

Table of Contents

1		
2	Extended Data.....	2
3	Extended Data Table 1. Lifestyle information and socioeconomic status of the Chinese	
4	subjects.....	2
5	Extended Data Table 2. Correlation between GM biodiversity and BMD variation.....	4
6	Extended Data Table 3. BMD-associated bacterial species and covariates in various skeletal	
7	sites.....	5
8	Extended Data Table 4. BMD-associated SCFAs and covariates in various skeletal sites.....	10
9	Extended Data Table 5. Valeric acid-associated bacterial species.....	11
10	Extended Data Table 6. Potential causality of <i>Bacteroides vulgatus</i> exposure on valeric acid	
11	outcome with MR approach.....	14
12	Extended Data Table 7. Characteristics of US post-menopausal Caucasian female cohort....	15
13	Extended Data Table 8. Regression coefficients between <i>B.vulgatus</i> and varies sites of BMD	
14	16
14	Extended Data Table 9. Chinese subjects exclusion criteria.....	18
15	Extended Data Table 10. SNPs characteristics used in MR analysis.....	19
16	Supplementary methods.....	21
17	Metagenomic shotgun sequencing and annotation.....	21
18	Measurement of short chain fatty acids (SCFAs).....	22
19	BMD measurement and bone morphology evaluation in mice in vivo.....	24
20	Whole genome sequencing (WGS).....	25
21	Online resources.....	27
22	Abbreviations.....	28
23		
24		

1 **Extended Data**

2 **Extended Data Table 1. Lifestyle information and socioeconomic status of the**
 3 **Chinese subjects**

Items	None	Low	High	
Calcium supplementation	301/517 (58.2%)	139/517 (26.9%)	77/517 (14.9%)	
Exercise	154/517 (29.8%)	48/517 (9.3%)	315/517 (60.9%)	
Family annual income	139/517 (26.9%)	221/517 (42.7%)	157/517 (30.4%)	
	Yes	No		
Alcohol consumption	146/517 (28.2%)	371/517 (71.8%)		
Smoking	0	100%		
Physically demanding jobs	218/517 (42.2%)	299/517 (57.8%)		
	Less than high school graduate	High school graduate	College	Graduate
Education	153/517 (29.6%)	153/517 (29.6%)	195/517 (37.7%)	16/517 (3.1%)

4 Note:

5 Calcium supplementation: None - never had calcium supplementation; Low -
 6 occasional calcium supplementation; High - daily calcium supplementation. Exercise:
 7 None - do not have exercise habit; Low - exercise less than 2.5 hours per week; High -
 8 exercise more than 2.5 hours per week. Family annual income: None - low family
 9 income; Low - lower than the local average income but higher than the low family

1 income; High - higher than the local average income.

2

1 **Extended Data Table 2. Correlation between GM biodiversity and BMD**
 2 **variation**

Phenotypes	Shannon index			Optimal Kernel	
	partial γ	<i>p</i> -value	<i>q</i> -value	<i>p</i> -value of MIREKAT	<i>q</i> -value of MIREKAT
UD-R BMD	-0.081	0.072	0.144	0.005	0.028
UD-U BMD	-0.098	0.029	0.093	0.014	0.028
UD-RU BMD	-0.097	0.031	0.093	0.010	0.028
HTOT BMD	-0.043	0.344	0.516	0.812	0.812
FN BMD	0.004	0.931	0.931	0.733	0.812
L1-L4 BMD	0.031	0.493	0.592	0.475	0.713

3 Note:

4 partial γ - coefficient of the correlation between the Shannon index and BMD
 5 variation with partial Spearman correlation analysis; *p*-value - *p*-value of the
 6 correlation coefficient γ ; *q*-value - false discovery rate of the partial Spearman
 7 correlation analysis; Optimal Kernel - optimal kernel based on weighted and
 8 unweighted UniFrac distance matrices and Bray-Curtis distance; *p*-value of MIREKAT
 9 - *p*-value of the association between GM biodiversity and BMD variation with
 10 MiRKAT; *q*-value of MIREKAT - false discovery rate of the MIREKAT analyses.

11 Bolded BMDs were the significant ones associated with GM biodiversity (*p*-values <
 12 0.05).

13 Abbreviations:

14 GM - gut microbiota; MiRKAT - microbiome regression based on the kernel
 15 association test; BMD - bone mineral density; UD-R - ultra-distal radius; UD-U -
 16 ultra-distal ulna; UD-RU - ultra-distal radius and ulna; HTOT - left total hip; FN -
 17 femoral neck; L1-L4 - lumbar spine.

1 Extended Data Table 3. BMD-associated bacterial species and covariates in various skeletal sites

Variables	L1-L4 BMD		UD-RU BMD		HTOT BMD		FN BMD	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
YSM	-0.028	< 0.001	-0.016	< 0.001	-0.016	0.008	-0.013	0.027
Age	-0.002	0.384	-0.001	0.136	-0.001	0.594	-0.002	0.228
BMI	0.014	< 0.001	0.006	< 0.001	0.015	< 0.001	0.011	< 0.001
Exercise	0.025	0.001	0.003	0.246	0.014	0.015	0.016	0.007
Family annual income	0.023	0.017	0.000	0.999	0.008	0.259	0.014	0.048
FSH	0.000	0.657	0.000	0.439	0.000	0.126	0.000	0.013
Bacteroides_vulgatus	-0.027	0.032	-0.004	0.369	-0.011	0.253	-0.012	0.216
Faecalibacterium_prausnitzii	0.003	0.789	0.002	0.564	0.000	0.972	0.005	0.520
Bacteroides_uniformis	-0.002	0.867	-0.003	0.489	-0.004	0.629	-0.006	0.483
Bacteroides_fragilis	0.000	0.986	-0.001	0.853	0.011	0.402	0.001	0.951
Escherichia_coli	0.000	0.840	0.000	0.840	0.000	0.784	0.000	0.969
Bacteroides_stercoris	0.000	0.975	0.002	0.428	0.006	0.198	0.006	0.192

Variables	L1-L4 BMD		UD-RU BMD		HTOT BMD		FN BMD	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
Bacteroides_ovatus	-0.003	0.850	0.011	0.046	0.005	0.639	0.007	0.497
Bacteroides_massiliensis	-0.001	0.896	0.006	0.088	-0.008	0.259	-0.006	0.382
Eubacterium_eligens_cag72	-0.002	0.697	-0.001	0.464	0.001	0.744	0.000	0.957
Firmicutes_bacterium_cag65	-0.001	0.803	-0.001	0.707	0.001	0.798	0.001	0.868
Eubacterium_rectale	0.005	0.586	0.005	0.150	0.006	0.381	0.003	0.640
Phascolarctobacterium_sp_cag2	-0.001	0.801	-0.001	0.570	0.000	0.916	-0.003	0.381
Clostridium_sp_cag7	0.007	0.359	0.005	0.059	0.009	0.104	0.008	0.176
Bacteroides_coprocola	0.007	0.403	0.002	0.624	0.008	0.185	0.004	0.483
Roseburia_sp_cag18	-0.001	0.787	0.000	0.976	0.000	0.921	0.001	0.851
Alistipes_putredinis	0.001	0.885	0.000	0.838	0.002	0.600	0.002	0.653
Roseburia_inulinivorans	0.008	0.415	0.002	0.496	0.010	0.145	0.006	0.387
Roseburia_intestinalis	-0.008	0.440	-0.002	0.701	-0.004	0.606	-0.009	0.268
Bacteroides_caccae	-0.003	0.776	-0.004	0.385	0.000	0.976	-0.001	0.881
Subdoligranulum_variabile	-0.005	0.477	0.001	0.655	0.001	0.857	0.000	0.999

Variables	L1-L4 BMD		UD-RU BMD		HTOT BMD		FN BMD	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
Bacteroides_xylanisolvans	0.003	0.839	-0.004	0.418	0.006	0.555	0.006	0.563
Bacteroides_thetaiotaomicron	0.027	0.021	0.004	0.410	0.003	0.700	0.010	0.266
Bacteroides_finegoldii	-0.018	0.192	-0.008	0.116	-0.012	0.267	-0.009	0.364
Bacteroides_eggerthii	0.009	0.313	0.002	0.480	-0.001	0.907	0.001	0.834
Parabacteroides_distasonis	-0.028	0.091	-0.004	0.558	-0.015	0.219	-0.015	0.215
Parabacteroides_merdae	0.004	0.615	-0.002	0.451	0.001	0.850	-0.002	0.717
Bacteroides_cellulosilyticus	0.014	0.202	0.005	0.292	0.013	0.129	0.004	0.613
Bacteroides_intestinalis	-0.015	0.278	-0.006	0.283	-0.021	0.043	-0.010	0.318
Hungatella_hathewayi	-0.003	0.843	-0.004	0.475	0.010	0.424	0.004	0.736
Butyrateproducing_bacterium_ss3/4	-0.024	0.040	-0.005	0.258	-0.022	0.009	-0.021	0.014
Clostridium_bolteae	0.045	0.036	0.014	0.072	0.023	0.152	0.020	0.201
Bacteroides_sp_9_1_42faa	0.033	0.031	0.006	0.289	0.017	0.150	0.021	0.063
Roseburia_hominis	0.015	0.179	-0.001	0.781	0.003	0.755	0.003	0.682
Clostridium_sp_cag43	0.005	0.601	0.004	0.215	-0.001	0.847	0.000	0.993

Variables	L1-L4 BMD		UD-RU BMD		HTOT BMD		FN BMD	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
uncultured_bacterium	0.000	NA	0.000	NA	0.000	NA	0.000	NA
Eubacterium_ventriosum	-0.003	0.801	-0.001	0.880	-0.002	0.798	0.005	0.506
Blautia_sp_cag37	-0.002	0.843	-0.005	0.122	-0.006	0.399	-0.001	0.857
Clostridium_clostridioforme	-0.028	0.418	-0.017	0.196	-0.025	0.326	-0.024	0.354
Peptoclostridium_difficile	-0.049	0.095	0.003	0.761	0.012	0.575	0.009	0.693
Blautia_wexlerae	-0.003	0.865	-0.003	0.646	-0.003	0.841	-0.012	0.356
Ruminococcus_lactaris	0.010	0.475	0.005	0.361	0.008	0.421	0.007	0.473
Bilophila_wadsworthia	0.000	NA	0.000	NA	0.000	NA	0.000	NA
Flavonifractor_plautii	-0.002	0.903	-0.003	0.491	-0.012	0.186	-0.010	0.273
Bacteroides_sp_3_1_33faa	0.000	NA	0.000	NA	0.000	NA	0.000	NA
Tyzzarella_nexilis	0.017	0.388	0.004	0.606	0.014	0.366	0.005	0.737
Dorea_formicigenerans	0.018	0.535	-0.003	0.779	-0.021	0.353	0.003	0.900
Lachnospiraceae_bacterium_a4	0.000	NA	0.000	NA	0.000	NA	0.000	NA

1 Note:

- 1 β - regression coefficient of the association between variable and human BMD variation; *p*-value - *p*-value of the regression coefficient.
- 2 Bolded species were the significant ones associated with BMD (*p*-values < 0.05).
- 3 Abbreviations:
- 4 YSM - years since menopause; BMD - bone mineral density; L1-L4 - lumbar spine; HTOT - left total hip; UD-RU - ultra-distal radius and ulna;
- 5 FN - femoral neck; BMI - body mass index; FSH - follicle stimulating hormone.
- 6

1 **Extended Data Table 4. BMD-associated SCFAs and covariates in various**
 2 **skeletal sites**

Variables	L1-L4 BMD		HTOT BMD		UD-RU BMD	
	β	<i>p</i> -value	β	<i>p</i> -value	β	<i>p</i> -value
Intercept	0.876	< 0.001	0.609	< 0.001	0.320	< 0.001
Caproic acid	0.008	0.726	-0.010	0.531	-0.002	0.854
Isovaleric acid	-0.015	0.284	0.005	0.660	-0.002	0.724
Butyric acid	-0.037	0.204	-0.024	0.264	-0.013	0.233
Acetic acid	-0.002	0.889	0.001	0.936	-0.001	0.884
Isobutyric acid	-0.011	0.468	-0.008	0.455	-0.001	0.871
Valeric acid	0.040	0.029	0.023	0.089	0.009	0.196
YSM	-0.025	0.001	-0.017	0.002	-0.016	< 0.001
Age	-0.003	0.286	-0.001	0.640	-0.001	0.157
BMI	0.015	< 0.001	0.017	< 0.001	0.007	< 0.001
Exercise	0.027	< 0.001	0.014	0.012	0.003	0.249

3 Note:

4 β - regression coefficient of the association between variable and human BMD variation;

5 *p*-value - *p*-value of the regression coefficient.

6 Bolded contents were the significant ones associated with BMD (*p*-values < 0.05).

7 Abbreviations:

8 SCFAs - short chain fatty acids; YSM - years since menopause; BMD - bone mineral

9 density; L1-L4 - lumbar spine; HTOT - left total hip; UD-RU - ultra-distal radius and

10 ulna; BMI - body mass index.

1 **Extended Data Table 5. Valeric acid-associated bacterial species**

Variables	β	<i>p</i> -value	Lower	Upper
YSM	-0.024	0.336	-0.072	0.025
Age	0.007	0.371	-0.008	0.023
BMI	-0.013	0.117	-0.029	0.003
FSH	0.000	0.560	-0.002	0.001
Bacteroides_vulgatus	-0.111	0.006	-0.189	-0.032
Faecalibacterium_prausnitzii	-0.003	0.916	-0.065	0.059
Bacteroides_uniformis	0.023	0.500	-0.044	0.091
Bacteroides_fragilis	0.060	0.267	-0.046	0.165
Escherichia_coli	0.000	NA	NA	NA
Bacteroides_stercoris	-0.010	0.585	-0.048	0.027
Bacteroides_ovatus	-0.054	0.225	-0.141	0.033
Bacteroides_massiliensis	0.028	0.325	-0.028	0.085
Eubacterium_eligens_cag72	0.005	0.719	-0.023	0.033
Firmicutes_bacterium_cag65	0.021	0.229	-0.014	0.056
Eubacterium_rectale	-0.023	0.409	-0.078	0.032
Phascolarctobacterium_sp_cag2	-0.004	0.759	-0.033	0.024
Clostridium_sp_cag7	-0.041	0.087	-0.089	0.006
Bacteroides_coprocola	-0.018	0.503	-0.069	0.034
Roseburia_sp_cag18	-0.013	0.421	-0.045	0.019
Alistipes_putredinis	0.054	< 0.001	0.024	0.084
Roseburia_inulinivorans	-0.005	0.856	-0.062	0.051
Roseburia_intestinalis	-0.029	0.403	-0.096	0.038
Bacteroides_caccae	0.089	0.009	0.023	0.155
Subdoligranulum_variabile	0.005	0.837	-0.040	0.049
Bacteroides_xylanisolvens	-0.021	0.625	-0.103	0.062

Variables	β	<i>p</i>-value	Lower	Upper
Bacteroides_thetaiotaomicron	0.022	0.544	-0.049	0.093
Bacteroides_finegoldii	-0.038	0.389	-0.124	0.048
Bacteroides_eggerthii	-0.006	0.828	-0.059	0.047
Parabacteroides_distasonis	-0.012	0.812	-0.115	0.090
Parabacteroides_merdae	-0.013	0.618	-0.067	0.040
Bacteroides_cellulosilyticus	0.083	0.021	0.013	0.153
Bacteroides_intestinalis	-0.121	0.006	-0.207	-0.034
Hungatella_hathewayi	-0.030	0.550	-0.129	0.069
Butyrateproducing_bacterium_ss3/4	0.030	0.399	-0.040	0.101
Clostridium_bolteae	-0.008	0.899	-0.140	0.123
Bacteroides_sp_9_1_42faa	0.045	0.359	-0.051	0.140
Roseburia_hominis	0.017	0.630	-0.052	0.086
Clostridium_sp_cag43	0.009	0.753	-0.048	0.067
uncultured_bacterium	0.000	NA	NA	NA
Eubacterium_ventriosum	0.022	0.520	-0.046	0.091
Blautia_sp_cag37	-0.003	0.913	-0.057	0.051
Clostridium_clostridioforme	0.134	0.216	-0.079	0.347
Peptoclostridium_difficile	0.056	0.543	-0.125	0.237
Blautia_wexlerae	-0.036	0.514	-0.143	0.072
Ruminococcus_lactaris	0.053	0.225	-0.033	0.139
Bilophila_wadsworthia	0.000	NA	NA	NA
Flavonifractor_plautii	-0.066	0.091	-0.142	0.011
Bacteroides_sp_3_1_33faa	0.000	NA	NA	NA
Tyzzarella_nexilis	-0.072	0.253	-0.195	0.051
Dorea_formicigenerans	-0.020	0.827	-0.203	0.162
Lachnospiraceae_bacterium_a4	0.000	NA	NA	NA

1 Note:

2 β - regression coefficient of the association between variables and valeric acid; *p*-value -
3 *p*-value of the regression coefficient; Lower - lower limit of 95% confidence interval;
4 Upper - upper limit of 95% confidence interval.

5 Bolded species were the significant ones associated with valeric acid (*p*-values < 0.05).

6 Abbreviations:

7 YSM - years since menopause; BMI - body mass index; FSH - follicle stimulating
8 hormone.

1 **Extended Data Table 6. Potential causality of *Bacteroides vulgatus* exposure on**
 2 **valeric acid outcome with MR approach**

MR methods	β	Standard Error	<i>p</i> -value
Weighted median method	-0.116	0.038	0.002
MaxLik	-0.075	0.033	0.022
IVW	-0.071	0.032	0.025
MR-Egger	-0.031	0.069	0.655
MR-Egger (intercept)	-0.023	0.035	0.517

3 Note:

4 β - regression coefficient of the association between *Bacteroides vulgatus* (as
 5 exposure) and valeric acid (as outcome) with various MR methods; *p*-value - *p*-value
 6 of the regression coefficient.

7 Abbreviations:

8 MR - Mendelian randomization; MaxLik - maximum likelihood estimation; IVW -
 9 inverse-variance weighted.

10

1 **Extended Data Table 7. Characteristics of US post-menopausal Caucasian female**
 2 **cohort**

Phenotypes	Max	Min	Mean	Standard Deviation
Age (years)	80.52	60.12	66.98	5.65
BMI (kg/m ²)	49.99	17.47	27.84	8.50
L1-L4 BMD (g/cm ²)	1.53	0.63	0.92	0.17
HTOT BMD (g/cm ²)	1.21	0.44	0.80	0.12
FN BMD (g/cm ²)	0.93	0.43	0.67	0.11
UD-R BMD (g/cm ²)	0.52	0.21	0.37	0.06
UD-U BMD (g/cm ²)	0.39	0.14	0.28	0.05
	Yes		No	
Alcohol drinking	47/59 (79.7%)		12/59 (20.3%)	
Smoking	23/59 (39.0%)		36/59 (61.0%)	
Regular exercise	44/59 (74.6%)		15/59 (25.4%)	

3 **Abbreviations:**

4 BMI - body mass index; BMD - bone mineral density; L1-L4 - lumbar spine; HTOT -
 5 left total hip; FN - femoral neck; UD-R - ultra-distal radius; UD-U - ultra-distal ulna.

6

1 **Extended Data Table 8. Regression coefficients between *B.vulgatus* and varies**
2 **sites of BMD**

Phenotype	Chinese cohort		the US cohort	
	β	<i>p</i> -value	β	<i>p</i> -value
Spine (L1-L4) BMD	-0.027	0.032	-0.002	0.862
L1 BMD	-0.027	0.027	-	-
L2 BMD	-0.025	0.057	-	-
L3 BMD	-0.031	0.024	-	-
L4 BMD	-0.029	0.042	-	-
R-UD BMD	-0.004	0.409	-0.007	0.115
U-UD BMD	-0.004	0.493	-0.003	0.411
RTOT BMD	-0.006	0.212	-0.008	0.147
RU-UD BMD	-0.004	0.369	-	-
R-33 BMD	-0.013	0.039	-	-
U-33 BMD	-0.005	0.520	-	-
RU-33 BMD	-0.009	0.153	-	-
UTOT BMD	-0.003	0.598	-	-
RUTOT BMD	-0.005	0.328	-	-
R-mid BMD	-	-	-0.007	0.171
R-13 BMD	-	-	-0.006	0.339
HTOT BMD	-0.011	0.253	-0.018	0.029
FN BMD	-0.012	0.216	-0.012	0.084
FN (upper) BMD	-0.012	0.200	-	-
FN (lower) BMD	-0.011	0.279	-	-
Hip (wards) BMD	-0.019	0.073	-	-
Hip (troch) BMD	-0.011	0.198	-	-
Hip (FS) BMD	-0.009	0.449	-	-

WB TOT	-	-	-0.014	0.133
WB (sub) TOT	-	-	-0.015	0.085

1 Note:

2 BMD - bone mineral density, L - lumbar spine, R - radius, U - ulna, UD - ultra-distal, TOT - total,

3 FN - femoral neck, FS - femoral shaft, WB - whole body, WB (sub) - whole body except head.

4

1 **Extended Data Table 9. Chinese subjects exclusion criteria**

Number	Exclusion criteria
1	Used antibiotics, oestrogens, anticonvulsant or proton pump inhibitor medications in the past three months;
2	Underwent hysterectomy or bilateral ovariectomy;
3	Serious residuals from cerebral vascular disease;
4	Diabetes mellitus, except for easily controlled, non-insulin dependent diabetes mellitus;
5	Chronic renal disease manifest by serum creatinine > 1.9 mg/dL;
6	Chronic liver diseases;
7	Significant chronic lung disease;
8	Alcohol abuse as defined by those who drink alcohol regularly and cannot control themselves and become intoxicated at least once a week.
9	Corticosteroid therapy at pharmacologic levels currently, or for more than 6 months duration at any time;
10	Treatment with anticonvulsant therapy currently, or for more than 6 months duration at any time;
11	Evidence of other metabolic diseases or inherited bone diseases such as hyper- or hypoparathyroidism, Paget's disease, osteomalacia, osteogenesis imperfecta, or others;
12	Rheumatoid arthritis, except for minor cases that involve only hand joint and wrist;
13	Recent major gastrointestinal disease (within the past year) such as celiac disease, post-gastrectomy, Crohn's disease, ulcerative colitis;
14	Any other disease, treatment (e.g., bisphosphonates), or condition that would be an apparent non-genetic factor underlying the variation of BMD.

2

1 **Extended Data Table 10. Characteristics of the SNP used in MR analysis**

SNP	Exposure (<i>B.vulgatus</i>)				Outcome (valeric acid)			
	effect_allele	β	<i>p</i> -value	se	effect_allele	β	<i>p</i> -value	se
rs688811	T	-1.288	3.95E-06	0.276	T	0.038	0.776	0.132
rs3219142	A	-1.019	3.63E-07	0.198	A	0	0.999	0.095
rs11682148	A	-1.089	6.40E-07	0.216	A	-0.060	0.555	0.101
rs77875633	G	-1.471	6.63E-07	0.292	G	0.194	0.165	0.139
rs36201997	A	-0.794	3.77E-06	0.170	A	-0.130	0.111	0.082
rs61870507	A	-0.510	2.77E-06	0.107	A	0.064	0.215	0.051
rs140807263	A	-1.146	3.70E-06	0.245	A	0.251	0.036	0.120
rs1077028	C	-0.401	1.34E-07	0.075	C	0.047	0.195	0.036
rs35903260	T	-0.316	8.10E-07	0.063	T	0.059	0.052	0.030
rs8049150	G	-1.069	4.27E-06	0.230	G	0.046	0.676	0.109
rs1452791	A	0.313	5.30E-06	0.068	A	0.053	0.105	0.032
rs4892102	G	-0.366	9.16E-06	0.082	G	0.092	0.016	0.038

rs17705279	G	-0.796	1.71E-06	0.164	G	0.094	0.231	0.078
rs6516034	A	0.351	4.25E-06	0.075	A	-0.045	0.2142	0.036
rs2051388	G	0.386	6.46E-06	0.085	G	-0.041	0.316	0.040

1

1 **Supplementary methods**

2 **Metagenomic shotgun sequencing and annotation**

3 *DNA library construction*

4 We constructed a faecal DNA library by the TruSeq Nano DNA LT Library
5 Preparation Kit (FC-121-4001, Illumina, San Diego, CA, USA). Faecal DNA was
6 fragmented by dsDNA Fragmentase (NEB, M0348S, Massachusetts, USA) and
7 incubated at 37 °C for 30 min. Then, we used the fragmented cDNA to construct
8 libraries. Blunt-end DNA fragmentation and size selection were performed with
9 provided sample purification beads. An A-base was added to the blunt ends of each
10 strand for the preparation of ligation to indexed adapters. These adapters also
11 contained sequencing primer hybridization sites for single, paired-end, and indexed
12 reads. The ligated products were amplified with polymerase chain reaction (PCR)
13 under the following conditions: initial denaturation at 95 °C for 3 min, followed by 8
14 cycles of 98 °C for 15 sec (denaturation), 60 °C for 15 sec, 72 °C for 30 sec
15 (extension), then a final elongation at 72 °C for 5 min.

16

17 *Raw data preprocessing and cleaning*

18 Raw sequencing reads were processed in multiple cleaning steps. First, we removed
19 sequencing adapters from sequencing reads by Cutadapt v1.9. Second, we trimmed
20 low quality reads by Fqtrim v0.94. We used a sliding-window (size = 6 bp) to
21 calculate average quality of the bases within this window, and trimmed 3' end of reads

1 when the average quality value dropped below 20. We also discarded the reads when
2 the length was less than 100 bp and the percentage of "N" was larger than 5% after
3 trimming. Third, we aligned reads to the host genome by using Bowtie2 v2.2.0, and
4 removed host genomic contamination. Once quality-filtered reads were obtained, they
5 were *de novo* assembled to construct metagenomes for each sample by SPAdes
6 v3.10.0. The coding sequences (CDS) of metagenomic contigs were predicted by
7 MetaGeneMark v3.26. The CDS of all samples were clustered by CD-HIT v4.6.1 to
8 obtain unigenes.

9

10 *Taxonomic and functional annotation of unigenes*

11 We used “blastp” function of DIAMOND for unigene alignments. It determines bit
12 score and expected value (*E*-value) of the computed alignment. The bit score gives an
13 indication of how good the alignment (hit) is; the higher the score, the better the
14 alignment. The *E*-value gives an indication of the statistical significance of a given
15 pairwise alignment; the lower the *E*-value, the more significant the hit. We selected
16 the best hit with the highest bit score from the all potential hits ($E\text{-values} \leq 1 \times 10^{-5}$)
17 as the respective KEGG Orthology (KO) for each unigene. KOs were further mapped
18 to GM-associated functional KEGG modules.

19

20 **Measurement of short chain fatty acids (SCFAs)**

21 *Gas chromatography-tandem mass spectrometry (GC-MS/MS)*

1 Serum samples were vortex-mixed with 36% phosphoric acid solution and further
2 extracted supernatants by liquid-liquid extraction with methyl tert-butyl ether which
3 containing SCFA stock solutions. Detailed conditions for GC-MS/MS were as follows:
4 setting the injector temperature at 240 °C, keeping the initial oven temperature at
5 90 °C for 1 min, then gradually raising it to 140 °C, 160 °C, 200 °C, and 240 °C at a
6 rate of 10 °C/min, 5 °C/min, 15 °C/min, and 10 °C/min, respectively. Pure helium was
7 used as a carrier gas at 1.0 ml/min rate. The main conditions of
8 mass spectrometry included electron impact ion source and multi-reaction monitoring
9 scan mode. The temperature of the transfer line, ion source, and quad were 240 °C,
10 230 °C, and 150 °C, respectively. The electron energy was 70 eV, and the solvent
11 delayed 2.4 min.

12

13 *Quality control (QC) and intra-day/inter-day accuracy*

14 Process of quality control (QC) and intra-day/inter-day accuracy were as follows: the
15 QC samples were composed of SCFA stock solutions dissolved in methyl *tert*-butyl
16 ether. Then the QC samples were processed in parallel with test samples to analyze
17 detection stability and repeatability under the same process with an injection volume
18 of 2 µl. Three QC samples were continuously injected to test the instrument stability.
19 For every 10 test samples injected, a QC sample was inserted to check for the
20 repeatability of the instrument.

21 Intra-day and inter-day accuracy were evaluated and reported as CV% of repeatability
22 at the concentration of each SCFA. The intra-day accuracy is 0.42% - 3.64%, and the

1 inter-day accuracy is 1.12% - 3.40%, indicating a good stability of the instrument.

2

3 *Qualitative and quantitative analyses of SCFAs*

4 We performed qualitative and quantitative analysis of SCFAs with Agilent Mass
5 Hunter software. By using a stock solution containing mixtures of SCFAs, five
6 calibration standards (concentration range from 0.005 to 8 mg/L) were prepared. Then
7 we performed GC-MS/MS measurement as described above and integrated the
8 obtained signals (e.g., retention time and peak area) to calculate relative retention time
9 and area ratios. Meanwhile, the calibration curves were constructed by plotting the
10 peak area versus concentration for each individual SCFA. And the slopes of the
11 calibration curves were determined by performing linear regression analysis. In
12 addition, average area ratio of blank samples was used as background signal/intercept.
13 Finally, we calculated the concentrations of SCFAs with the area ratio, average area
14 ratio blank samples, and slope (see Formula 1).

$$15 \quad \text{Concentration} = \frac{\text{area ratio} - \text{average area ratio blank samples}}{\text{slope}} \quad (1)$$

16

17 **BMD measurement and bone morphology evaluation in mice in vivo**

18 *Micro-computed tomography (μ CT)*

19 The scanning parameters were as follows: standard model, low X-ray, 96 $\mu\text{m} \times 361$
20 slice, scan conditions with 48 mm, 360° rotation, and asynchronous.

21

1 *Histomorphometry*

2 The bones were prefixed in 4% paraformaldehyde (Solarbio, Beijing, China) at first.
3 Samples were then incubated with 10% EDTA (pH 7.4, Solarbio, Beijing, China) until
4 the bone was easily penetrated through by a needle without any force. Subsequently,
5 samples were dehydrated in graded ethanol solutions and embedded in paraffin. Then,
6 for each specimen, 5 serial sections (5 μm thick) were cut on a microtome and stained
7 with hematoxylin-eosin. Finally, the specimens were examined and photographed
8 using a high-quality microscope.

9

10 **Whole genome sequencing (WGS)**

11 Qualified genomic DNA samples were randomly fragmented by Covaris technology
12 and the fragment of 350 bp was obtained after fragment selection. End repair of DNA
13 fragments was performed and an "A" base was added at the 3'-end of each strand.
14 Adapters were then ligated to both ends of the DNA fragments, and ligation-mediated
15 PCR was performed for amplification, single strand separation and cyclization. The
16 rolling circle amplification was performed to produce DNA Nanoballs. The qualified
17 pair-end reads were read through on the BGISEQ-500 platform. High throughput
18 sequencing was performed for each library to ensure that each sample meets the
19 average sequencing coverage requirement. Sequencing-derived raw image files were
20 processed by BGISEQ-500 Base-calling software for base-calling with default
21 parameters. Sequence data of each individual were generated as paired-end reads,

1 which is defined as "raw data" and stored in FASTQ format. Clean reads were
2 produced by data filtering, including removed reads containing sequencing adapter,
3 removed reads whose low-quality base ratio (base quality less than or equal to 5) was
4 more than 50%, removed reads whose unknown base ("N" base) ratio was more than
5 10%. Then, clean reads of each sample were mapped to the human reference genome
6 (GRCh38/HG38). Burrows-Wheeler Aligner (BWA) software was used for sequence
7 alignment. To ensure accurate variant calling, we followed recommended best
8 practices for variant analysis with the Genome Analysis Toolkit (GATK).

9

1 Online resources

Softwares	URL
Agilent Mass Hunter	https://www.agilent.com/en/promotions/masshunter-mass-spec
BWA	https://sourceforge.net/projects/bio-bwa/
CD-HIT v4.6.1	http://weizhongli-lab.org/cd-hit/
Cutadapt v1.9	https://cutadapt.readthedocs.io/
DIAMOND	http://ab.inf.uni-tuebingen.de/software/diamond/
Fqtrim v0.94	http://ccb.jhu.edu/software/fqtrim/
GATK	https://www.broadinstitute.org/gatk/
IOF	https://www.iofbonehealth.org/
KEGG dataset	https://www.genome.jp/kegg/
MetaGeneMark	http://exon.gatech.edu/GeneMark/
PLINK 1.9	https://www.cog-genomics.org/plink/1.9/
R software v3.5.1	https://www.r-project.org/
SPAdes v3.10.0	http://cab.spbu.ru/software/spades/
Stata 14	https://www.stata.com/

2

1 Abbreviations

Abbreviations	Full forms	Abbreviations	Full forms
γ	correlation coefficient	β	regression coefficient
μ CT	micro-computed tomography	ALP	alkaline phosphatase
AR-S	alizarin red S	BMD	bone mineral density
BMI	body mass index	BWA	Burrows-Wheeler Aligner
CDS	coding sequence	CTX-I	C-telopeptide of type I collagen
CV	coefficient of variation	DMEM	Dulbecco's Modified Eagle Medium
DXA	dual energy X-ray absorptiometry	FBS	fetal bovine serum
ECM	extracellular matrix	ED	Extended Data
FN	femoral neck	FSH	follicle stimulating hormone
GC-MS/MS	gas chromatography-tandem mass spectrometry	GM	gut microbiota
GWAS	genome-wide association study	HDAC	histone deacetylase
HTOT	left total hip	IMP	inosine monophosphate
IOF	International osteoporosis foundation	IVW	inverse-variance weighted
IVs	Instrumental variables	KO	KEGG Orthology
L1-L4	lumbar spine	LD	linkage disequilibrium
MaxLik	maximum likelihood estimation	MiRKAT	Microbiome Regression-based Kernel Association Test

Abbreviations	Full forms	Abbreviations	Full forms
MR	Mendelian randomization	NS	normal saline
OC	osteocalcin	OP	osteoporosis
OVX	ovariectomy	PCR	polymerase chain reaction
PMOP	postmenopausal osteoporosis	QC	quality control
RANKL	receptor activator of nuclear factor- κ B ligand	UD-RU	ultra-distal radius and ulna
UD-R	ultra-distal radius	UD-U	ultra-distal ulna
SCFAs	short chain fatty acids	SD	standard deviation
SNPs	single nucleotide polymorphisms	TPM	transcripts per kilobase million
TRAP	tartrate-resistant acid phosphatase	VA	valeric acid
WGS	whole genome sequencing	YSM	years since menopause

1