Cell Systems, Volume *3*

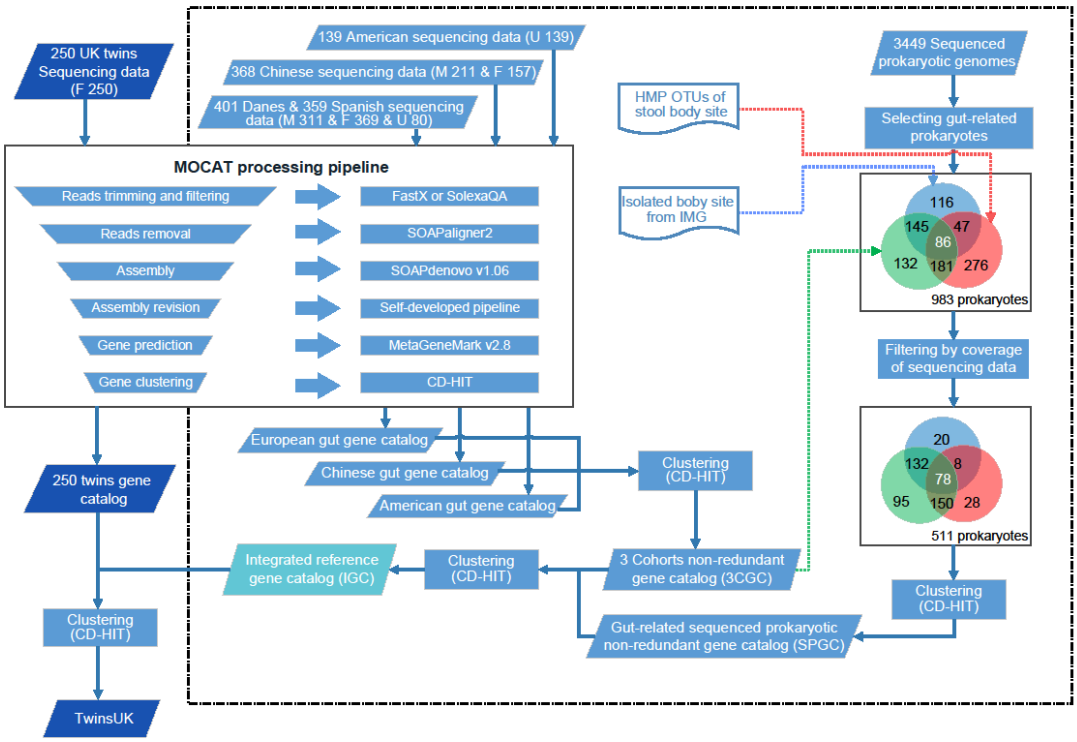
Supplemental Information

Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome

Hailiang Xie, Ruijin Guo, Huanzi Zhong, Qiang Feng, Zhou Lan, Bingcai Qin, Kirsten J. Ward, Matthew A. Jackson, Yan Xia, Xu Chen, Bing Chen, Huihua Xia, Changlu Xu, Fei Li, Xun Xu, Jumana Yousuf Al-Aama, Huanming Yang, Jian Wang, Karsten Kristiansen, Jun Wang, Claire J. Steves, Jordana T. Bell, Junhua Li, Timothy D. Spector, and Huijue Jia

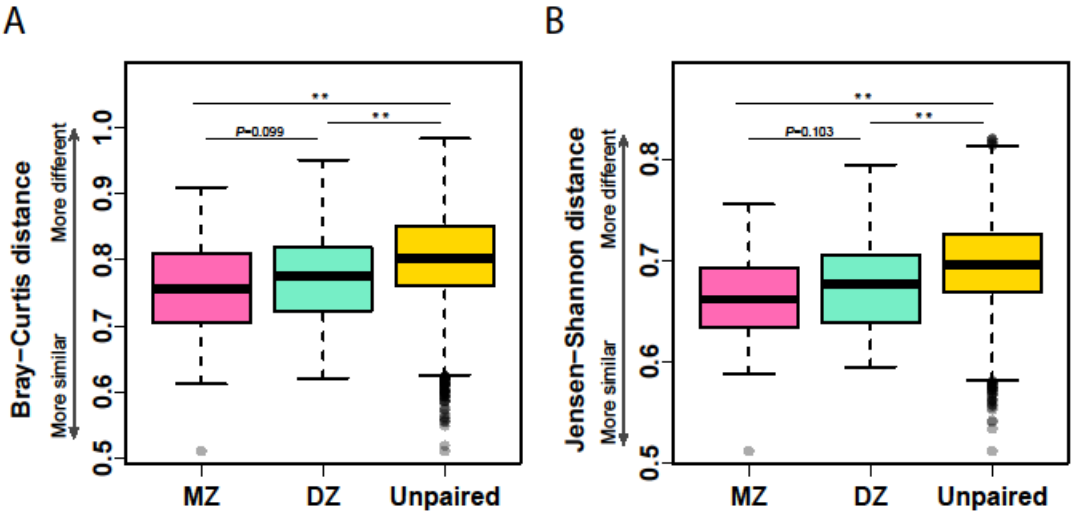
SUPPLEMENTAL FIGURES

Figure S1 related to Figure 1. Construction of the TwinsUK gut microbial reference gene catalog.



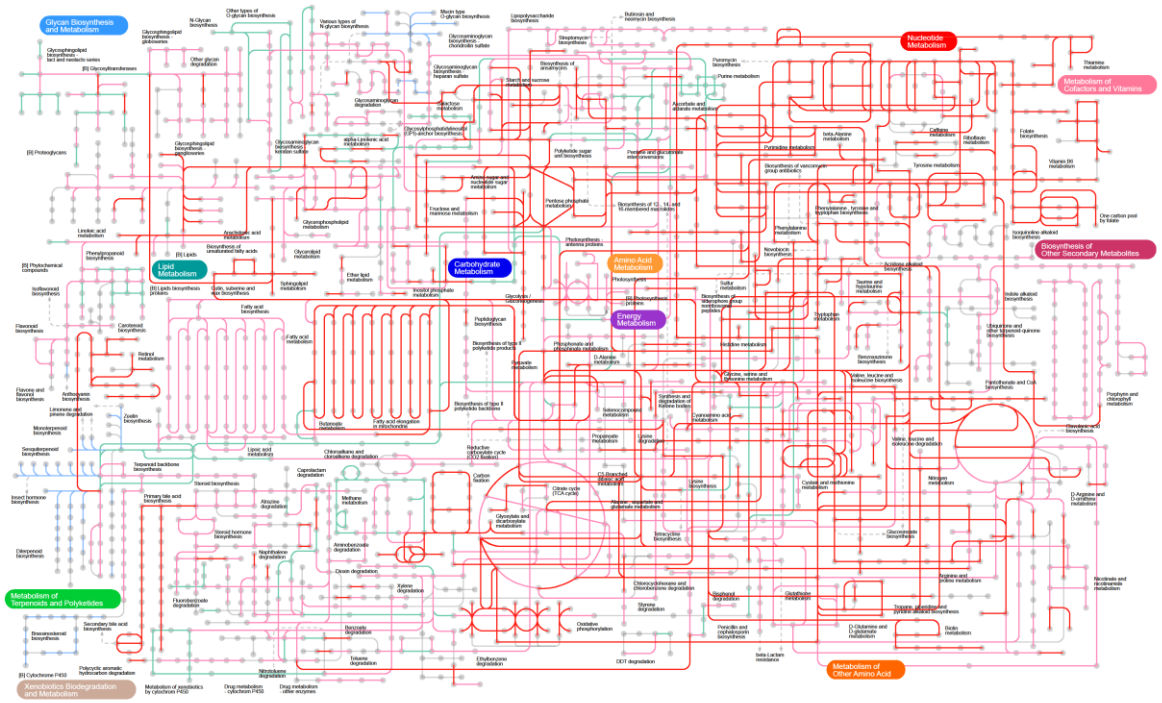
Metagenomic sequencing data from the 250 TwinsUK samples were processed with the MOCAT pipeline and the resulting gene catalog was merged with IGC using CD-HIT as previously described (Li et al., 2014), leading to a non-redundant catalog of 11.4 million genes. The pipeline in the dashed box was adapted from Figure 1 of the IGC publication (Li et al., 2014). M, number of male samples; F, number of female samples; U, number of gender unkonwn samples.

Figure S2 related to Figure 2. Greater resemblance of the gut microbiome between twins.



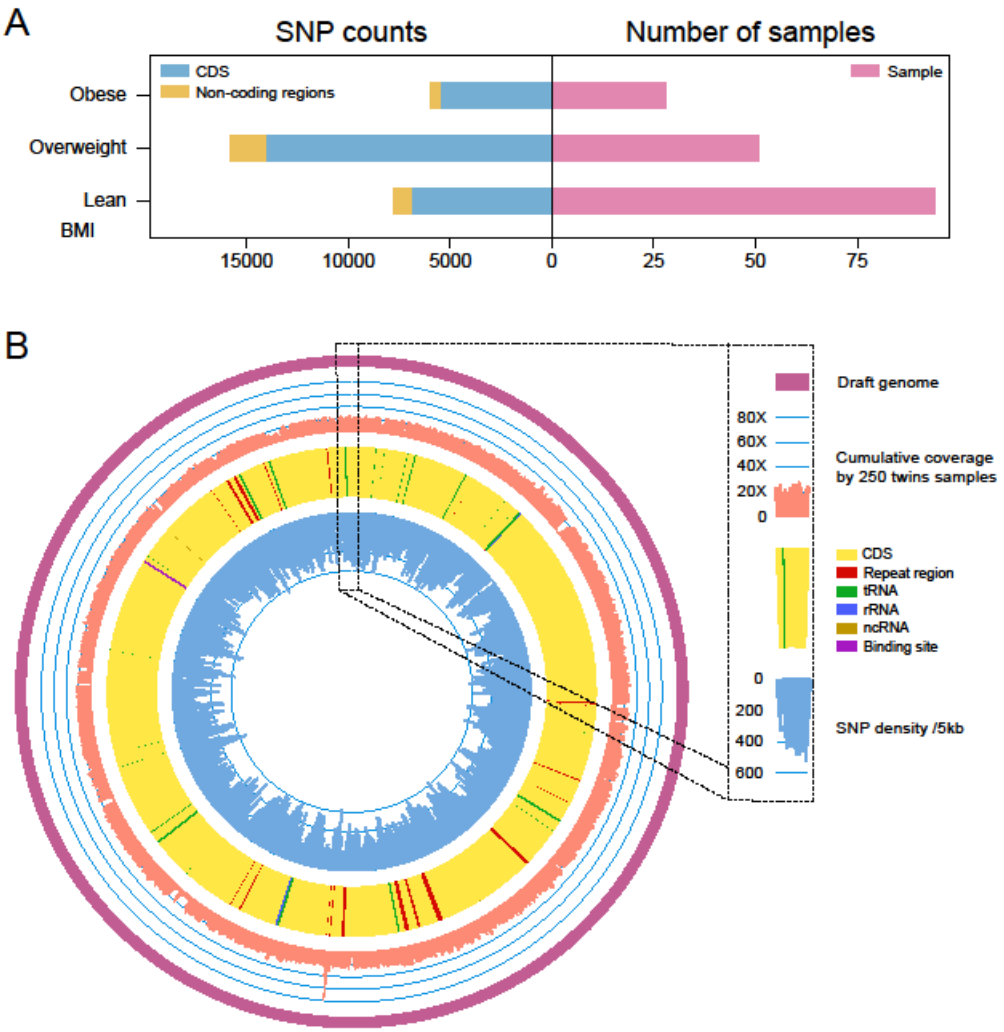
(A) Bray-Curtis and **(B)** Jensen-Shannon distances of the gut microbial gene abundance profile between samples. MZ, between paired MZ twins; DZ, between paired DZ twins; unpaired, between any two unrelated samples. **, $p < 0.01$; *, $p < 0.05$, Wilcoxon rank sum test.

Figure S3 related to Figure 4. Heritability of KO pathways.



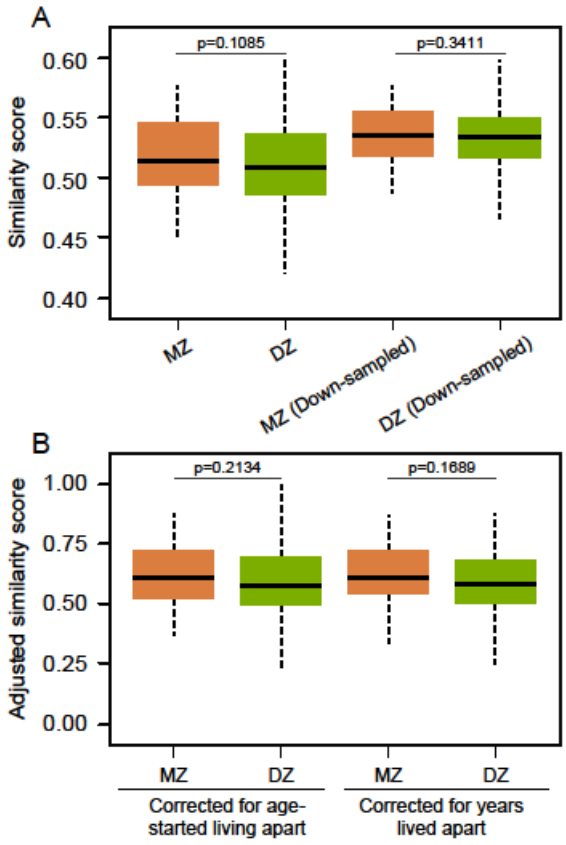
KEGG metabolic pathways identified in the TwinsUK cohort are colored to their ACE calculations. Blue, present in less than 50% of the samples; green, ICC MZ < DZ; pink, ICC MZ > DZ; red, ICC MZ > DZ and $p < 0.1$ between ACE and CE models (**Table S3B**). Pathways were drawn in the order: blue, green, pink and red**.**

Figure S4 related to Figure 6. Distribution of SNPs in *A. muciniphila*.



(A) Total number of SNPs in all samples in each BMI group, and their distribution in protein coding and non-coding regions. The number of samples in each BMI group (obese, overweight, lean) is shown on the right. All subjects in the lean group have a normal BMI (18.50 ~ 24.99) except for 4 underweight subjects (< 18.5). **(B)** Cumulative coverage of the *A. muciniphila* draft genome by sequencing data from all 250 samples, and distribution of SNPs in different genomic regions.

Figure S5 related to Figure 7. **slightly higher SNP similarity score between MZ than DZ twins.**



(A) To compare all samples at the same sequencing coverage, data in each sample was down-sampled to 10x coverage (Schloissnig et al., 2012). p = 0.1085 before and p = 0.3411 after down-sampling, one-tailed Wilcoxon rank sum test. **(B)** After correcting influences related to live apart,p = 0.2134 for age-started living apart and p = 0.1689 for years lived apart.

SUPPLEMENTAL TABLES

Table S2 related to Figure 2. Potential influences of phenotypes on the gut microbiome.

(A) PERMANOVA for the influence of phenotypes on the gut microbial gene profile. 9999 permutations, Bray-Curtis distance. As one PERMANOVA test was performed for each phenotype, multiple testing was controlled using the Benjamini-Hochberg procedure. Twin pair information used for permutation of the 'order of birth' only.

Phenotypes	Number of twins	Degree of freedom	Sums of squares	Mean square	F model	Pseudo-R2	P (>F)	p.adjust (BH)
Twin pair number	246	122	43.280	0.355	1.196	0.543	0.000	0.002
BMI	249	1	0.511	0.511	1.571	0.006	0.002	0.017
Drugs (Diabetic Tablets)	230	1	0.463	0.463	1.428	0.006	0.015	0.083
Has a doctor ever diagnosed or treated you for any of the following conditions? \	222	1	0.429	0.429	1.319	0.006	0.028	0.091
Year of birth	250	1	0.424	0.424	1.303	0.005	0.035	0.091
Current location	247	3	1.133	0.378	1.161	0.014	0.037	0.091
Vegan or vegetarian	198	1	0.420	0.420	1.291	0.007	0.041	0.091
Age at metagenomics sample	250	1	0.417	0.417	1.280	0.005	0.043	0.091
Number of units of alcohol drunk per	241	1	0.389	0.389	1.193	0.005	0.100	0.173
Menopausal Status	240	2	0.734	0.367	1.129	0.009	0.102	0.173
Smoking status	249	1	0.367	0.367	1.126	0.005	0.160	0.247
Currently, how many minutes per week do you spend walking briskly /	196	1	0.354	0.354	1.096	0.006	0.214	0.303
Currently, how many minutes per week do you spend in non-weight bearing activity? e g swimming, cycling, yoga,	188	1	0.323	0.323	0.998	0.005	0.447	0.559
Insulin drugs	230	1	0.321	0.321	0.988	0.004	0.460	0.559
Exercise \ Currently, how many minutes per week do you spend in weight bearing activity? e g aerobics, running, dance, football, basketball, racquet sports etc (do not include walking or	191	1	0.309	0.309	0.953	0.005	0.589	0.667
Outdoor sports	108	1	0.290	0.290	0.885	0.008	0.798	0.848
Birth order	220	1	0.261	0.261	0.800	0.004	0.856	0.856

(B) Correlation between continuous phenotypes and Bray-Curtis distance between the gut microbial gene profiles of paired twins. These phenotypes are available for most of the 34 MZ and 89 DZ pairs. P-values from Pearson, Spearman, or Kendall's correlation coefficients.

Phenotype	Number of twin pairs	cc (Pearson)	cc (Spearman)	cc (Kendall)	P-value (Pearson)	P-value (Spearman)	P-value (Kendall)
Age started living apart	123	-0.2662	-0.2082	-0.1449	0.0029	0.0208	0.0212
Years lived apart	123	0.1682	0.0509	0.0372	0.0629	0.5759	0.5419
Age at metagenomics sample	123	-0.0461	-0.0564	-0.0454	0.6126	0.5347	0.4560
BMI difference	123	0.1193	0.0676	0.0476	0.1887	0.4570	0.4351
Weight difference	123	-0.0401	0.0162	0.0093	0.6599	0.8593	0.8800
Height difference	123	0.0372	-0.0316	-0.0172	0.6826	0.7288	0.7779

Table S7 related to Figure 7. Correlation between age-related phenotypes and SNP similarity score of paired twins. These phenotypes are available for most of the 34 MZ and 89 DZ pairs. P-values from Pearson, Spearman, or Kendall's correlation coefficients. * for twin pairs that separated between 16-24 years old only.

Phenotype	Number of twin pairs	cc (Pearson)	cc (Spearman)	cc (Kendall)	P-value (Pearson)	P-value (Spearman)	P-value (Kendall)
Age started living apart	123	0.1293	0.1938	0.1352	0.1541	0.0317	0.0315
Years lived apart	123	-0.1160	-0.2100	-0.1374	0.2015	0.0199	0.0242
Age at metagenomics sample	123	-0.0211	-0.0755	-0.0505	0.8172	0.4062	0.4074
Age started living apart*	109	0.2240	0.2271	0.1575	0.0192	0.0176	0.0195
Years lived apart*	109	-0.1182	-0.2115	-0.1325	0.2208	0.0274	0.0411
Age at metagenomics sample*	109	-0.0662	-0.1457	-0.0962	0.4938	0.1306	0.1383

Please see the following Supplemental Tables in the Excel files provided.

Table S1 related to Figure 1. The TwinsUK cohort for metagenomic sequencing.
(A) Phenotypic information for the twins.
(B) Statistics for the sequencing data. 1.9 TB of raw Illumina sequencing reads led to 1.8 TB of high-quality non-human reads. Percentage of reads commonly or uniquely mapped to IGC (Li et al., 2014) or the 250 twins gene catalog is also shown (**Fig. 1**).
(C) The updated reference gene catalog compared to IGC.

Table S3 related to Figure 3. Heritability of gut microbial taxa. The p-values of the ICC-filtered (r(MZ) > r(DZ); r(MZ) > 0 and p < 0.01) taxa were controlled for multiple testing according to Storey's FDR method.
(A) Phyla;
(B) Genera;
(C) mOTU;
(D) Species.

Table S4 related to Figure 4. Heritability of gut microbial functions. The p-values of the ICC-filtered (r(MZ) > r(DZ); r(MZ) > 0 and p < 0.01) KOs, modules or pathways were controlled for multiple testing according to Storey's FDR method.
(A) KOs;
(B) KO modules;
(C) KO pathways.

Table S5 related to Figure 5. Heritability of T2D or obesity related strains or functions. The p-values of the ICC-filtered MLGs (r(MZ) > r(DZ); r(MZ) > 0 and p < 0.01) were controlled for multiple testing according to Storey's FDR method.
(A) T2D MLGs from Qin et al. (Qin et al., 2012).
(B) Pathways for butyrate synthesis.
(C) GMMs enriched in low microbial gene richness, i.e. low gene count (LGC), or high gene count (HGC) individuals according to Le Chatelier et al. (Le Chatelier et al., 2013).

Table S6 related to Figure 6. Non-redundant references genomes and SNP counts.