

SUPPLEMENTAL MATERIAL

Data S1.

Supplemental Methods

Induced TMAO Production in Healthy Adults

Elastic Net Modeling

The goals of the present analyses were first, to test whether plasma TMAO is associated with the abundance of specific strains of bacteria in the fecal microbiome and second, identify microbial enzymes or strains whose associations with TMAO are comparable to more easily measured predictors such as sex and diet. We were especially focused on identifying simple linear relationships amenable to future experiments in model systems. Finally, we expected a high degree of collinearity when comparing closely related strains and gene homologs.

Elastic net accomplishes both goals by “selecting” a relatively small number of predictors that minimize model error and reporting an effect size (β coefficient) for each predictor that can be directly compared.

For each elastic net run, we predicted either plasma TMAO (Figure 4), the change in plasma TMAO between non-meat and red meat diet (Δ TMAO, Figure 4), or choline consumption in a fecal culture (18-hour d6-choline, Figure 6). The predictors included either the clinical measurements listed in Table 1, the homologs of a TMA-related microbial gene cluster (cai, cut, tor, yea), fecal TMA-lyase enzyme activities, or, as a sanity check, TMAO fractional excretion rate (see Response #2). Interactions and other nonlinear terms were not included, as they would make interpretation more difficult.

Each outcome-predictor set pair was analyzed as follows. The pseudocode below summarizes the elastic net analysis.

```
main
1 Define a predictor set  $P = \{ \dots \}$  and outcome  $\vec{O}$ 
2 Define training sets  $T_{\text{train}} = \{ T_1, T_2, \dots, T_{100} \}$  and  $T_{\text{test}}$ 
   where  $T_i = \{ p_i \subset P, \vec{o}_i \subset \vec{O} \}$ 
3 for each training set  $T_i$  do
4   | Train linear model
5   | Predict testing set  $T_{\text{test}_i}$ 
6   | Compute adjusted  $R^2$  and RMSE
7 end
```

```

Train
1 Define a grid of alpha and lambda values
2 for each {α, λ}:
3   Randomly divide training data into ten folds
4   for each fold do
5     Hold out the nth fold
6     Preprocess the remaining nine folds
7     Fit linear model with preprocessed data
8     Predict data in the nth fold
9   end
10  Compute average RMSE across 10 folds
11 end
12 Select the {α, λ} set with smallest RMSE
13 Fit final model using all training data and optimal {α, λ}

```

From the set of 461 clinical visits, those visits with complete data were randomly divided into training (70%) and testing (30%) sets. Due to the wide variation in model performance, we repeated training/testing assignment with replacement to produce a total of 100 training/testing sets. Within each training/testing set, model error and variance explained were quantified as follows.

Elastic net models were trained and tested in R 3.5.0 by invoking the `glmnet` function via the `train` function of the `caret` package. Briefly, model parameters are tuned by minimizing a cost function for each of combination of 9 out of 10 folds within the training set, and the root-mean-square error (RMSE) is determined for the prediction of the tenth, held-out fold. The optimal parameters producing the lowest RMSE are used to produce a model using the entire training set, and this model is then used to predict the held-out testing set. The error and variance explained by the model in the testing set is then computed by comparing predicted to measured values.

A grid of λ and α values based on preliminary testing was defined for $\lambda = \{0, 0.1, 0.2 \dots 12\}$ and $\alpha = \{0, 0.1, 0.2 \dots 1\}$. These two parameters were optimized, or tuned, by quantifying prediction error for all 1200 combinations of λ and α values using 10-fold cross-validation.

For cross-validation, the training set was divided into ten folds, and each fold was iteratively held out. The remaining nine folds are preprocessed independently of the held out fold: continuous predictors are scaled and centered, categorical predictors are recoded as “dummy variables”, and predictors with missing or zero data in >30% samples are excluded.

Within each cross-validation, $\hat{\beta}$ coefficients were computed to minimize the cost function C using the given values of λ and α .

$$C(\hat{\beta}, \lambda, \alpha) = \frac{1}{2n} \sum_i (\hat{y}_i - y_i)^2 + \lambda \left(\frac{1-\alpha}{2} \sum \hat{\beta}^2 + \alpha \sum |\hat{\beta}| \right)$$

Each 10-fold cross validation produced 10 models tested on held out folds to compute error as root-mean-square error (RMSE), a value used when predicting data with a skewed distribution, such as TMAO.

$$RMSE = \sqrt{\frac{1}{n} \sum_i (\hat{y}_i - y_i)^2}$$

Nearly all the measured values in our manuscript appear to follow a right-skewed distribution. We wanted to give more weight to larger absolute errors to construct a more clinically useful model. Larger absolute errors will hamper the clinical goal of risk stratification more than smaller absolute errors.

The values of λ and α with the least average RMSE were deemed optimal, and the final model was trained as above with all training data using the optimal values of λ and α .

The resulting linear model was then used to predict the data in the held-out testing data. These predicted values are compared to actual values using RMSE and adjusted R-squared. Predictor sets that produce higher median adjusted R-squared and/or lower median RMSE across 100 testing sets are said to improve the model. The Wilcox test is used to assess a difference in medians.

Table S1. Proportion of samples in which gene or gene cluster is detected. Metagenomic sequencing reads from 405 fecal samples were aligned to genes related to TMA production as well as the associated gene cluster. In almost every case, conserved gene clusters are more frequently detected than the genes of interest. P values represent Bonferroni-adjusted Chi-square tests comparing each gene to the related gene cluster.

Gene Cluster	Gene	Detection Rate (%)	P (Chi-Square)
cai		99.8 (404/405)	
	caiA	76.8 (311/405)	< 0.0001
	caiB	94.8 (384/405)	0.0004
	caiC	93.6 (379/405)	< 0.0001
cnt		13.1 (53/405)	
	cntAB	12.8 (52/405)	>0.99
cut		97.3 (394/405)	
	cutC	94.8 (384/405)	>0.99
	cutD	38.8 (157/405)	< 0.0001
grd		97.8 (396/405)	
	grdH	94.3 (382/405)	0.19
tor		95.3 (386/405)	
	torA	64.9 (263/405)	< 0.0001
yea		95.1 (385/405)	
	yeaW	65.2 (264/405)	< 0.0001
	yeaX	45.2 (183/405)	< 0.0001

Table S2. Universal single copy genes used for alternative normalization. Manor and Banorstein (1) selected prokaryotic orthologous groups with exactly one copy in almost every prokaryotic genome listed by KEGG. We used only the top ten universal single copy genes to quantify relative bacterial genome count in fecal samples as an alternative form of normalization.

KEGG Orthology	COG	Annotation	Prokaryotic Genomes With Gene (%)	Prokaryotic Genomes With >1 Copy (%)
K02988	COG0098	small subunit ribosomal protein S5	6112 (99)	5 (0.08)
K02994	COG0096	small subunit ribosomal protein S8	6111 (99)	5 (0.08)
K02933	COG0097	large subunit ribosomal protein L6	6108 (99)	8 (0.13)
K02931	COG0094	large subunit ribosomal protein L5	6097 (99)	13 (0.21)
K02863	COG0081	large subunit ribosomal protein L1	6096 (99)	5 (0.08)
K02876	COG0200	large subunit ribosomal protein L15	6094 (99)	8 (0.13)
K02967	COG0052	small subunit ribosomal protein S2	6093 (98)	7 (0.11)
K02864	COG0244	large subunit ribosomal protein L10	6081 (98)	8 (0.13)
K02881	COG0256	large subunit ribosomal protein L18	6076 (98)	6 (0.10)
K03110	COG0552	fused signal recognition particle receptor	5961 (96)	8 (0.13)

Table S3. Spearman correlations with plasma TMAO among all samples. Four candidate genes were quantified using KEGG SSDB gene clusters. Of 274 gene cluster homologs, 60 were detected in >70% of samples. Spearman correlations were calculated between these 60 homolog abundances and plasma TMAO. Gene clusters are named with KEGG org identifier and genomic coordinates. Universal single-copy gene normalization (USCG) utilized ten USCGs.

Gene Cluster	RPKM			USCG		
	Rho	P	P-Bonferroni	Rho	P	P-Bonferroni
cai-eih-27336-52938	0.17	0.00065	0.04	0.17	0.00086	0.053
cai-senv-57098-100887	0.16	0.0012	0.073	0.16	0.0016	0.099
cai-sena-57098-100887	0.16	0.0018	0.11	0.16	0.002	0.12
cai-seeb-3884824-3928761	0.16	0.0011	0.069	0.17	0.00094	0.057
cai-cwe-3656832-3683530	0.16	0.0011	0.07	0.17	0.00083	0.051
cai-ecv-27317-52919	0.15	0.0026	0.16	0.15	0.0032	0.19
cai-sek-57350-101254	0.15	0.0028	0.17	0.15	0.0035	0.21
cai-senq-57098-100887	0.15	0.0033	0.2	0.15	0.0036	0.22
cai-sel-57340-101063	0.15	0.0029	0.18	0.15	0.0029	0.18
cai-senc-1519486-1563447	0.15	0.0029	0.18	0.15	0.0034	0.21
cai-ses-2844264-2873996	0.14	0.0043	0.26	0.14	0.004	0.24
cai-ebt-3409854-3439396	0.14	0.005	0.31	0.14	0.0041	0.25
cai-edh-3821075-3845551	0.13	0.013	0.77	0.12	0.016	0.98
cai-ecz-27336-52937	0.13	0.0074	0.45	0.14	0.0055	0.34
cai-setc-3466245-3510218	0.12	0.02	1	0.11	0.023	1
cai-cko-3102867-3134413	0.12	0.018	1	0.12	0.018	1
cai-ecl-3947888-3972396	0.11	0.022	1	0.12	0.02	1
cai-eci-27336-52938	0.11	0.034	1	0.11	0.036	1
cai-sea-57341-101275	0.11	0.029	1	0.11	0.025	1
cai-senj-2127653-2171775	0.11	0.023	1	0.12	0.019	1
cutC-Pooled	0.11	0.029	1	0.11	0.023	1
cai-ecoi-27459-53061	0.1	0.042	1	0.1	0.041	1
cai-seep-57183-100834	0.1	0.039	1	0.1	0.038	1
cai-senl-57098-100886	0.1	0.042	1	0.1	0.044	1
cai-seno-57098-100887	0.097	0.055	1	0.095	0.058	1
cai-sens-57342-100598	0.097	0.054	1	0.097	0.054	1
grd-cace-803252-808152	0.096	0.056	1	0.1	0.048	1
grd-shi-2784756-2816008	0.091	0.071	1	0.09	0.072	1
tor-eck-1158853-1178787	0.089	0.078	1	0.091	0.069	1

cai-elu-4788091-4813693	0.088	0.082	1	0.087	0.083	1
cai-sene-57344-101308	0.087	0.085	1	0.088	0.079	1
cai-eal-1458476-1483446	0.079	0.12	1	0.083	0.1	1
grd-cfm-2173856-2178773	0.077	0.12	1	0.081	0.11	1
cai-sega-57339-101062	0.073	0.15	1	0.069	0.17	1
cai-cama-23975-56364	0.069	0.17	1	0.07	0.16	1
cai-cyo-4142522-4169176	0.066	0.19	1	0.071	0.16	1
cai-ema-2472536-2497872	0.064	0.2	1	0.065	0.19	1
cai-eco-25826-50302	0.063	0.21	1	0.065	0.2	1
cai-sec-51598-95536	0.062	0.22	1	0.069	0.17	1
cai-eko-4058406-4082885	0.055	0.28	1	0.056	0.27	1
cai-seec-1052456-1096396	0.053	0.29	1	0.053	0.29	1
cai-ena-14772-39563	0.051	0.31	1	0.053	0.29	1
cai-sew-57348-101465	0.047	0.36	1	0.048	0.34	1
cai-ebc-294299-320969	0.045	0.37	1	0.05	0.32	1
cai-spq-51689-95628	0.039	0.44	1	0.04	0.43	1
cai-cif-1056571-1081433	0.037	0.46	1	0.041	0.42	1
cai-sed-57342-96817	0.033	0.51	1	0.034	0.5	1
cai-ebf-3985021-4010024	0.033	0.51	1	0.039	0.44	1
cai-sbg-52520-83258	0.028	0.58	1	0.031	0.53	1
yea-eco-1878874-1894731	0.027	0.59	1	0.03	0.55	1
cai-cro-32133-66456	0.021	0.68	1	0.024	0.64	1
cai-sbv-49201-79934	0.019	0.71	1	0.021	0.67	1
cai-cfar-891731-923473	0.0064	0.9	1	0.007	0.89	1
cai-sbz-49701-81116	0.0061	0.9	1	0.011	0.83	1
cai-cbra-2695495-2727410	0.0022	0.96	1	0.0038	0.94	1
cai-seo-57344-102030	-0.0039	0.94	1	-0.003	0.95	1
cai-efe-21905-59475	-0.014	0.78	1	-0.011	0.83	1
cai-kin-3809521-3836197	-0.03	0.55	1	-0.028	0.58	1
cai-cir-287785-319507	-0.056	0.27	1	-0.054	0.28	1
cai-caf-4639362-4680627	-0.062	0.22	1	-0.056	0.27	1
grd-cdrk-2091059-2111627	-0.066	0.19	1	-0.071	0.16	1

Table S4. Spearman correlations with plasma TMAO among those assigned to non-meat before red meat. Four candidate genes were quantified using KEGG SSDB gene clusters. Of 274 gene cluster homologs, 60 were detected in >70% of samples. Spearman correlations were calculated between these 60 homolog abundances and plasma TMAO. Gene clusters are named with KEGG org identifier and genomic coordinates. Universal single-copy gene normalization (USCG) utilized ten USCGs.

Gene Cluster	RPKM			USCG		
	Rho	P	P-Bonferroni	Rho	P	P-Bonferroni
cai-ebt-3409854-3439396	0.23	0.00092	0.056	0.22	0.0017	0.1
cutC-Pooled	0.22	0.0016	0.1	0.21	0.0019	0.12
cai-eih-27336-52938	0.21	0.0019	0.11	0.2	0.0037	0.23
cai-cwe-3656832-3683530	0.21	0.0029	0.18	0.2	0.0035	0.22
cai-ecl-3947888-3972396	0.2	0.0036	0.22	0.19	0.0057	0.35
cai-sel-57340-101063	0.19	0.0059	0.36	0.19	0.0071	0.43
cai-ecz-27336-52937	0.18	0.0098	0.6	0.18	0.0088	0.54
cai-sena-57098-100887	0.17	0.012	0.73	0.17	0.015	0.91
cai-ecv-27317-52919	0.16	0.019	1	0.15	0.03	1
cai-edh-3821075-3845551	0.15	0.026	1	0.14	0.04	1
cai-seeb-3884824-3928761	0.15	0.028	1	0.14	0.037	1
cai-senc-1519486-1563447	0.15	0.034	1	0.14	0.046	1
cai-senv-57098-100887	0.14	0.045	1	0.13	0.063	1
cai-ses-2844264-2873996	0.14	0.047	1	0.13	0.059	1
cai-eci-27336-52938	0.13	0.058	1	0.13	0.069	1
cai-senj-2127653-2171775	0.13	0.064	1	0.12	0.083	1
cai-sbv-49201-79934	0.13	0.068	1	0.12	0.077	1
cai-ebc-294299-320969	0.13	0.063	1	0.12	0.079	1
cai-setc-3466245-3510218	0.12	0.077	1	0.11	0.11	1
cai-elu-4788091-4813693	0.11	0.11	1	0.1	0.15	1
cai-senq-57098-100887	0.11	0.1	1	0.11	0.12	1
cai-sens-57342-100598	0.1	0.15	1	0.094	0.18	1
grd-shi-2784756-2816008	0.1	0.13	1	0.099	0.16	1
cai-cama-23975-56364	0.098	0.16	1	0.093	0.18	1
cai-sek-57350-101254	0.093	0.18	1	0.086	0.22	1
cai-eko-4058406-4082885	0.092	0.18	1	0.085	0.22	1
cai-sea-57341-101275	0.085	0.22	1	0.079	0.26	1
cai-sbz-49701-81116	0.082	0.24	1	0.079	0.26	1
cai-sene-57344-101308	0.08	0.25	1	0.072	0.3	1

grd-cfm-2173856-2178773	0.078	0.26	1	0.072	0.3	1
cai-sbg-52520-83258	0.077	0.27	1	0.07	0.32	1
cai-seo-57344-102030	0.073	0.29	1	0.063	0.37	1
cai-cyo-4142522-4169176	0.073	0.29	1	0.061	0.38	1
grd-cace-803252-808152	0.067	0.33	1	0.057	0.41	1
cai-ecoi-27459-53061	0.065	0.35	1	0.061	0.38	1
cai-seec-1052456-1096396	0.059	0.39	1	0.059	0.4	1
cai-eal-1458476-1483446	0.057	0.41	1	0.044	0.53	1
cai-seno-57098-100887	0.057	0.41	1	0.055	0.43	1
cai-seep-57183-100834	0.053	0.44	1	0.05	0.47	1
cai-senl-57098-100886	0.052	0.46	1	0.044	0.52	1
cai-cif-1056571-1081433	0.044	0.53	1	0.038	0.59	1
cai-ena-14772-39563	0.032	0.65	1	0.03	0.67	1
cai-sega-57339-101062	0.029	0.68	1	0.016	0.82	1
cai-cko-3102867-3134413	0.029	0.68	1	0.023	0.74	1
tor-eck-1158853-1178787	0.024	0.73	1	0.024	0.73	1
cai-cfar-891731-923473	0.022	0.75	1	0.015	0.83	1
cai-sec-51598-95536	0.017	0.8	1	0.017	0.8	1
cai-spq-51689-95628	0.013	0.85	1	0.0057	0.94	1
yea-eco-1878874-1894731	0.0076	0.91	1	0.007	0.92	1
cai-ema-2472536-2497872	0.00064	0.99	1	-0.0014	0.98	1
cai-sew-57348-101465	-0.0013	0.99	1	-0.0066	0.92	1
cai-cir-287785-319507	-0.0036	0.96	1	-0.011	0.88	1
cai-eco-25826-50302	-0.0045	0.95	1	-0.0075	0.91	1
cai-cro-32133-66456	-0.011	0.88	1	-0.02	0.78	1
cai-efe-21905-59475	-0.018	0.79	1	-0.022	0.76	1
cai-cbra-2695495-2727410	-0.02	0.78	1	-0.032	0.64	1
cai-ebf-3985021-4010024	-0.034	0.62	1	-0.038	0.59	1
cai-caf-4639362-4680627	-0.036	0.61	1	-0.037	0.6	1
cai-sed-57342-96817	-0.045	0.52	1	-0.056	0.42	1
grd-cdrk-2091059-2111627	-0.076	0.27	1	-0.097	0.16	1
cai-kin-3809521-3836197	-0.08	0.25	1	-0.09	0.2	1

Table S5. Spearman correlations with plasma TMAO among those assigned to red meat before non-meat. Four candidate genes were quantified using KEGG SSDB gene clusters. Of 274 gene cluster homologs, 60 were detected in >70% of samples. Spearman correlations were calculated between these 60 homolog abundances and plasma TMAO. Gene clusters are named with KEGG org identifier and genomic coordinates. Universal single-copy gene normalization (USCG) utilized ten USCGs.

Gene Cluster	RPKM			USCG		
	Rho	P	P-Bonferroni	Rho	P	P-Bonferroni
cai-cko-3102867-3134413	0.24	0.0011	0.064	0.24	0.00089	0.055
cai-sek-57350-101254	0.23	0.0016	0.1	0.23	0.0018	0.11
cai-senq-57098-100887	0.19	0.0089	0.54	0.2	0.0064	0.39
cai-senv-57098-100887	0.18	0.012	0.72	0.18	0.012	0.72
cai-seeb-3884824-3928761	0.18	0.011	0.7	0.2	0.0058	0.36
cai-seep-57183-100834	0.17	0.021	1	0.18	0.014	0.83
tor-eck-1158853-1178787	0.16	0.027	1	0.17	0.02	1
cai-ecoi-27459-53061	0.15	0.039	1	0.16	0.033	1
cai-senl-57098-100886	0.15	0.038	1	0.16	0.025	1
cai-ses-2844264-2873996	0.15	0.046	1	0.15	0.036	1
cai-eco-25826-50302	0.14	0.053	1	0.15	0.042	1
cai-sea-57341-101275	0.14	0.06	1	0.15	0.042	1
cai-senc-1519486-1563447	0.14	0.049	1	0.15	0.037	1
cai-eih-27336-52938	0.13	0.087	1	0.13	0.067	1
cai-ecv-27317-52919	0.13	0.067	1	0.14	0.053	1
cai-ema-2472536-2497872	0.13	0.085	1	0.13	0.071	1
cai-seno-57098-100887	0.13	0.076	1	0.13	0.069	1
cai-sena-57098-100887	0.13	0.083	1	0.13	0.067	1
cai-sed-57342-96817	0.13	0.082	1	0.14	0.054	1
cai-sega-57339-101062	0.12	0.11	1	0.12	0.096	1
cai-sec-51598-95536	0.12	0.12	1	0.13	0.075	1
cai-ebf-3985021-4010024	0.12	0.1	1	0.13	0.067	1
grd-cace-803252-808152	0.12	0.11	1	0.14	0.06	1
cai-eal-1458476-1483446	0.11	0.14	1	0.13	0.071	1
cai-setc-3466245-3510218	0.11	0.13	1	0.12	0.094	1
cai-cwe-3656832-3683530	0.11	0.14	1	0.13	0.082	1
cai-edh-3821075-3845551	0.1	0.17	1	0.11	0.14	1
cai-sens-57342-100598	0.1	0.16	1	0.11	0.15	1
cai-senj-2127653-2171775	0.099	0.18	1	0.11	0.12	1

cai-sel-57340-101063	0.098	0.18	1	0.1	0.16	1
cai-sene-57344-101308	0.098	0.18	1	0.12	0.11	1
cai-sew-57348-101465	0.095	0.19	1	0.1	0.15	1
cai-ecz-27336-52937	0.088	0.23	1	0.098	0.18	1
cai-ena-14772-39563	0.077	0.3	1	0.085	0.25	1
grd-cfm-2173856-2178773	0.076	0.3	1	0.091	0.22	1
cai-eci-27336-52938	0.074	0.32	1	0.08	0.28	1
grd-shi-2784756-2816008	0.074	0.31	1	0.082	0.26	1
cai-spq-51689-95628	0.067	0.36	1	0.076	0.3	1
cai-elu-4788091-4813693	0.064	0.38	1	0.072	0.33	1
cai-cro-32133-66456	0.061	0.41	1	0.079	0.29	1
cai-cyo-4142522-4169176	0.057	0.44	1	0.083	0.26	1
yea-eco-1878874-1894731	0.057	0.44	1	0.067	0.36	1
cai-seec-1052456-1096396	0.042	0.57	1	0.053	0.47	1
cai-kin-3809521-3836197	0.04	0.59	1	0.051	0.49	1
cai-cama-23975-56364	0.038	0.61	1	0.046	0.53	1
cai-ebt-3409854-3439396	0.038	0.61	1	0.054	0.46	1
cai-cbra-2695495-2727410	0.034	0.64	1	0.047	0.53	1
cai-efe-21905-59475	0.022	0.77	1	0.034	0.65	1
cai-ecl-3947888-3972396	0.02	0.79	1	0.033	0.65	1
cai-cif-1056571-1081433	0.02	0.78	1	0.038	0.61	1
cai-eko-4058406-4082885	0.011	0.88	1	0.02	0.79	1
cai-cfar-891731-923473	-0.015	0.83	1	-0.0066	0.93	1
cutC-Pooled	-0.018	0.81	1	-0.0023	0.97	1
cai-sbg-52520-83258	-0.03	0.68	1	-0.013	0.86	1
grd-cdrk-2091059-2111627	-0.041	0.58	1	-0.032	0.66	1
cai-ebc-294299-320969	-0.051	0.49	1	-0.036	0.63	1
cai-sbz-49701-81116	-0.084	0.25	1	-0.065	0.37	1
cai-seo-57344-102030	-0.09	0.22	1	-0.075	0.31	1
cai-caf-4639362-4680627	-0.091	0.22	1	-0.073	0.32	1
cai-sbv-49201-79934	-0.12	0.1	1	-0.11	0.15	1
cai-cir-287785-319507	-0.12	0.11	1	-0.1	0.16	1

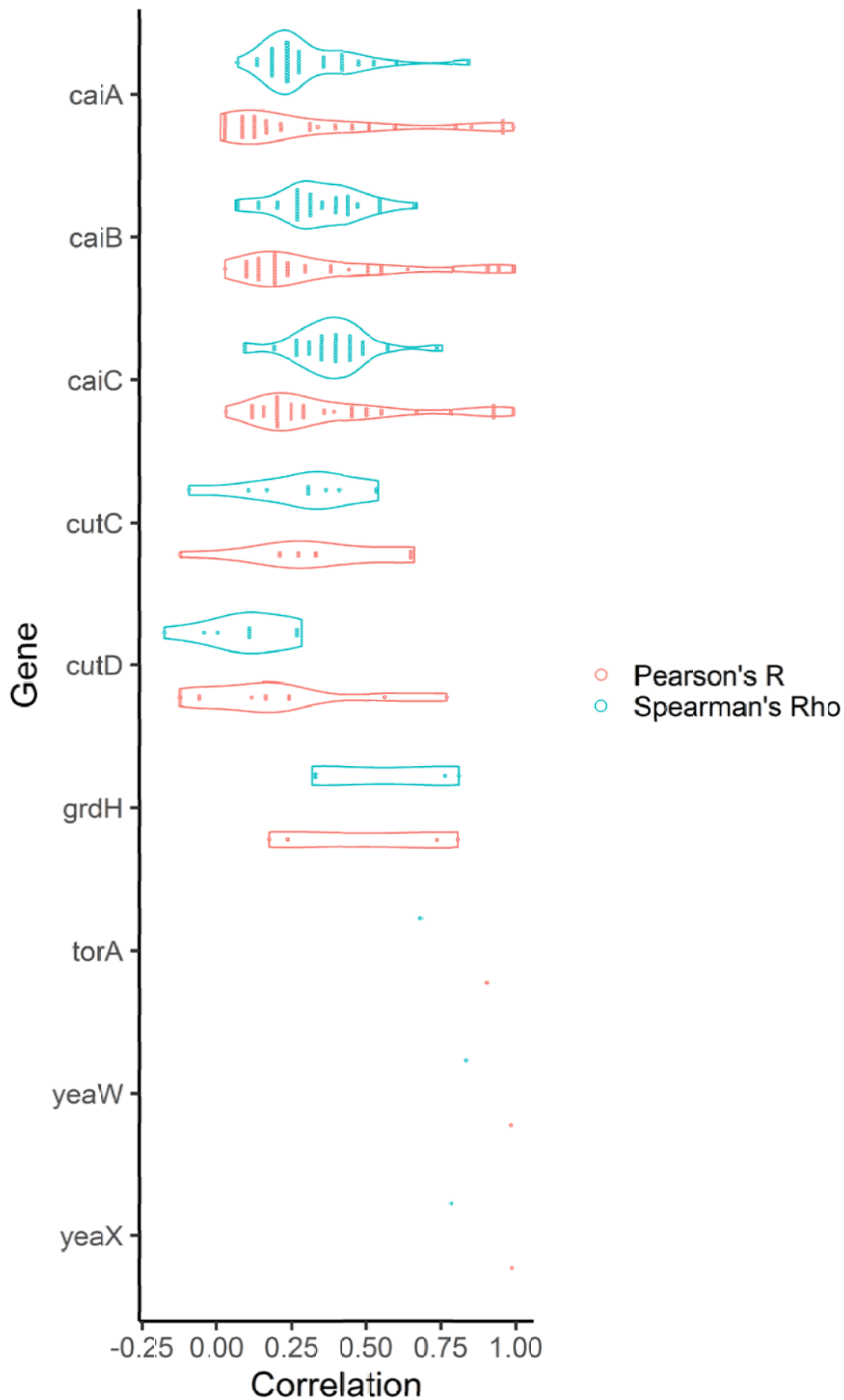


Figure S1. Gene cluster abundances are correlated with gene abundances. Genes known to be related to TMAO synthesis were quantified as RPKM according to the count of reads aligning to any gene homolog. Gene cluster homologs present in >70% of samples were quantified individually, and multiple homologs of each gene cluster were highly correlated with and detectable in more fecal samples than gene-level quantification.

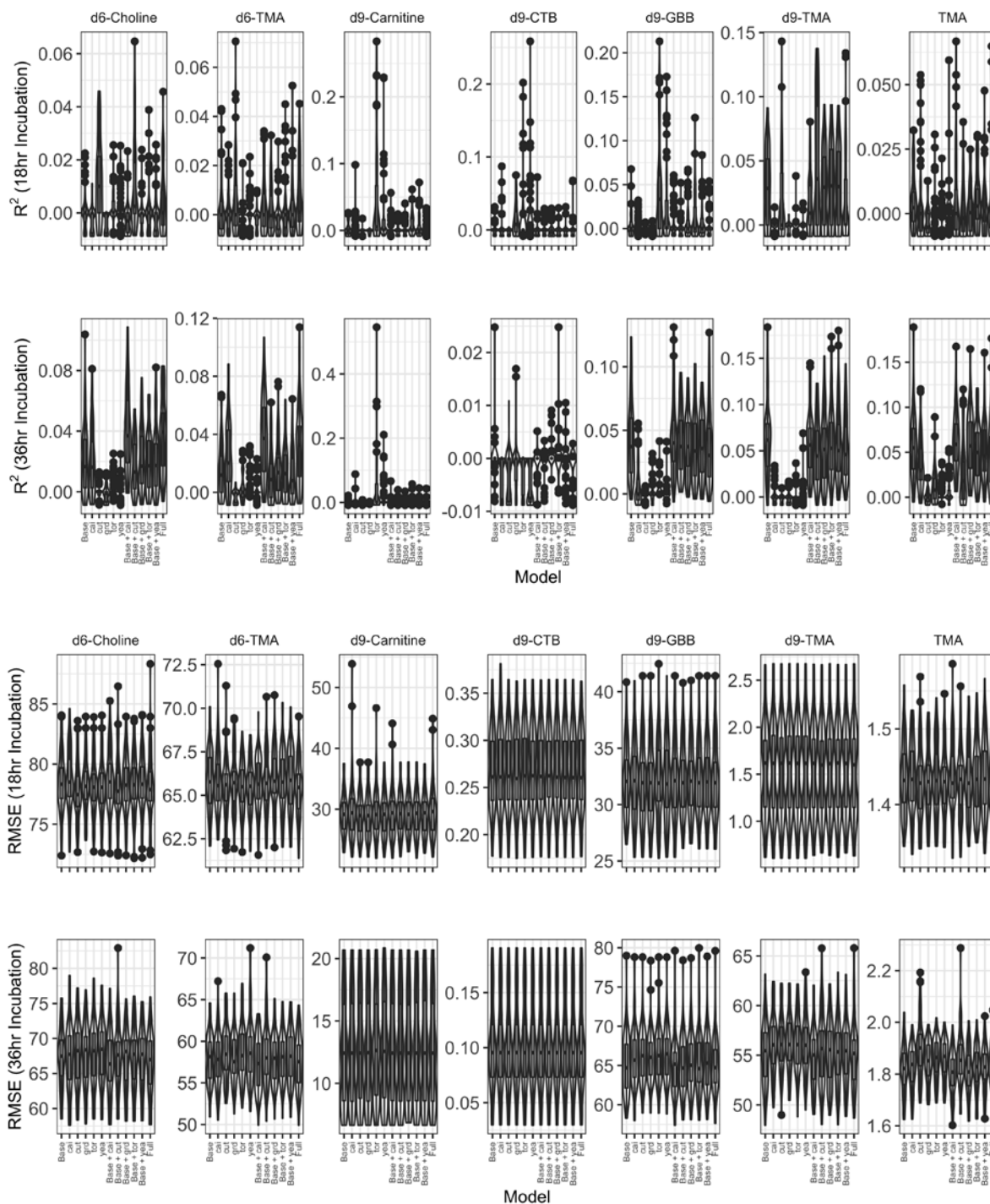


Figure S2. Prediction of fecal enzyme activities with gene abundance. Performance of elastic net linear models in prediction of choline and carnitine metabolites following incubation with labeled choline and carnitine. Root mean square error was reduced and correlation between actual and predicted values were slightly improved when predicting 18hr d6-choline with cut abundance and prediction of all other metabolites at both time points was not improved by any gene's abundance.

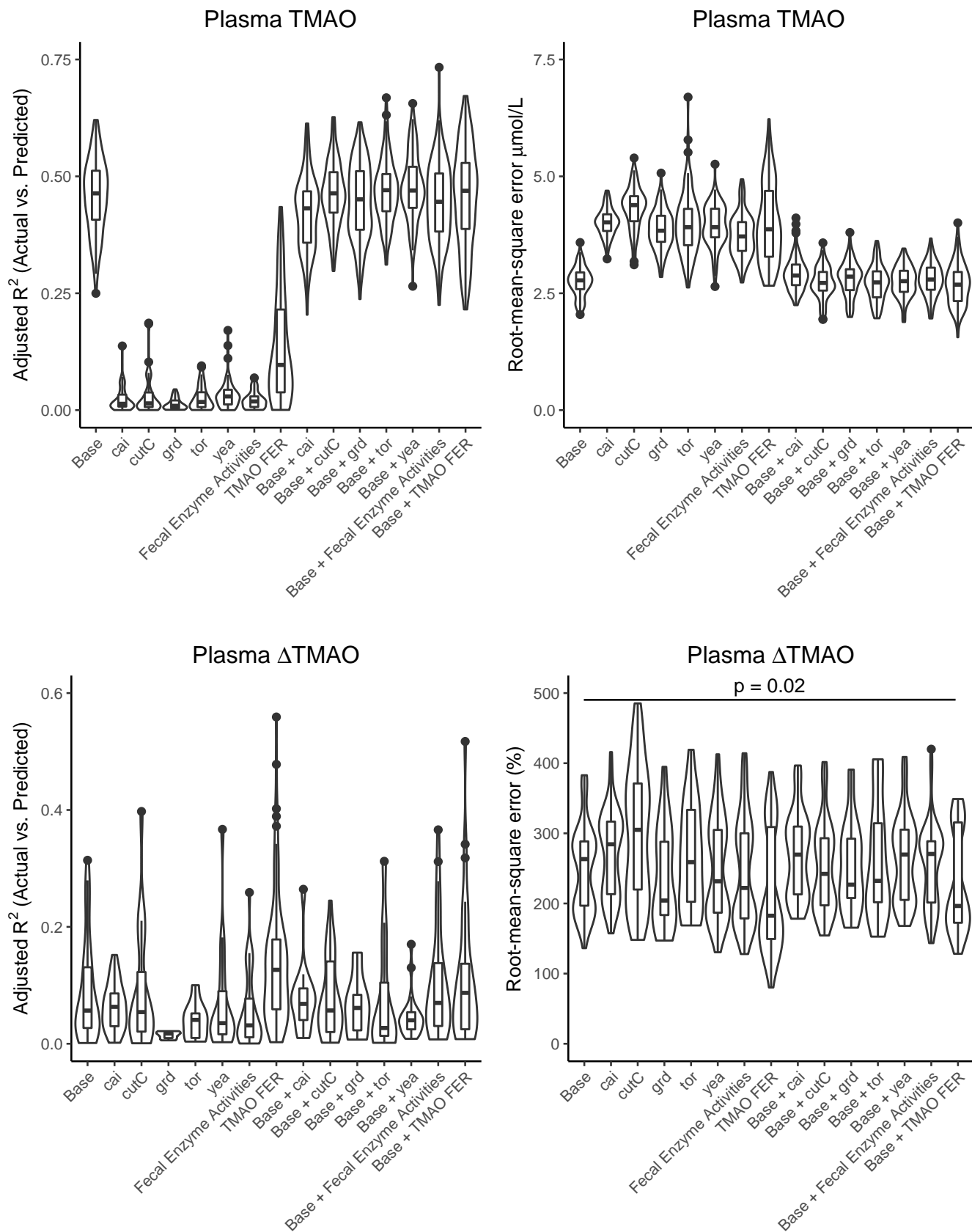


Figure S3. Abundance of genes related to TMA synthesis do not improve random forest models of TMAO. Random forest was used to train and test models of plasma TMAO, as well as percent changes in TMAO from non-meat to red meat diets (Δ TMAO). Models were tested with 100 randomly selected training/testing groups, using adjusted R^2 and root-mean-squared error (RMSE) as performance metrics. The base model, using data summarized in Table 1, predicted 46% of the variance in TMAO, and the addition of TMA-related gene abundances did not improve model performance. As a sanity check, TMAO fractional excretion rate (FER) outperformed the base model in predicting changes in plasma Δ TMAO, in terms of RMSE ($p = 0.02$, Wilcox test).

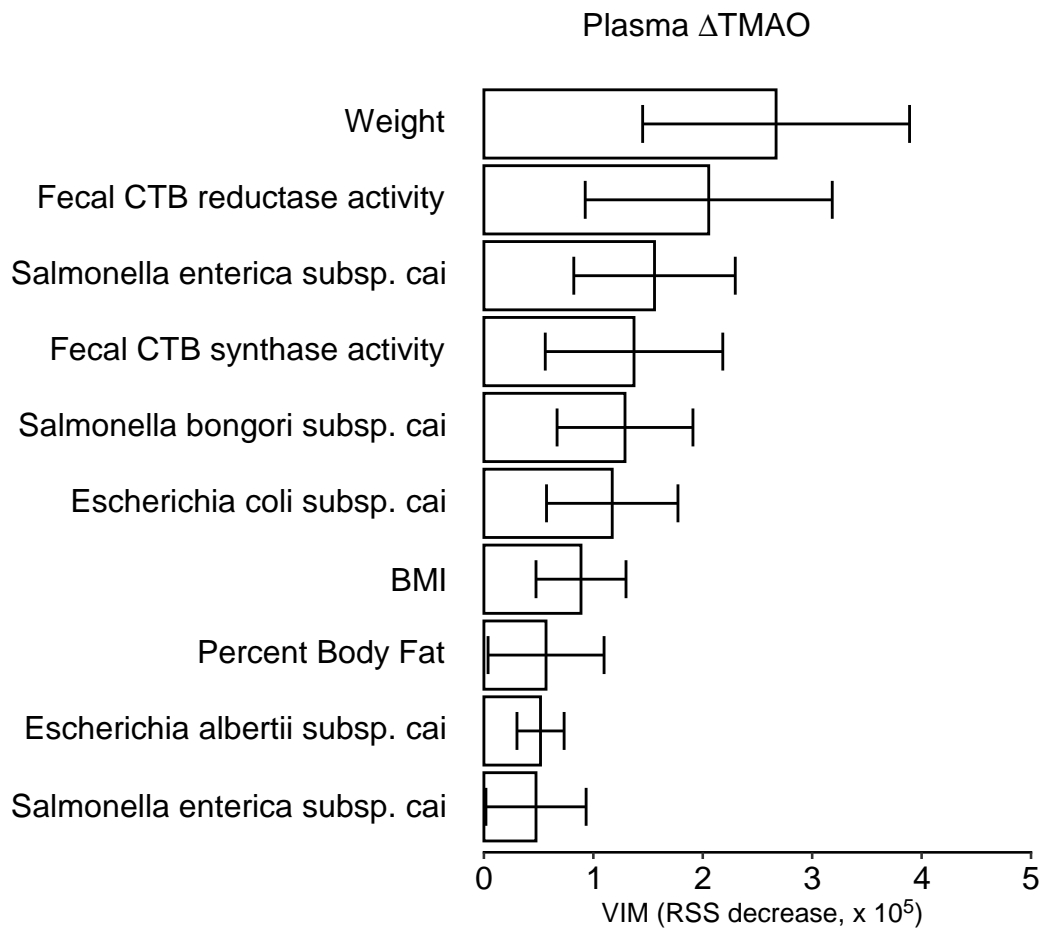
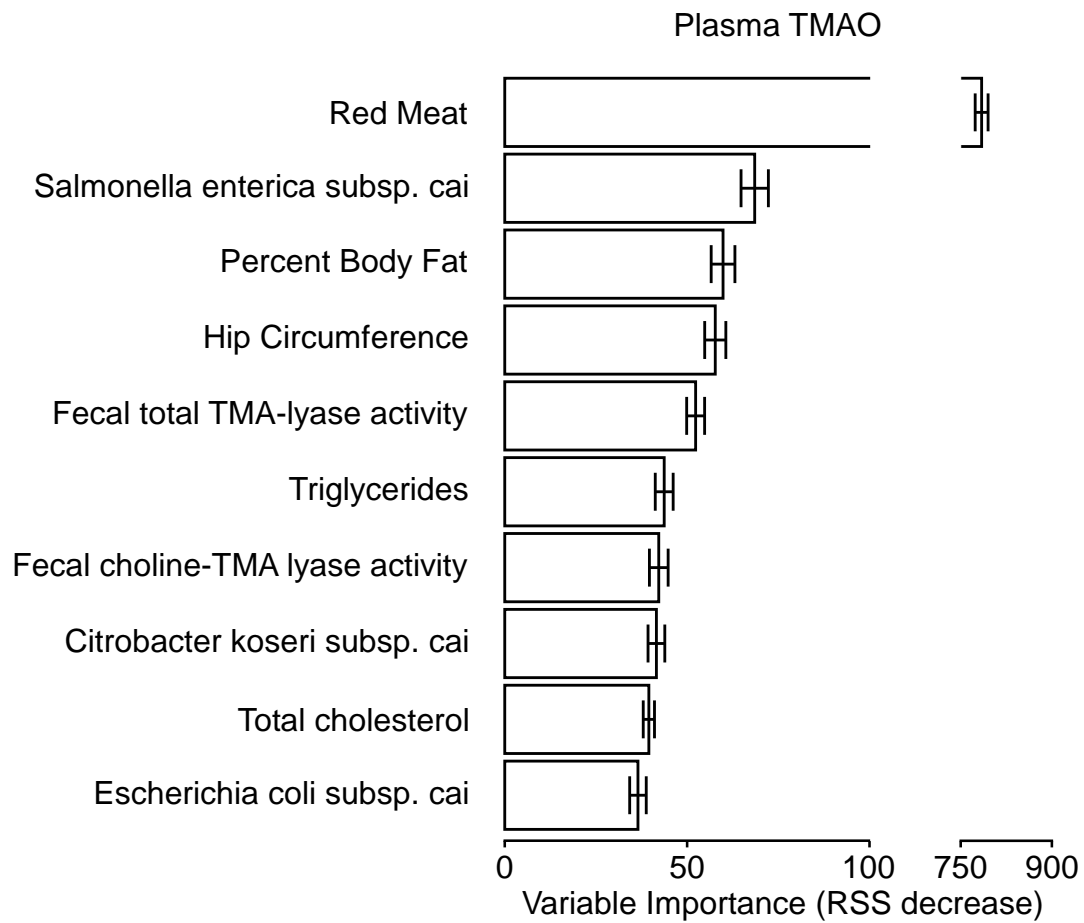


Figure S4. cai gene cluster ranks among the most important predictors of plasma TMAO. Random forest was used to train and test models of plasma TMAO, as well as percent changes in TMAO from non-meat to red meat diets (Δ TMAO). Models were tested with 100 randomly selected training/testing groups. During training, the cross-validation error (residual sum of squares, RSS) was computed after permuting each predictor. Variable importance is reported as the increase in RSS after permutation. Bars represent mean and standard error.