. Rectero idina io . Rectero identes f. Rectero identes	26-06	6.77
			p Bacteroidetesy: Bacteroid is yo Bacteroid a les; f Bacteroidaceae; g Bacteroides	21-09	-6,77
			p_Bacteroidetes;c_Bacteroid ia;o_S phingotacteriates	98-06	-6,42
			p_Bacte roide tesy_Bacte roid is yo_S phingo tacte rists; f_S phingo tacte riscess	11-04	-3,39
		OTU_62	p_Firmicutesy_Clostridie yo_Clostridie les;f_Clostridie ceaefitg_Clostridium & ensu& trictofits_Multi-affilietion 	98-05	-5,04
		OTU 49	p Firmicutes): Clostrial algo Clostrial alles). Peptos reproducede en Rombo utales metage nome	11-06	9,20
		-	p_Prote obsecte risy:_Alpha prote o bacte risy:_Ace to bacte risks	56-06	7,38
			p_Proteobacteria;c_Alpina proteo tacteria;o_Acetotacteria les;f_Acetobacteria cene	58-06	7,33
		OTU_44	p_Protechadze na y_Alpha protechadze na y_Canu lobadze na les (f_Canulobadze na cenery), Canulobadze na y_Un known A Rostechadze na y_Alpha protechadze na you can biometric and a red Stabilizatione reduced a subscriptione biome	18-02	-3, 53
		0701.42	p_Protectadade mayAlpha protectadede into sphingsmonadales / Sphingomonadadeae.g_roves prings alum p_Protectade mayAlpha protectade (exposition of the state of the sphing on the state of the sphing of the stat	21-02	-124
		0.0_4	p Protochacteria: De ha protochacteria: Maccocanies	21-08	4,46
			p_Proteobacteria;c_Eeita proteo tacteria;o_Oligofexales;f_Oligofexaceae	91-05	7,86
		OTU_37	p_Prote obsche nisz_Gamma prote obsche niszo_Bata prote obsche nisks, f_Burkhold aniscene ag_Cuprievid us z_Multi-affiliation	28-07	-7,02
			p_Proteconscience,_Gemme proteconscience, Bene proteconscience in Est_Burknond e macene,g_kaston m n. Proteconscience, Gemme proteconscience, Bene proteconscience is lest. Ne isse riscense	46-02	-4,71
			p Proteobacterina y Gamma proteobacterina yo Bena proteobacterina les f Neisserina cenergi Neisserina	21-02	4,12
			p_Prote obacteria x_Gamma prote obacteria xo_Enterobacteria les ;1_Enterobacteria ceae,g_Pantoea	21-04	7,73
		OTU_42	p_Prote obacteria x_Gamma prote obacteria xo_legione lla les,f_legione lla ceae,g_legione lla s_Unknown	3 E-08	4,36
		010_22	p_Proteobacteria;c_Gamma proteobacteria;o_Legione la les;f_Legione la ceae.g_Legione la z_Unknown Roteobacteria: Gamma proteobacteria: Roteobacteria da Manacateria da Manacateria da Santa da Manacateria da Man	31-08	5,35
		010_46	p_proteconscients;c_Gamma proteconscients;o_Pseudomonadales;f_monake increas;g_acineto uncter;s_muntramilation p_proteconscients;: Gamma proteconscients;o_Pseudomonadales;f_Monase increas; Acineto uncter;s_Multianfiliation	28-15	10,64
		OTU 30	p Proteobacteria: Gamma proteobacteria:o Reudomonada las f Moravella za ez Acineto bacters. Multi-affiliation	21-02	4.13
14	CRC	#OTU	Tax.ono my	padj	log2 Fold Change
		OTU_48	p_Actino bacteria x_Actino bacteria xy_Cory ne bacteria les,f_Cory ne bacteria ceae,g_Cory ne bacterium£13_Unknown	98-07	7,82
		OTU_94	p_Actino bacha ria x_Actino bacha ria xy_Micrococca bas/f_Microbacha riazana g_Microbacha riums_Microbacha riumás p. A Actino bacha ria x_Actino bacha ria x_Box sion ibacha ria bacha ria Xuona nio Manana.	31-08	5,23
		оти в	p_Actino bacteria y_Actino bacteria xo_propioni bacteria est_revolutio bacteria a Actino bacteria y_Actino bacteria xo_propioni bacteria est? Progioni bacteria cene z_Cutibacteriums_Multieffiliation	78-06	1.16
			p_Bacte ro ide tes x_Bacte roid is xy_Bacte roid a les ;f_Bacte roidace ae	11-05	-4,62
			p_Bacte roldetzsy_Bacte rold a ky_Bacte rold a ky_Bacte roldaczaczg_Bacte roldes	18-05	-4,62
		0TU_52	p_Bacte roide testBacte roid is to_Bacte roid a lest_Bacte roidaceae.g_Bacte roidests_Multi-affiliation and the Multi-affiliation	31-06	6,65
			p_Bactero detest_Bactero de po_chronophege es	11-06	(,4) 5 57
			p Back to be the full and a second program and the	21-05	3.85
			p_Bacte roide tes;c_Bacte roid is;o_Cytophagales;f_Hyme notacte raceae;g_Hyme notacte r	58-04	4,86
			p_Bacteroidletesy_Bacteroidley_Flavobacteriales;#_Wealsellaceae.g_Cloacitacterium	48-05	-3,64
		OTU_26	p_Bacteroidetesy_Bacteroid is yo_Flavobacteria les,f_Wheelse Naceae,g_Close (bacterium,z_Multi-affiliation a namber instance, namber is is a faction to state in the	71-05	3,24
			p_actero detes;actero di a;o_sphingo tacteria es p_actero detes;;_actero di a;o_sphingo tacteria es d'Sphingo tacteriacene	48-02	8.75
				31-02	-4,12
			p_De pendenties x_Babelies	3E-02	-4,12
			p_De pendentinex_Babelinexo_Babelin les	3 E-02	-4,12
			p_FirmCutesy_Bacility_Bacilia esy_femily2i 	91-07	5,90
			p Firmicutes c Bacillio Iacto tacilla test I Iacto tacilla cene g Iacto tacillus	11-07	7,70
			p_Firmicutes x_C kostrid in xy_C kostrid in les; f_C kos trid in cenerg_C kostrid ium ise ns u is tricto i	31-09	7,93
			p_Firmicutes;c_C lostrid is;c_Clostrid is les;f_lachnos piaceae	46-02	-4,01
			p_Firmicutes)c_Clostrid is (o_Clostrid is les)f_Peptostre ptococcascese Significations - Clostrid is (o_Clostrid is build - De strategy strategy and a - De subscript is	56-12	9,59
		OTU 48	p_Firmicutes;c_C.kstribie;o_C.kstribie;s;f_Perfostreptococcedere;e_Komboutsie;s_metaeenome p_Firmicutes;c_C.kstribie;o_C.kstribie;s;f_Perfostreptococcedere;e_Romboutsie;s_metaeenome	78-04	3,29
			p Firmicutes;: Clostridia;: Clostridia; Es; f Rum inococasceae	21-06	-3,41
			p_Petescibecte rie	28-08	4,76
			p_Peterscibecte rie x_Seache nimoned in	SE-07	8,13
			p_retescioadae reta_bacane nimoreda eto 2.860 ne rimo rede es	21-07	8,15
			p. Proteobacte nayAlp ha proteo tacte nayAce to bacte na les f. Ace to bacte na cene	48-06	4,70
		ວານ_ສ	p_Prote obacte na y_Alp ha prote o bacte na yo_Cau iobacte na les;#_Caulobacte aceae;#_Un known y_Unknown	28-02	1,27
		OTU_14	p_Prote obsche nisy_Alphie prote obsche nisyo_Rhizobiskes,f_Rhizobiscene z_Phyllobache nium,s_Multi-stifilistion	31-02	-1,71
		010_33	p_prote consistentia):_Alpha prote consistentia):_knoa consistentia; knoa consistentia en prote consistentia; Alpha prote consistentia; Sobiaecon consistentia; Alpha prote consistentia; Sobiaecon consistentia; Sobiaecon consistentia; Alpha prote consistentia; Sobiaecon consistentia; Sobiaecon consistentia; Alpha prote consistentia; Sobiaecon consistentia; S	11-06	-101
		OTU 37	p Proteobacteria;: Gamma proteobacteria;o Beta proteobacteria les;f Burkhobleriaceae;g Cupriavidus; Multi-affiliation	41-05	5,77
		-	p_Prote obsche nisy_Gamma prote obsche nisy_Beta prote obsche nisks; f_Burkhoble niscese; g_Polynuc te obsche n	21-02	-6,00
		010_91	p_Prote obsche nie x_Gem nasprote obsche nie xo_Bete prote obsche nie kesn_Burkhobble niecese xg_Rekton is x_Multieffilistion	81-06	4,77
			p_monzona caz na "_ Learnina protecona caz na jo_ sena protecona caz na les;/_ Ne sse na case ag_ Ne sse na a Protecoha che ria y, Garmina protecoha che ria yo Epteroha che ria le na coha che ria case ag_ Ne sse na	18-02	4,43
		OTU 32	p Proteobactering: Gamma proteobactering: Legione lates; functional enderging legione las Multi-affiliation	11-04	5,74
		OTU_46	p_Prote obacte na y_Gamma prote obacte na yo_Pseudomonad a les;f_Monaxe la ceae,g_Acineto tacte rs_Multianfiliation	28-10	8,85
		OTU_31	p_Prote obacte risy:_Gamma prote obacte risy:_Pseudomonada les;f_Moraxe Naceae.g_Acineto tacte rs_Multi-affiliation 	31-03	-6,40
		010_30	µ_rious vasuus na, _ vaimma proteotaste ratio_ise valomonavia est, Morake Haceae g_Acineto taste (3_Muthamilation n. Proteobacteria v. Gaimma proteotasteria v. Vainthomonada les 1 Vainthomonada sease. Sterotio abomonas v. Muthamilation	5E-02 5E-10	5,29
IA Co	ontrols	#OTU	1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.	padj	log2 Fold Change
			p_Actino texte rie x_Actino bacte rie x_Bifd o texte rie les	18-06	9,25
			p_ActinobacteriesActinobacteriesBinfoobacteries;f_Binfoobacteriescee	18-06	9,25
		onu 🐝	румских маскатер у "Actinobacte responsory na macta releast Convertance relate a globy na bacta rium. na Actinobacte risty, Actinobacte risty: Convertance ristes of Convertance riskanes a Convertance riums. Convertance riums risk	51-03 51-04	-8,05
		OTU_42	p_actino tacta ris p_actino bacta ris p_Cory na tacta ris tast_Cory na tacta ris casaCory na tacta ris modelCory na tacta ris tast_Cory na tacta ris casaCory na tacta ris mati s	61-03	-6,42
		-	p_Actino bacteria x_Thermo leophilia	21-04	7,78
			p_Bacte roide tesx_Bacte roid is >>_Chitin op haga tes	31-09	-8,69
			p_Bactero de tes x_Bactero de xy_Christophaga les ,f_Christophaga cese	68-09	-8,83
			p Bacheroidetes: Bacheroid is:::::::::::::::::::::::::::::::::::	48-06	-4,85
			p_Bacte roide tes;c_Bacte roid ia;o_Cytophaga tes;f_Spirosom aceae	48-06	-5,22
			p_Firmicutesx_Bacilliyo_lacto tacilla les;f_lactotac illaceae	18-05	-6,8.5
			p_Firmicutesy_Bacilijo_lacto tacilla estf_lacto tacilla ceneg_lacto tacillus a Rimicutesy_Chatridia y, Chatridia la ff_Chatridia cana a ChatridiumZeneguZethisto 1	SE-06 96-12	-6,99
			p Firmicutes: Clostridia: Costridiales; T Clostridiaceae11	21-02	3,74
		OTU_62	p_Firmicutesy_Ckstridie.yo_Ckstridie.es,f_Cks tridie.ceeeitg_Ckstridium/sensu/strictofits_Multi-effilietion	61-05	-3, 55
_			p_Firmicutesy_C kstridie.yo_Ckstridie.les;f_femilydXl	3 E- 10	9,13
			p_Patescibectere n. Patescibectere v. Saccharimonal in	31-02	-3,90
			p_recessives.come rec_second minored es p. Petessibeste rieu: Second minored ieux: Second rimoneda les	21-05	-1,13
			p_ProtzobactzniaAlpha protzobactz niaRizobia izs.#_Beije nin ckaczasza Bosea	38-02	-3,05
		OTU_33	p_Prote obsche nie x_Alpha prote obsche nie y_Rhodobache naies, * Rhodobache nacese z_Pascoccus z_Multismilistion	SE-06	-3,41
		OTU_37	p_Prote obacter risy:_Gamma prote obacter risy:_Beta prote obacter risks;f_Burkhobie risceae,g_Cuprisvidus;s_Multi-affiliation Restances in the company of the test of the company of the test of the company of the test of test	3E-10	-8,31
		010_71	p_vromo o accerna,∠Gamma proteco accerna,⊘, aena proteconaccerna les,f_Burkhold e naceae,g_Massila,≦_Massila,≦sp. a. Protecohacterna y, Gamma protecohacterna yo Benta o teotavterna kast Naksaria cena	18-06	2,21
		OTU_45	p_Protzobactznia z_Gamma protzobactznia zo Diplonic lettsia les, f_Diplonic lettsia czasz, Unknown s. Unknown	31-02	4,29
		ວານ_ສະ	p_Prote obacte risy_Gamma prote obacte risyo_legione listes, flegione liscene g_legione lis z_Unknown	21-02	4,36
		OTU_67	p_Protecobacte risy:_Gamma protecobacte risyo_Reludomon ad a les;f_Morave laceae.g_Acineto bacte rs_Unknown Relation of the second secon	38-08	10,00
		010_1 0110_12	p_vromovance:ne,c_usmine protected response uncomponente es;f_Pseudomonada zeaz,g_Pseudomonas;_Multieffiliation a Protecta de cristica Gamma a rotecta de constantionomada les filanthomonada ceaz a Stenoton domonas; Multieffiliation	6E-08 3E-10	0,44
		-10 <u>-10</u>		21, 10	- 1 M - 1
-1,3	Base Mean		4		

В

Α

С

*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.

