

Supplementary Table 2) Classification of reads from whole genome shotgun sequencing.

Sample location	Involved	Uninvolved	Involved	Uninvolved	Involved	Uninvolved	Involved	Uninvolved	Involved	Uninvolved	Involved	Uninvolved
Sequencing sample ID	CC-M-067	CC-M-066	CC-M-071	CC-M-070	CC-M-073	CC-M-072	CC-M-075	CC-M-074	CC-M-077	CC-M-076	CC-M-088	CC-M-094
Sample ID	CD-I-5	CD-U-2	CD-I-25	CD-U-22	CD-I-65	CD-U-63	CD-I-56	CD-U-53	CD-I-59	CD-U-58	CD-I-11	CD-U-9
Read Length	100	100	100	100	100	100	100	100	100	100	100	100
Total number of reads	65196446	58497580	89099732	81481942	49797264	57998298	66843642	75664790	50405638	107033092	107071062	91506374
Low quality reads (removed)	17163097	15283097	25905262	21456670	11698725	12526822	13152964	15088967	10521477	21884440	30172510	26369234
Duplicate / repeat reads	1703651	1610262	3255461	2786171	957897	898760	754027	1254386	580562	1344731	7864823	5989392
Human reads	46322100	41597052	59924269	57230333	37137198	44568226	52931642	59309764	39300509	83775123	68998495	59113085
Known bacterial reads	671	742	2206	795	260	509	473	2629	277	19445	7871	7586
Known other reads (Archaea, Fungi, etc)	11	13	17	15	6	13	5	15	3	18	32	16
Known viral / phage reads	1	7	10	11	72	3	3	7	2	89	8	9
Unmapped reads	6915	6407	12507	7948	3106	3965	4528	9022	2808	9246	27323	27052

Sample location	Involved	Uninvolved	Involved	Uninvolved	Involved	Uninvolved*	Involved	Uninvolved	Involved	Uninvolved	**Proximal margin
Sequencing sample ID	CC-M-090	CC-M-089	CC-M-092	CC-M-091	CC-M-096	CC-M-095	CC-M-097	CC-M-093	CC-M-099	CC-M-098	CC-M-069
Sample ID	CD-I-16	CD-U-13	CD-I-51	CD-U-47	CD-I-36	CD-U-32	CD-I-78	CD-U-76	CD-I-82	CD-U-81	CD-I-20
Read Length	100	100	100	100	100	100	100	100	100	100	100
Total number of reads	46063398	57306784	57505252	83525914	45601516	35595524	160018990	111195604	67282074	67506916	75904148
Low quality reads (removed)	13286967	16512255	15573178	22630712	10352888	8739436	39002030	29469424	19049423	19718651	19745534
Duplicate / repeat reads	3937351	6242635	5397054	6876296	3167204	3255911	8731082	4572432	4772306	3644714	2150294
Human reads	28819807	34492091	36489494	53968002	32067900	23575877	112258683	77127127	43414446	44100996	53998459
Known bacterial reads	7841	14248	10024	10757	3168	9704	5858	5837	12381	10876	1323
Known other reads (Archaea, Fungi, etc)	30	48	28	106	14	60	17	26	21	30	11
Known viral / phage reads	4	28	18	15	857	639	7	6	20	23	2
Unmapped reads	11398	45479	35456	40026	9485	13897	21313	20752	33477	31626	8525

Sample location	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free	IBD-free
Sequencing sample ID	CCM146	CCM147	CCM148	CCM149	CCM150	CCM151	CCM152	CCM153	CCM154	CCM155	CCM156	CCM157
Sample ID	TI-N02	TI-N04	TI-N06	TI-N08	TI-N10	TI-N12	TI-N14	TI-N16	TI-N18	TI-N20	TI-N22	TI-N24
Read Length	101	101	101	101	101	101	101	101	101	101	101	101
Total number of reads	298152056	28561816	26112868	30443438	32621480	20919238	21599392	24398862	37615244	20988402	33136006	27305124
Low quality reads (removed)	78768968	8219766	7127159	7863157	9633610	5473606	5960687	5645193	11021830	5700834	9358261	6787264
Duplicate / repeat reads	30468221	894398	719835	1420165	2222398	890885	758666	758418	2954760	770623	1846587	814272
Human reads	188871539	19444184	18263349	21150617	20754851	14550225	14875327	17992893	23554203	14507905	21920313	19698369
Known bacterial reads	4256	250	301	2330	1943	475	951	244	5925	2645	2208	805
Known other reads (Archaea, Fungi, etc)	78	9	9	5	13	6	12	2	50	23	13	6
Known viral / phage reads	8	1	0	12	5	4	5	2	10	5	9	0
Unmapped reads	38986	3208	2215	7152	8660	4037	3744	2110	78466	6367	8615	4408

* histopath uninv. Included

** don't have uninvolved sample

Known other reads - Archaea, Fungi, Cloning vector, etc