

SUPPORTING INFORMATION

Bacterial biogeography of the human digestive tract

Jennifer C. Stearns^{a†}, Michael D.J. Lynch^{a,b}, Dilani B. Senadheera^c, Howard C. Tenenbaum^c,
Michael B. Goldberg^c, Dennis G. Cvitkovitch^c, Kenneth Croitoru^d, Gabriel Moreno-Hagelsieb^b,
and Josh D. Neufeld^{a*}.

^aDepartment of Biology, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

^bDepartment of Biology, Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada.

^cFaculty of Dentistry, University of Toronto, Toronto, Ontario, M5G 1G6, Canada.

^dDepartment of Medicine, University of Toronto, Zane Cohen Center for Digestive Diseases,
Mount Sinai Hospital, Toronto, Ontario, M5G 1X5, Canada.

[†]Present address: Farncombe Family Center for Microbiome Research, Health Sciences Center,
McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada

*Corresponding author: jneufeld@uwaterloo.ca

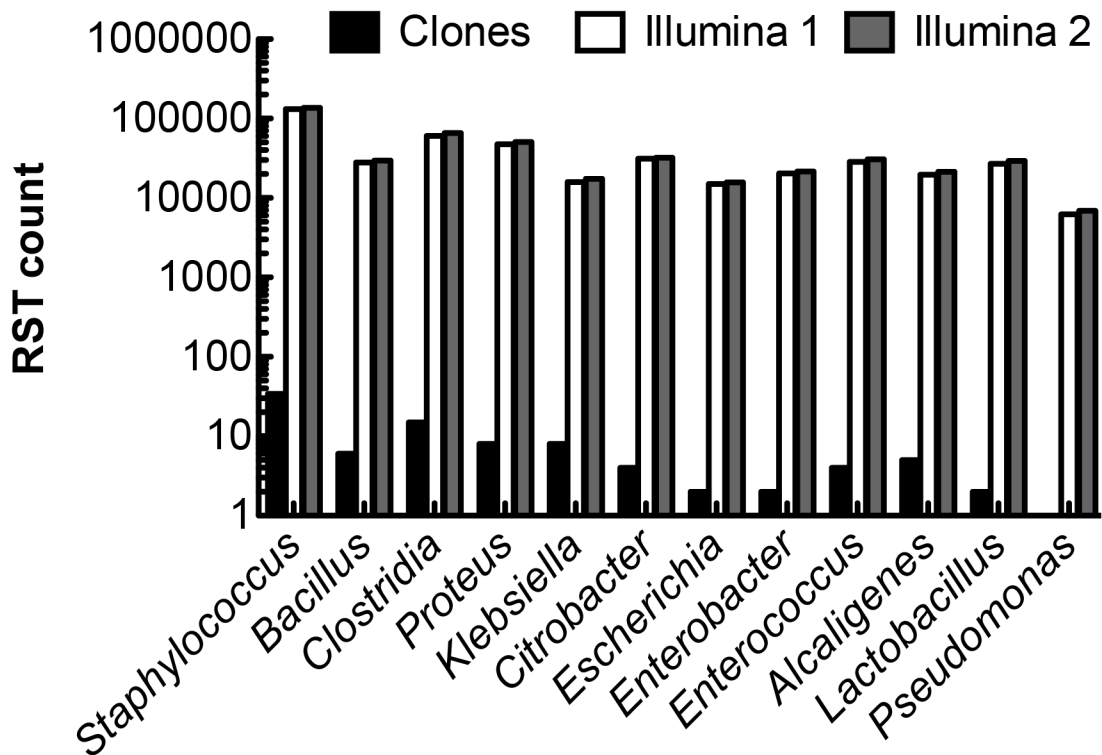


Fig. S1. Comparison of two sequencing methods for measurement of populations within a simulated bacterial community that included 12 organisms. A library of 96 clones, sequenced with the traditional Sanger method was compared to a library of approximately 500,000 sequences returned with the Illumina method (shown in duplicate). No results were obtained for *Pseudomonas* within the clone library, however many sequences for this organism were returned within the Illumina libraries. DNA for all organisms was added in equal amounts, except for *Staphylococcus* for which more was added.

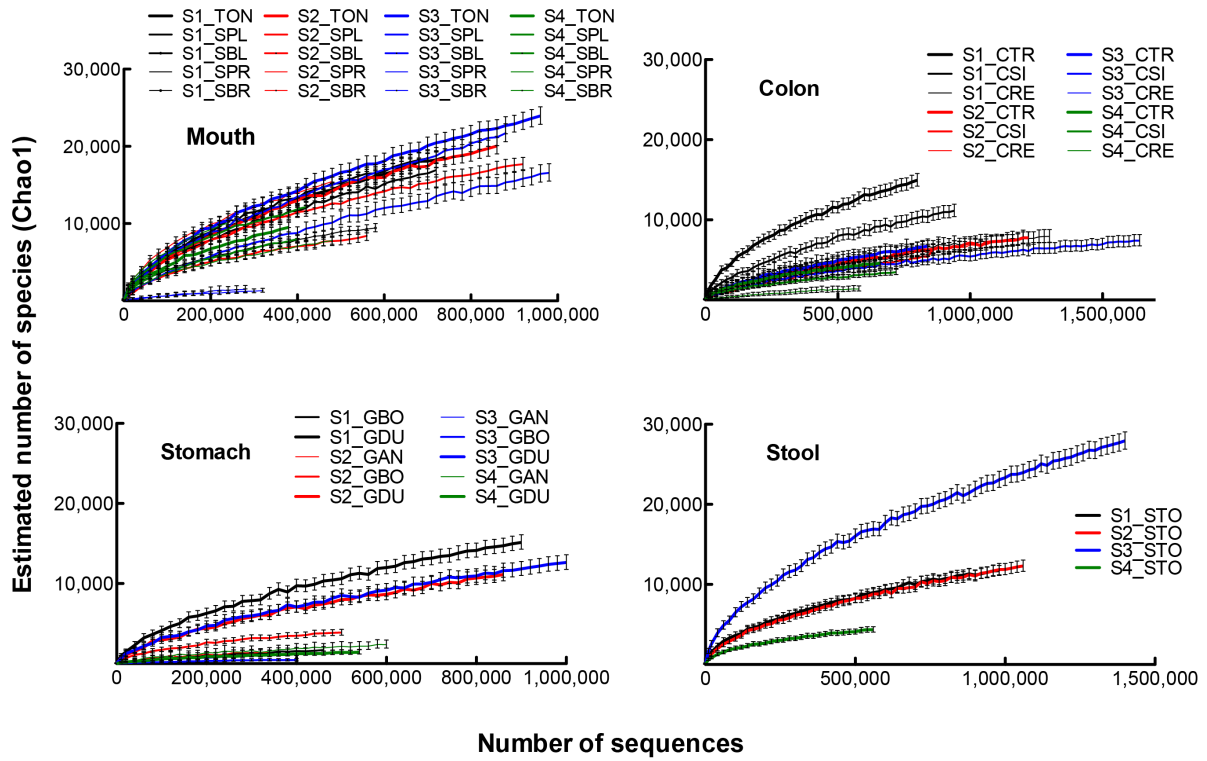


Fig. S2. Species estimates for all samples grouped by site. Average Chao1 is presented from five iterations of rarified subsets of all sequences (on the x-axis) and error bars represent the upper and lower limits of the 95% confidence interval.

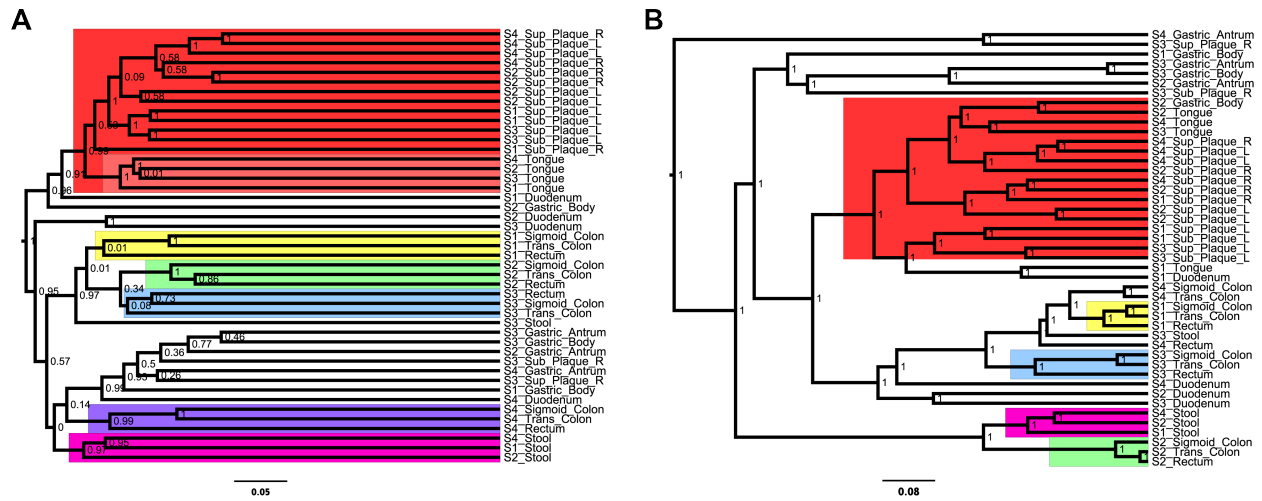


Fig. S3. UPGMA trees of A) unweighted UniFrac and B) weighted UniFrac distances. Nodes are labeled with Jackknife values calculated with 100 iterations of 200,000 sequences. Branch colours: mouth (red), tongue plaque (light red), stool (magenta), subject 1 colon (yellow), subject 2 colon (green), subject 3 colon (blue), subject 4 colon (purple).

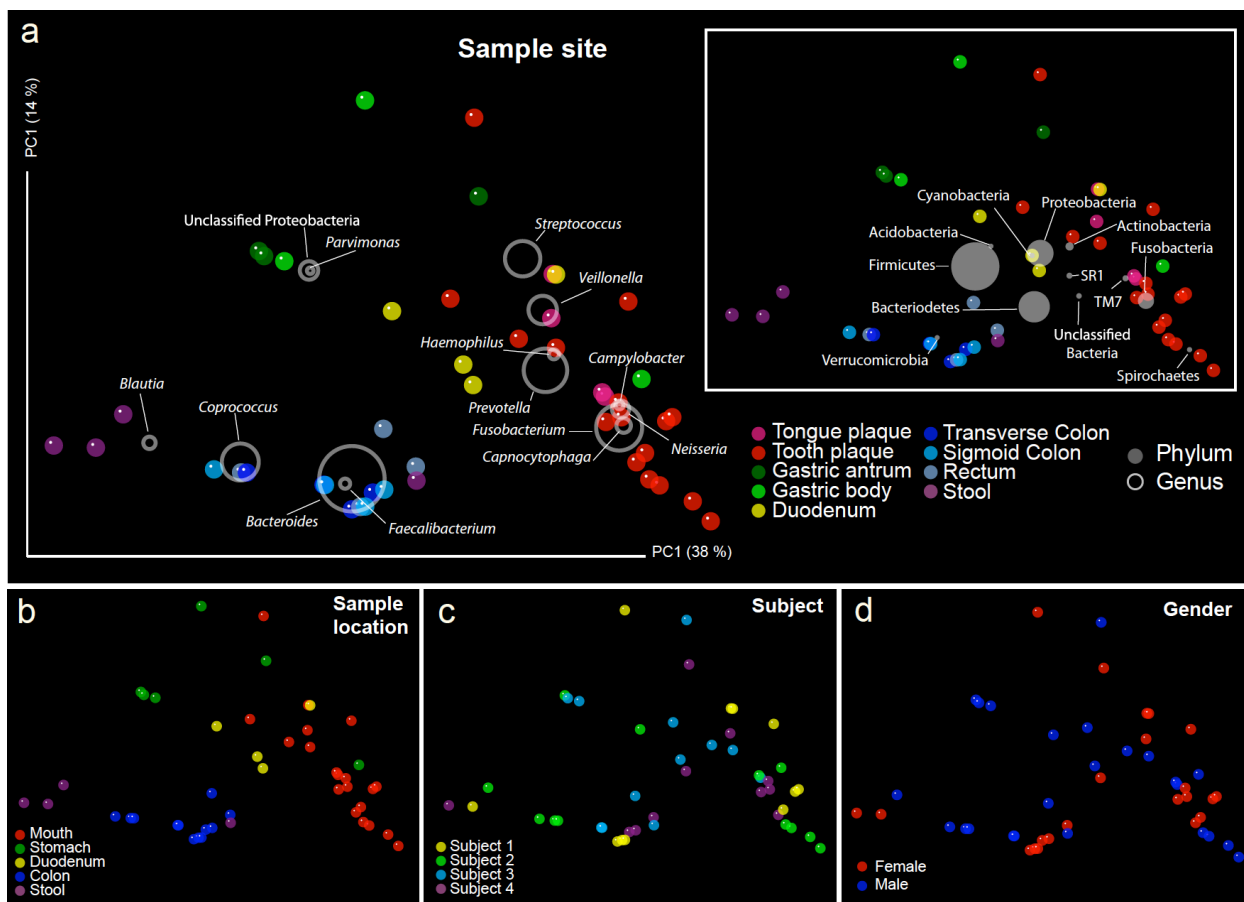


Fig. S4. Contribution of different taxonomic groups to separation of samples based on bacterial abundance and phylogenetic information. The contribution of each group is represented by the size of the circles (grey) overlaid onto a PCoA of weighted UniFrac distances for all samples within the oral and digestive tract. Panels (a-d) represent variations in sample colouration to highlight potential relationships between sample clustering and metadata.

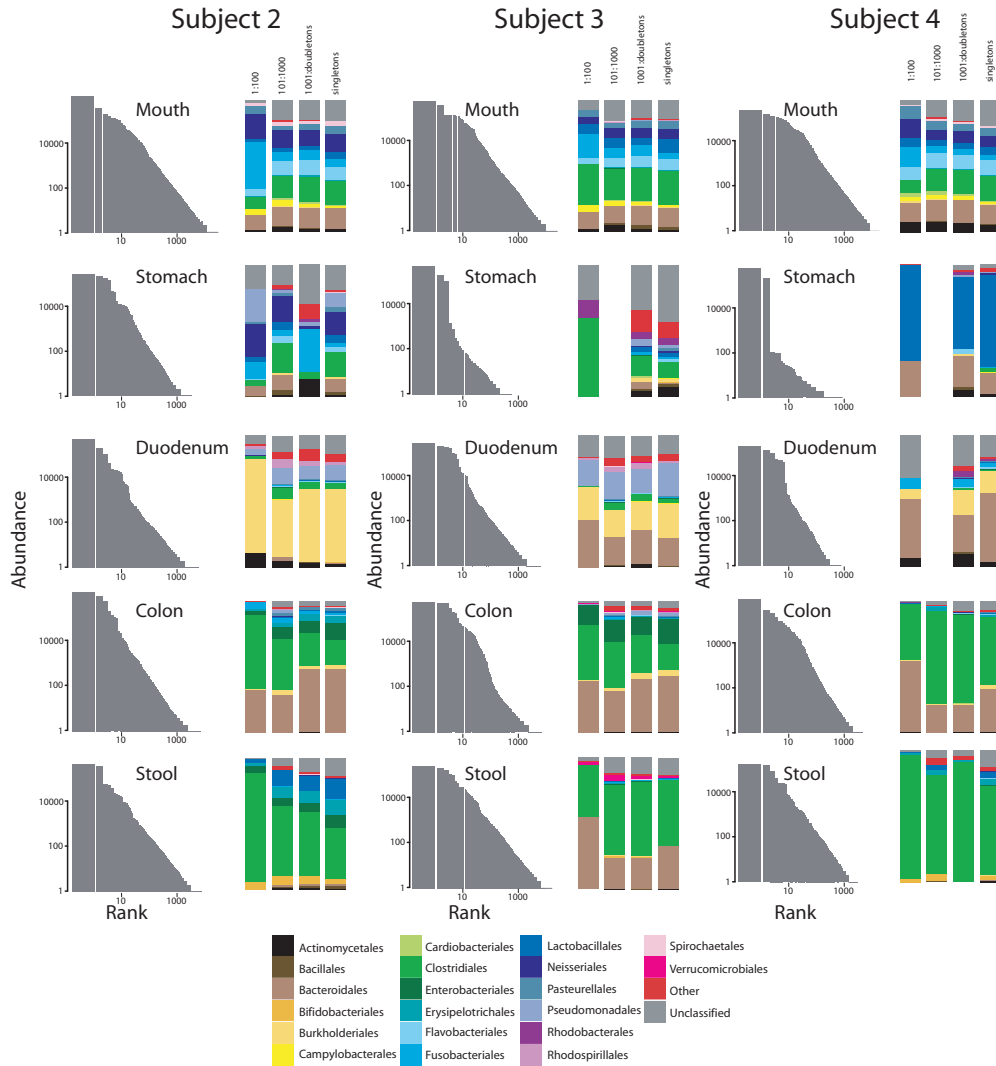


Fig. S5. Rank abundance plots and proportional taxonomic bar plots for sequences from Subjects 2-4 pooled within various body sites and assigned to order. To demonstrate taxonomic distribution at decreasing ranks phylotypes were split according to different logarithmic abundance ranges as appropriate. The low abundance orders comprising the ‘Other’ category included Synergistales, Halanaerobiales, Mycoplasmatales, Xanthomonadales, Sphingobacteriales, Caulobacterales, Desulfobacterales, Legionellales, Oceanospirillales, Deinococcales, Methanobacteriales, Myxococcales, Anaerolineales, Methylophilales, Chromatiales, Thermales, Bdellovibrionales, Desulfuromonadales, Solirubrobacteriales, Methanomicrobiales, Planctomycetales, Methylococcales, Anaeroplasmatales, Coriobacteriales, Desulfovibrionales, Rhizobiales, Rhodocyclales, Sphingomonadales, Victivallales.

Table S1. Number of clusters (phylotypes) over various sequence identity thresholds using the CD-HIT analysis package for all sequences over all treatments.

Cluster (%)	CD-HIT
99	665,076
98	203,148
97	97,251
96	49,355
95	32,017

Table S2. Summary table of samples and sequencing results. The number of sequences refers to the number of assembled sequences after error filtering. The number of phylotypes is the number of sequence clusters with 97% identity.

Subject	Gender	Site	Sample	Sequence counts	Phylotype counts	Good's coverage			
1	Female	Mouth	Tongue	749,110	7,249	0.994			
			Subgingival - left side	712,873	7,749	0.994			
			Supragingival - left side	580,927	4,794	0.996			
			Supragingival - right side	725,916	6,971	0.994			
			Supragingival - right side	23,555	427	0.990			
		Stomach	Gastric body	466,105	543	0.999			
			Duodenum	Duodenum	905,619	6,039	0.996		
				Colon	Transverse	806,472	6,634	0.995	
					Sigmoid	947,940	4,467	0.997	
					Rectum	1,305,935	2,550	0.999	
					Stool	Stool	985,754	5,681	0.997
						Mouth	Tongue	865,725	7,892
					2		Male	Mouth	Subgingival - left side
Supragingival - left side	417,865	5,572	0.993						
Supragingival - right side	928,856	7,921	0.995						
Supragingival - right side	494,229	7,428	0.992						
Stomach	Gastric antrum	358,061	588	0.999					
	Duodenum	Gastric body	505,693	1,883		0.998			
		Duodenum	879,898	4,155		0.997			
Colon			Transverse	1,239,858		2,916		0.999	
			Sigmoid	852,442		1,669		0.999	
			Rectum	1,195,591		2,666		0.999	
			Stool	Stool		1,063,904		5,541	0.997
				Mouth		Tongue		976,495	9,648
3	Male	Mouth	Subgingival - left side			892,482		8,268	0.994
			Supragingival - left side	338,944	565	0.999			
			Supragingival - right side	995,946	5,853	0.996			
			Supragingival - right side	281,634	411	0.999			
			Stomach	Gastric antrum	269,779	348	0.999		
		Duodenum		Gastric body	405,549	202	1.000		
				Duodenum	1,006,884	4,635	0.997		
		Colon			Transverse	848,073	2,547	0.998	
					Sigmoid	1,652,488	2,663	0.999	
					Rectum	424,944	1,408	0.998	
					Stool	Stool	1,408,777	12,426	0.995
						Mouth	Tongue	390,262	4,357
		4	Female	Mouth	Subgingival - left side		406,422	3,858	0.995
Supragingival - left side	243,717				4,858	0.989			
Supragingival - right side	421,279				5,861	0.993			
Supragingival - right side	489,144				3,829	0.996			

Stomach	Gastric antrum	607,152	657	0.999
Duodenum	Duodenum	551,726	602	0.999
Colon	Transverse	678,894	2,248	0.998
	Sigmoid	738,341	1,776	0.999
	Rectum	589,654	494	0.999
Stool	Stool	564,162	2,359	0.998
		Total	Total	Average
		32,770,833	184,752	0.996

Table S3. Sequence counts for phyla present in each sample. See Excel spreadsheet.

Table S4. Sequence counts for genera present in each sample. See Excel spreadsheet.

Table S5. List of sequence counts for 97% OTU clusters shared across all subjects and all sampling regions. Taxonomy is based on the RDP classifier. See Excel spreadsheet.

Table S6. Prevalence in each sample of medically important species. Counts derived from blastn results with 99% sequence identity

		<i>Faecalibacterium prausnitzii</i>	<i>Clostridium difficile</i>	<i>Streptococcus mutans</i>	<i>Treponema denticola</i>	<i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>	Segmented filamentous bacterium	<i>Helicobacter hepaticus</i>
Species:									
Strain:		ATCC 27768	630	NCTC 10449	ATCC 35405	LMG 1242T	ATCC 10145		ATCC 51449
Accession:		AJ413954	AM180355	AJ243965	AF139203	Z76651	AF094713	X77814	AE017125
Subject 1	Mouth Tongue	21	0	0	0	0	0	0	0
	Mouth Supra_R	0	0	3	0	0	0	0	0
	Mouth Supra_L	7	0	309	0	0	0	0	0
	Mouth Sub_R	1	0	902	0	0	0	0	0
	Mouth Sub_L	13	0	0	0	0	0	0	0
	Stomach Gastric_Body	0	0	0	0	1	1	0	0
	Duodenum Duodenum	20	0	0	0	0	0	0	0
	Colon Transverse	55580	0	0	0	0	0	0	0
	Colon Sigmoid	66769	0	0	0	0	0	0	0
	Colon Rectum	59736	0	0	0	0	0	0	0
	Stool Stool	16497	0	0	0	0	0	0	0
Subject 2	Mouth Tongue	16	0	0	36	0	0	0	0
	Mouth Supra_R	0	0	0	625	0	0	0	0
	Mouth Supra_L	14	0	6	625	0	0	0	0
	Mouth Sub_R	0	0	0	83	0	0	0	0
	Mouth Sub_L	8	0	0	1949	0	0	0	0
	Stomach Gastric_Body	0	0	0	0	0	0	0	0
	Stomach Gastric_Antrum	3	0	0	0	0	2	0	0
	Duodenum Duodenum	18	0	0	0	1	3	0	0
	Colon Transverse	37	105	0	0	52	136	0	0
	Colon Sigmoid	12	0	0	0	0	0	0	0
	Colon Rectum	18	162	0	0	0	0	0	0
Stool Stool	28	155	206	0	0	0	0	0	
Subject 3	Mouth Tongue	20	0	0	1	0	0	0	0
	Mouth Supra_R	0	0	1	0	0	0	0	0
	Mouth Supra_L	17	0	0	0	0	0	0	0
	Mouth Sub_R	28	0	0	0	0	1	0	0
	Mouth Sub_L	67	0	0	1	0	0	0	0
	Stomach Gastric_Body	0	0	0	0	0	0	0	0
	Stomach Gastric_Antrum	0	0	0	0	1	1	0	0

	Duodenum	Duodenum	24	0	0	0	0	1	0	0
	Colon	Transverse	125994	0	0	0	0	0	0	0
	Colon	Sigmoid	220481	0	1	0	0	0	0	0
	Colon	Rectum	22972	0	0	0	1	1	0	0
	Stool	Stool	60563	1	0	0	1	2	0	0
	Mouth	Tongue	0	0	16	10	0	0	0	0
	Mouth	Supra_R	0	0	0	0	0	0	0	0
	Mouth	Supra_L	1	0	83	0	0	0	0	0
	Mouth	Sub_R	9	0	0	0	0	0	0	0
	Mouth	Sub_L	0	0	0	0	0	0	0	0
	Stomach	Gastric_Antrum	1	0	0	0	0	0	0	0
	Duodenum	Duodenum	0	0	0	0	0	0	0	0
Subject 4	Colon	Transverse	17086	0	0	1	0	0	0	0
	Colon	Sigmoid	10108	0	0	0	0	0	0	0
	Colon	Rectum	0	0	0	0	0	0	0	0
	Stool	Stool	24	2	1	0	0	0	0	0