Sample	Group	Raw Reads[a]	Cleaned Reads[b]	Decontam Reads[c]	Classified Reads[d]	Avg Read Len[e]	Uniquely Mapped Reads[f]	Uniquely Mapped Reads (%)[g]	Avg Mapped Len[h]	Multi Mapped Reads[i]	Too Many Mapped Reads[j]
SUBJECT12A	Breast-fed	12657133	12351034	12339489	7270725	203	9219198	74.71%	201.85	118862	120
SUBJECT12B	Breast-fed	12936604	12630335	12622859	7777684	208	9408386	74.53%	206.72	109466	93
SUBJECT13A	Breast-fed	11975209	11679060	11664826	7382648	204	9034835	77.45%	202.97	101319	64
SUBJECT13B	Breast-fed	13066054	12768140	12758643	7692273	208	10075514	78.97%	207.03	102455	60
SUBJECT14A	Breast-fed	15053426	14745011	14735687	11918114	211	11174734	75.83%	209.87	94207	190
SUBJECT14B	Breast-fed	13822355	13532828	13522147	10763117	211	10407265	76.96%	209.97	88903	173
SUBJECT15A	Breast-fed	11789722	11501043	11437870	8234878	199	9463599	82.74%	198.28	92730	160
SUBJECT15B	Breast-fed	12979568	12665711	12625070	9225786	203	10513543	83.28%	202.17	95107	164
SUBJECT17A	Breast-fed	11380343	11114475	11025775	9547213	208	6820861	61.86%	205.03	141756	242
SUBJECT17B	Breast-fed	14655796	14279589	13098785	10927614	210	7963423	60.80%	207.25	149252	267
SUBJECT18A	Breast-fed	13705186	13331276	12821732	9801896	207	10162009	79.26%	207	65543	156
SUBJECT18B	Breast-fed	11990896	11666233	11341767	8698848	210	8992209	79.28%	209.74	55137	121
SUBJECT1A	Breast-fed	13365572	13018552	12998305	10551311	196	9755913	75.06%	194.54	91337	300
SUBJECT1B	Breast-fed	12781762	12467526	12456890	10225857	203	8820735	70.81%	200.81	81438	287
SUBJECT5A	Breast-fed	13281471	12983934	12963442	9141529	203	9673165	74.62%	202.05	146408	237
SUBJECT5B	Breast-fed	12645142	12369779	12360002	8394481	210	9181845	74.29%	208.53	117330	145
SUBJECT7A	Breast-fed	11398789	10973328	8896563	8339483	206	5549725	62.38%	203.21	79032	8
SUBJECT7B	Breast-fed	11720701	11366003	10614338	10016505	213	6582560	62.02%	210.38	83728	14
SUBJECT9A	Breast-fed	11204847	10914410	10628553	9301773	213	6455188	60.73%	209.74	114943	127
SUBJECT9B	Breast-fed	11478766	11148240	10496969	9075236	213	6371452	60.70%	209.31	108837	113
SUBJECT10A	Formula-fed	11788309	11538713	11530193	9661213	207	8606080	74.64%	206.21	50419	298
SUBJECT10B	Formula-fed	11473609	11230372	11218399	9817550	212	8330119	74.25%	210.13	44372	207
SUBJECT16A	Formula-fed	12824459	12502271	12492961	11290652	205	7544713	60.39%	201.46	165614	230
SUBJECT16B	Formula-fed	12829805	12523603	12518269	11563018	210	7720225	61.67%	206.6	157036	221
SUBJECT2A	Formula-fed	12127916	11740300	11723314	8532270	201	7540092	64.32%	198.45	90829	380
SUBJECT2B	Formula-fed	11651159	11300971	11289641	8286360	206	7449144	65.98%	203.89	81076	166
SUBJECT3A	Formula-fed	11060062	10807117	10797847	2835985	204	8895746	82.38%	202.46	51802	152
SUBJECT3B	Formula-fed	16470871	16051157	16026213	4966310	199	12754901	79.59%	197.94	83927	198
SUBJECT4A	Formula-fed	10931262	10650957	10639707	9162589	203	7050997	66.27%	201.08	89438	223
SUBJECT4B	Formula-fed	11852743	11541081	11520917	9021218	204	7895885	68.54%	201.95	102816	195

Supplementary Table 1. (a) Number of raw reads after sequencing. (b) Number of reads after adapter trimming. (c) Number of reads after human and PhiX decontamination. (d) Number of reads classified by Kraken1. (e) Average read length. (f) Number of reads which mapped to one unique location in the metagenome. (g) Percentage of reads which mapped to one unique location in the metagenome. (h) Average mapped read length. (i) Number of reads that mapped to multiple locations in the metagenome. (j) Number of reads that were discarded due to being mapped to more than 20 location in the metagenome. The means of all columns shown in this table were compared between Breast-fed and Formula-fed groups using a Welch's t-test with a significance level of 0.01; no statistically significant

differences were found.