Correlation Network Analyses Based on Metagenomics and Multi-type Metabolomic Data Identified Biomarkers of Coronary Artery Disease

Chaoran Dong
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica
https://orcid.org/0000-0003-0424-9746

Zilong He
Institute of Microbiology Chinese Academy of Sciences

Qianhui Zhu
Institute of Microbiology Chinese Academy of Sciences

Jianxun Liu
Institute of Microbiology Chinese Academy of Sciences

Feng Gao
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica

Kexin Li
Institute of Microbiology Chinese Academy of Sciences

Shanshan Sun
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica

Qifeng Liu
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica

Yinghong Wang
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica

Yida Tang
Chinese Academy of Medical Sciences & Peking Union Medical College Fuwai Hospital

Songnian Hu
Institute of Microbiology Chinese Academy of Sciences

Haibo Zhu (zhuhaibo@imm.ac.cn)
Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Materia Medica

---

Research article

Keywords: coronary artery disease, metagenomics, metabolomics, correlation network, biomarkers

Posted Date: September 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-65514/v1

Abstract

Background: Coronary artery disease (CAD) is a complex, multifactorial disease and the underlying pathogenesis is unclear. It is essential to improve our understanding of the aetiology and pathogenesis of CAD for developing effective methods of early diagnosis and treatment.

Results: We recruited 190 participants including normal coronary artery (n = 49), stable coronary artery disease (n = 93) and acute myocardial infarction patients (n = 48). We combined metagenomics (16S rRNA sequencing and gut microbiome whole-genome sequencing) and multi-type metabolomics (serum, faeces and urine) analyses to determine the correlations among metabolites, the microbiota and clinical indexes to identify biomarkers of CAD. Compared with the faecal metabolites, serum and urine metabolites exhibited strong correlation with clinical indexes. Comparing the three types of metabolome, we discovered that the faecal and urine metabolome was more suitable than the serum metabolome for correlation analysis with the microbiota. Furthermore, we found that the serum and urine metabolome was more suitable than the faecal metabolome for correlation analysis with the clinical indexes. We constructed, for the first time, the relationship networks among the microbiota, metabolites and pathways. Through the relationship network analysis, we identified some important differential metabolites and delineated how these metabolites are related to the differentially abundant microbes.

Conclusions: In our research, we used metagenome and multi-type metabolome profiling based on hundreds of samples to figure out the correlations among metabolites, microbiota and clinical indexes. In addition, we firstly constructed the relationship networks of microbiota, metabolites and pathways. We believe our findings can help the researchers to further understand the pathogenesis of CAD.

1. Background

Coronary artery disease (CAD) is a heart disease caused by the accumulation of atherosclerotic plaques in blood vessels[1]. Based on clinical symptoms, the extent of arterial blockage and the degree of myocardial injury, CAD is divided into different categories, including stable coronary artery disease (sCAD) and acute myocardial infarction (AMI)[2]. Stable coronary artery disease (sCAD) refers to the syndrome of recurrent, transient episodes of chest pain reflecting demand-supply mismatch, that is, angina pectoris[3]. AMI is characterized by acute myocardial ischaemia and the elevated cardiac troponin values (cTn) with at least one value above the 99th percentile upper reference limit (URL)[4]. Epidemiological studies of CAD have shown that some clinical indexes such as age, male gender and smoking can increase the risk of myocardial infarction. Simultaneously, low-density lipoprotein cholesterol (LDL-C), triglyceride-rich lipoproteins, or high-density lipoprotein cholesterol (HDL-C) are associated with CAD[1]. Although Invasive coronary angiography (CA) is the gold standard for diagnosis of coronary artery disease (CAD), it is so invasive that it can cause various complications[5]. Improving our understanding of CAD aetiology and pathogenesis is essential to developing effective methods for early diagnosis and treatment.

Recent studies have shown an association between the gut microbiome and cardiovascular diseases[6]. Several metagenome-wide association studies have revealed that the composition of the human intestinal microbiome can shape human health, and that specific gut microbes are associated with many diseases[7–9]. Many studies have proven that the composition of the gut microbiome is influenced by the diet, lifestyle and genetic susceptibility of the host[10–12]. Researches show that the disordered human gut microbiome can lead to diseases, such as cardiovascular disease, metabolic syndrome and type 2 diabetes [7, 10, 13].

In terms of the contribution of the microbiome to cardiovascular diseases, several breakthroughs have been made by performing 16rRNA-seq or metagenome sequencing. For example, Li et al. performed comprehensive metagenomic and metabolomic analyses, described a novel causal role of aberrant gut microbiota in hypertension pathogenesis and also verified the results in mouse models[14]. Liu et al. identified 29 metabolite modules that were correlated with CAD phenotypes and constructed a classifier that could discriminate cases from controls[6]. Zhou et al. provided the first evidence that cardiovascular outcomes following myocardial infarction (MI) are driven by the intestinal microbiota[15].

Despite these advances, no one has compared the differential correlations between the gut microbiota/clinical indexes and three key metabolome profiles: serum, faeces and urine metabolomes. In addition, few studies have described the relationships among metabolites, the gut microbiota and pathways. To better understand the pathogenesis of CAD, we analyzed metagenome (16S rRNA sequencing and gut microbiome whole-genome sequencing) and multi-type metabolome (serum, faeces and urine) data to determine the correlations among metabolites, the microbiota and clinical indexes, and to identify biomarkers of CAD in 190 participants. Additionally, we systematically compared the results for three types of metabolome (serum, faeces and urine). We hope our findings will help the researchers to understand the pathogenesis of CAD much better.

2. Results

2.1 Clinical characteristics of CAD groups

We collected 190 participants including 49 individuals with a normal coronary artery (NCA) profile, 93 patients with stable coronary artery disease (sCAD) and 48 patients with acute myocardial infarction (AMI) from Fuwai Hospital (Beijing, China) (Table 1). In terms of patient demographics, we found that sex was significantly associated with the three groups (p <0.05), while smoking and drinking were only significantly different in NCA vs. sCAD and NCA vs. AMI (p <0.05). Only hypertension was significantly different in NCA vs. sCAD and sCAD vs. AMI. In laboratory parameters, C-reactive protein (CRP) levels was significantly different among the three groups (p <0.05), LDL cholesterol (LDL-C) and total cholesterol (TC) levels were significantly different in NCA vs. sCAD and NCA vs. AMI (p <0.05). In terms of the Gini coefficients of all clinical indexes (Figure S2), lactic acid dehydrogenase (LDH), aspartate transaminase (AST), CRP, LDL-C, TC and creatine kinase-MB (CK-MB) were the top ranked, demonstrating that these indexes have auxiliary significance for the clinical diagnosis of CAD.

Table 1
Summary of clinical indexes

	NCA (n=49)	sCAD (n=93)	AMI (n=48)	*p* value (NCA vs. sCAD)	*p* value (NCA vs. AMI)	*p* value (sCAD vs. AMI)
Demographics						
Sex (Male/Female)	19/29	71/22	46/3	< 0.001	< 0.001	0.010
Age	57.06±9.23	58.43±9.26	58.04±13.21	0.428	0.639	0.950
Smoking, case(%)	12 (25.00)	61 (65.59)	32 (65.31)	< 0.001	< 0.001	1.000
Drinking, case(%)	9 (18.75)	42 (45.16)	26 (53.06)	0.002	< 0.001	0.480
Medical history						
Hypertension, case(%)	20 (41.67)	68 (73.12)	25 (51.02)	< 0.001	0.418	0.017
Type 2 diabetes, case(%)	13 (27.08)	32 (34.41)	18 (36.73)	0.566	0.385	0.710
Hyperlipidemia, case(%)	31 (64.58)	63 (67.74)	35 (71.43)	0.710	0.518	0.706
Blood glucose	7.25±3.61	6.38±1.81	6.74±2.02	0.534	0.735	0.302
Laboratory parameters						
PN	59.34±11.54	65.49±8.18	79.58±99.06	< 0.001	< 0.001	0.828
ANC	3.85±1.51	4.72±1.67	6.27±9.95	< 0.001	< 0.001	0.537
PL	31.2±10.18	25.46±7.21	24.65±6.77	< 0.001	< 0.001	0.629
ALC	1.99±0.77	1.77±0.62	1.8±0.6	0.093	0.135	0.988
PM	6.38±1.72	6.62±2.17	7.27±2.13	0.519	0.017	0.037
AMC	0.39±0.11	0.59±0.9	0.53±0.18	0.010	< 0.001	0.053
RBC	4.45±0.51	4.57±0.82	4.53±0.5	0.036	0.350	0.320
PLT	219.46±57.39	227.44±71.95	234.59±69.06	0.848	0.444	0.517

D-dimer	0.46±0.48	0.67±1.52	0.75±1.13	0.776	0.581	0.350
FDP	2.47±0.68	2.79±2.34	3.12±1.55	0.981	0.071	0.020
BMI	24.82±3.6	25.55±2.83	26.41±3.4	0.124	0.043	0.219
HbA1c	6.43±1.59	9.12±20.62	9.51±20.84	0.618	0.699	0.299
TC	4.78±1.18	4.12±1.98	3.83±0.84	< 0.001	< 0.001	0.797
HDL	1.35±0.41	1.16±0.46	1.06±0.3	< 0.001	< 0.001	0.189
Triglycerides	1.8±1.02	1.81±1.01	1.45±0.98	0.900	0.047	0.010
LDL-C	3.09±1.07	2.38±0.91	2.2±0.77	< 0.001	< 0.001	0.459
CRP	1.24±1.11	3.79±6.05	9.32±14.61	< 0.001	< 0.001	< 0.001
Creatinine	71.85±10.1	82.14±20.51	77.88±15.52	< 0.001	0.032	0.241
WBC	7.48±8.31	7.33±2.62	7.31±2.05	0.022	0.004	0.952
ALT	24.76±14.27	31.95±31.69	36.45±21.2	0.129	< 0.001	0.013
AST	22.96±16.99	23.41±21.95	61.24±69.14	0.231	< 0.001	< 0.001
CK-MB	11.83±3.06	19.54±41.39	32.63±47.89	0.617	0.214	0.425
CK	83.17±40.51	113.72±183.55	419.63±563.23	0.927	< 0.001	< 0.001
LDH	171.58±44.5	176.94±44.49	316.63±228.12	0.573	< 0.001	< 0.001
FT3	2.92±0.27	2.92±0.3	2.62±0.55	0.766	< 0.001	< 0.001
FT4	1.09±0.18	1.15±0.32	1.08±0.42	0.676	0.018	0.003
T3	1.08±0.16	1.15±0.39	1.01±0.38	0.863	0.007	0.002
T4	7.86±1.8	7.74±1.86	7.43±1.95	0.789	0.442	0.409
TSH	4.18±12.35	2.62±2.82	1.76±1.28	0.481	0.098	0.009

The wilcox test was applied for continuous variables, and Categorical variables were compared by the Fisher's test.

2.2 Identification of differential microbiotas in the gut microbiome between the CAD subgroups

We first aimed to explore the changes in intestinal microbial community composition in the three CAD subgroups. We used QIIME2 and LEfSe to analyze the 16S rRNA sequencing data from the 190 participants[20]. The alpha diversity plot (Figure S3) showed that the sequencing data from all the samples was saturated, and the richness was diverse. We then used the cPCoA method to reveal the beta diversity of microbial composition among the CAD subgroups (Figure S4). In addition, the taxonomic classification at the phylum level of the CAD subgroups is shown in Figure S5. Here, we found that the 190 samples were mainly composed of the phyla *Bacteroidetes* and *Firmicutes*.

We focused next on the microbes with significant differential abundances among the three CAD subgroups (LDA > 2) (Figure 1). We identified 25 different taxa between the NCA and sCAD groups (Figure 1A). For example, *Bacteroides* and *Escherichia* were found at a higher abundance in the NCA group (LDA > 3), while *Desulfovibrio* and *Parabacteroides* were found at a higher abundance in the sCAD group (LDA > 2). We also identified 34 significantly different taxa between the NCA and AMI groups (Figure 1B). Here, *Desulfovibrio*, *Streptococcus*, and *Lacobacillus* had higher abundances in the AMI group (LDA > 2), while *Clostridium* had a higher abundance in the NCA group (LDA > 3). Finally, we found 69 significantly different taxa between the sCAD and AMI groups (Figure 1C). Notably, *Streptococcus*, *Alistipes*, *Olsenella*, *Actinomyces*, and *Escherichia* had higher abundances in the AMI group (LDA > 2), and *Prevotella*, *Clostridium*, and *Lactobacillus* had higher abundances in the sCAD group (LDA > 3). Together, we have identified taxa with significantly different abundances in different CAD subgroups.

2.3 Identification of differential metabolites in multi-type metabolomes between the CAD subgroups

We next used NMR (Nuclear Magnetic Resonance) to detect metabolites in the three types of metabolome (faeces, serum and urine) in samples from NCA, sCAD and AMI participants. We detected total of 29, 33 and 31 metabolites in the faeses, serum and urine, respectively. In NCA vs. sCAD, we identified 3, 18 and 18 metabolites that were differential detectable in the faeces, serum and urine, respectively (Table 2) (p <0.1). The methanol level was also significantly different in the three types of metabolome, and the levels of other four metabolites, including creatinine, were significantly different in the serum and urine metabolome. In sCAD vs. AMI, we detected 7, 21 and 18 differentially detectable metabolites in the faeces, serum and urine, respectively (Table 2) (p < 0.1). Similar to NCA vs. sCAD, the levels of five metabolites, including creatinine, were significantly different in the serum and urine metabolomes. Finally, in NCA vs. AMI, we detected 6, 18 and 14 differentially detectable metabolites in the faeces, serum and urine, respectively (Table 2) (p <0.1). Again, the levels of creatinine and alanine were significantly different in the serum and urine metabolome. Overall, we found that citrate and creatinine levels in the serum and

urine metabolomes were significantly different among the CAD subgroups and thus, we infer that citrate and creatinine are associated with the pathogenesis of CAD.

2.4 Integrated analyses reveal the relationships between the gut microbiota and differential faecal metabolites in the CAD subgroups

To identify the relationships among the microbiota, metabolites and metabolic pathways, we performed an integrated analysis based on 16S rRNA gene, metagenomic, and metabolomics (faecal) sequencing data. 20 samples of each CAD subgroups were randomly selected for metagenomic sequencing. We identified the relationships between metabolites and metabolic pathways in the MetaCyc database[25] and obtained the relative abundances of taxa and metabolic pathways by analyzing metagenomic data. Then, we constructed relationship networks based on the metabolic pathways, differential metabolites and differentially abundant genera.

In the NCA vs. sCAD network (Figure 2), the differential abundant metabolite alanine and differentially abundant genus *Escherichia*/*Streptococcus* were linked through the metabolic pathways: L-alanine biosynthesis. In the sCAD vs. AMI network, we observed 6 differential metabolites, 13 metabolic pathways, and 12 differentially abundant genera. According to the previous studies, methionine and acetate are associated with the development of CAD [27-30]. And in our research, methionine and acetate were connected to differentially abundant genera, such as *Escherichia*, *Streptococcus* and *Coprococcus* through the L-methionine cycle I, pyruvate fermentation to acetate and lactate II metabolic pathways. Notably, we found that *Escherichia* was present in both the NCA vs. sCAD and sCAD vs. AMI networks. Finally, the NCA vs. AMI network contained 3 differential metabolites, 11 metabolic pathways and 6 genera. Interestingly, methionine and acetate were present in both the sCAD vs. AMI and NCA vs. AMI networks. With the exception of *Desulfovibrio*, the five other differentially abundant genera also existed in the sCAD vs. AMI network. These metabolites and genera might, therefore be associated with AMI occurrence.

Table 2A
Differential metabolites in fecal samples among NCA, sCAD and AMI

Feces Metabolites	Fold Change	*p* Value
sCAD vs NCA		
Hypoxanthine	1.198	0.0307
Methanol	0.654	0.0887
Alanine	1.126	0.0963
AMI vs NCA		
1,3-Dihydroxyacetone	0.841	0.0685
Acetate	0.852	0.0835
Glucose	0.794	0.0980
Malonate	0.811	0.0506
Methionine	1.183	0.0473
Phenylacetate	1.315	0.0153
AMI vs sCAD		
1,3-Dihydroxyacetone	0.884	0.0686
Acetate	0.801	0.0035
Alanine	0.872	0.0702
Citrulline	0.875	0.0538
Methionine	1.165	0.0286
Phenylacetate	1.323	0.0022
Tyrosine	0.856	0.0605

Table 2B
Differential metabolites in urine samples among NCA, sCAD and AMI

Urine Metabolites	Fold Change	*p* Value
sCAD vs NCA		
2-Hydroxyisobutyrate	1.143	0.0268
3-Hydroxyisovalerate	0.858	0.0203
Betaine	1.548	0.0265
Citrate	0.779	0.0049
Creatinine	1.275	0.0001
Galactose	0.004	<0.0001
Glycine	1.131	0.0419
Glycolate	1.128	0.0459
Isoleucine	0.559	<0.0001
Methanol	0.821	0.0073
Methylsuccinate	0.599	0.0071
N-Acetylglutamate	0.811	0.0044
Phenylalanine	1.428	0.0131
Tartrate	0.627	<0.0001
Taurine	1.210	0.0863
Threonine	1.125	0.0634
Tyrosine	0.785	0.0164
Valine	0.397	<0.0001
AMI vs NCA		
2-Oxoglutarate	0.650	0.0002
3-Hydroxyisovalerate	0.836	0.0221
Acetate	4.041	0.0016
Alanine	0.840	0.0209
Choline	0.820	0.0167
Citrate	0.534	<0.0001

Creatinine	0.805	0.0107
Glycolate	0.777	0.0007
Hippurate	0.487	0.0004
Isoleucine	0.796	0.0364
Methanol	0.791	0.0190
Pyruvate	0.825	0.0170
Taurine	1.413	0.0078
Trigonelline	0.178	0.0542
AMI vs sCAD		
2-Hydroxyisobutyrate	0.829	0.0113
Acetate	4.041	<0.0001
Alanine	0.766	0.0005
Betaine	0.491	0.0013
Choline	0.746	0.0004
Citrate	0.686	0.0007
Creatinine	0.631	<0.0001
Formate	1.769	0.0137
Galactose	187.176	<0.0001
Glycolate	0.689	<0.0001
Hippurate	0.503	0.0009
Hypoxanthine	0.797	0.0855
Isoleucine	1.425	0.0043
N-Acetylglutamate	1.148	0.0950
Tartrate	1.443	0.0072
Threonine	0.831	0.0260
Tyrosine	1.834	0.0091
Valine	2.189	<0.0001

Table 2C
Differential metabolites in serum samples among NCA, sCAD and AMI

Serum Metabolites	Fold Change	*p* Value
sCAD vs NCA		
Tyrosine	0.915	0.0176
Proline	1.104	0.0087
Methionine	1.089	0.0124
Methanol	0.630	<0.0001
Isobutyrate	1.186	<0.0001
Glycine	0.962	0.0667
Glycerol	1.825	<0.0001
Formate	1.848	0.0026
Creatinine	1.210	<0.0001
Citrate	0.933	0.0261
Alanine	0.917	0.0113
3-Hydroxybutyrate	1.355	0.0577
2-Hydroxybutyrate	1.208	<0.0001
VLDL/LDL	0.909	0.0008
Ptdcho	0.890	0.0007
N-Ac	0.949	0.0002
lipid-CH_2CO	0.950	0.0020
lipid-CH_2CH_2CO	0.925	<0.0001
AMI vs NCA		
Propylene glycol	0.844	<0.0001
Phenylalanine	1.093	0.0374
Lactate	1.263	<0.0001
Isoleucine	0.875	0.0037
Isobutyrate	0.902	0.0867

Glycine	0.900	<0.0001
Glycerol	1.479	0.0420
Formate	1.648	0.0840
Dimethyl sulfone	0.818	0.0041
Creatinine	1.115	0.0341
Citrate	0.872	0.0032
Betaine	0.848	0.0026
Alanine	0.852	0.0007
Acetone	1.403	0.0166
3-Hydroxybutyrate	1.811	0.0062
2-Hydroxybutyrate	1.148	0.0381
lipid-CH=CH	0.929	0.0513
VLDL/LDL	0.914	0.0165
AMI vs sCAD		
Propylene glycol	0.875	0.0001
Proline	0.865	0.0004
Phenylalanine	1.107	0.0013
Methionine	0.902	0.0019
Methanol	1.299	0.0230
Lactate	1.289	<0.0001
Isoleucine	0.848	<0.0001
Isobutyrate	0.761	<0.0001
Glycine	0.935	0.0024
Glycerol	0.810	0.0757
Dimethyl sulfone	0.864	0.0141
Creatinine	0.921	0.0810
Citrate	0.934	0.0578
Betaine	0.820	<0.0001
Alanine	0.929	0.0563

3-Hydroxybutyrate	1.336	0.0860
Ptdcho	1.060	0.0844
N-Ac	1.062	<0.0001
lipid-CH2CO	1.032	0.0899
lipid-CH2CH2CO	1.064	<0.0001
lipid-CH=CH	0.935	<0.0001

2.5 Correlation analysis between the gut microbiota and CAD clinical indexes

To uncover the relationships between the gut microbiota and clinical indexes, we performed a spearman correlation analysis of the relative abundance of the gut microbiota with the clinical indexes of all samples (Figure 3). As the figure shows, a lots of significant correlations were found, for example, AST was significantly positively correlated with *S.agalactiae*, *Streptococcus*, and *Streptococcaceae* (rho > 0.25 and p <0.05); FDP was significantly negatively correlated with *Prevotella* and *Prevotellaceae* (rho < -0.25 and p <0.05); LDL-C was significantly negatively correlated with *Desulfovibrio* (rho < -0.25 and P < 0.05); type 2 diabetes was significantly positively correlated with *Lactobacillus* and *L. hamsteri* (rho > 0.25 and p <0.05), and significantly negatively correlated with *Erysipelotrichaceae* (rho < -0.25 and p <0.05); and HbA1c was significantly positively correlated with *L. hamsteri* (rho > 0.25 and p <0.05). Twelve taxa such as Streptococcus, Prevotella and Bacteroides were significantly different in NCA vs. sCAD, sCAD vs. AMI and NCA vs. AMI. In addition, according to the Gini coefficients of clinical indexes, we found that AST and LDL-C were highly significant for CAD grouping (Figure S2). Together, these data show that some differentially abundant taxa are significantly related to some key clinical indexes such as LDL-C, AST and HbA1c, which provided a good theoretical basis for these differential taxa as biomarkers in future, and pending replication in later researches.

2.6 Correlation analysis between clinical indexes and three types of metabolites

To explore and compare the relationships between clinical indexes and the metabolites in serum, urine and faeces, we again performed a spearman correlation analysis between the three types of metabolites and clinical indexes (Figure 4). In the analysis of the correlation between faecal metabolites and clinical indexes, we found that FT3 was significantly positively correlated with hypoxanthine (rho > 0.25 and p <0.05). In the correlation analysis of serum metabolites and clinical indexes, CRP was significantly negatively correlated with alanine and citrate (rho < -0.25 and p <0.05); AST was significantly positively correlated with lactate (rho > 0.25 and p <0.05), and negatively correlated with glycine (rho < -0.25 and p

<0.05). Type 2 diabetes, blood glucose and HbA1c was significantly negatively correlated with most metabolites (rho < -0.25 and p <0.05) (Figure 4A).

In the correlation analysis between urine metabolites and clinical indexes, CRP was significantly positively correlated with taurine (rho > 0.25 and p <0.05), and significantly negatively correlated with citrate, hippurate. (rho < -0.25 and p <0.05); AST was significantly positively correlated with metabolites such as urea and galactose (rho > 0.25 and p <0.05); FT3 was significantly negatively correlated with taurine (rho < -0.25 and p <0.05) and significantly positively correlated with creatinine, hippurate, glycolate and choline (rho > 0.25 and p <0.05) (Figure 4B).

When comparing the results of the correlation analysis between the three types of metabolites and clinical indexes, we found that more metabolites in serum and urine were correlated with more clinical indexes than the faecal metabolome, this may indicate the serum and urine metabolome were more closely linked to the pathogenesis of CAD. It is worth noting that both type 2 diabetes vs. methanol and blood glucose vs. methanol showed significant negative correlation not only in the correlation analysis between serum metabolome and clinical indexes, but also in the correlation analysis between urine metabolome and clinical indexes. This may imply methanol may be correlated with both CAD and type 2 diabetes.

2.7 Correlation analysis between gut microbiota and multi-type metabolites

To reveal and compare the relationships between the gut microbiota and the three CAD subgroups of faecal, serum and urine metabolites. In our spearman correlation analysis (Figure 5), we found that in the correlation analysis between the faecal metabolome and gut microbiota (Figure 5A), *Desulfovibrio* was significantly positively correlated with sarcosine, dimethylamine, isovalerate (rho > 0.3 and p <0.05); Consistent with a previous study[31], *Faecalibacterium* were significantly positively related to butyrate (rho > 0.3 and p <0.05). We also found that *Prevotella* and *Prevotellaceae* were significantly positively correlated with hypoxanthine and xanthine (rho > 0.3 and p <0.05); *A.finegoldii*, *Alistipes* and *Rikenellaceae* were significantly positively correlated with p-cresol (rho > 0.3 and p <0.05), which was consistent with the previous report[31].

When analyzing the correlation between the serum metabolome and gut microbiota (Figure 5B), *Prevotella* and *Prevotellaceae* showed significantly positive correlations with isoleucine and propylene-glycol (rho > 0.3 and p <0.05); *Lactobacillus* and *Lactobacillaceae* showed significantly positive correlations with mannose (rho > 0.3 and p <0.05); Lactobacillaceae was significantly negatively correlated with alanine (rho < -0.3 and p <0.05); *C.aldenense* was significantly positively correlated with lactate (rho > 0.3 and p <0.05).

In the analysis of the correlation between urine metabolome and gut microbiota (Figure 5C), *Prevotella* and *Prevotellaceae* were significantly positively related to hippurate (rho > 0.3 and p <0.05); *M.fermentans*

and *S.termitidis* were significantly positively related to hypoxanthine and N-phenylacetylglycine, respectively (rho > 0.3 and p <0.05); *S.variabile* was significantly positively correlated with galactose and urea (rho > 0.3 and p < 0.05).

Finally, when analyzing the correlation between the three types of metabolome and the gut microbiota, we found that 70% of the taxa in which correlated with the urine metabolites (rho > 0.3 and p <0.05) were also correlated with the faecal metabolites. With the exception of *Prevotella* and *Prevotellaceae*, the gut microbes associated with the serum metabolome were different from those associated with the other two metabolomes. In addition, *Prevotella* and *Prevotellaceae* which were related to the three metabolomes were differentially abundant in sCAD vs. AMI.

3. Discussion

The study has explored the relationships among the gut microbiota, three types of metabolome and clinical indexes in an attempt to find potential CAD biomarkers; the biological roles of our identified markers (such as the genus *Streptococcus* and *Prevotella*, the metabolite citrate and creatinine) now warrant further investigation. In support of these findings, Liu et al. found that *Prevotella* was a potential therapeutic target for patients with cardiovascular disease[32]. It has been proved that the citrate cycle enhances cardioprotection[33]. And creatinine has been proposed to be a suitable biomarker for diabetes, CAD and renal function[34]. Furthermore, we constructed relationship networks for the microbiota, metabolic pathways, and metabolites to try to explore CAD pathogenesis. We used the faecal metabolites to build the network because the metabolites detected in the faeces are more likely to come from intestinal microorganisms. Through network analysis, we delineated how these metabolites are related to the differentially abundant microbes. For example, we speculated that *Escherichia* releases alanine through the L-alanine biosynthesis pathway.

In the correlation analysis, serum/urine metabolites correlated more with clinical indexes than faecal; this finding constitutes a good option for researchers who want to further study the relationships between metabolites and clinical indexes. Compared with the serum metabolome, more correlations have been found between the faecal/urine metabolome and gut microbiota. By comparing the performance of the three metabolomes, we found that faeces/urine metabolites have stronger correlations with gut microbiota than serum metabolites.

190 faecal samples were sequenced by 16S rRNA sequencing, 60 faecal samples of which were also sequenced by metagenome sequencing. Compared with metagenomics sequencing, 16S rRNA sequencing is not expensive and suitable for large cohort analysis. But it is difficult to detect the abundance of microbiotas in species level of 16S data for technical reason. A similar situation occurs in the metabolome using NMR methods. Although metabolome sequencing based on NMR is highly specific, fewer metabolites can be detected. With the development of next generation sequencing (NGS) sequencing technology, metagenomic shotgun sequencing will become the mainstream, which will

promote the further study of CAD in the future. Simultaneously, a stable untargeted metabolomics sequencing technology with high specificity, sensitivity and resolution need to be developed.

4. Conclusions

In our research, we used metagenome and multi-type metabolome profiling based on hundreds of samples to figure out the correlations among metabolites, microbiota and clinical indexes. We found serum and urine metabolites have stronger correlations with clinical indexes than faecal metabolites. Faecal and urine metabolites have stronger correlations with gut microbiota than serum metabolites. In addition, we firstly constructed the relationship networks of microbiota, metabolites and pathways. We believe our findings can help the researchers to further understand the pathogenesis of CAD.

5. Methods

5.1 Participants and study design

This study was approved by the Ethics Committee of Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (Beijing, China) in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. Written informed consent was obtained from all participants.

This is a single-center cross sectional study. We continuously recruited 190 participants under 65 years old with complete information on medical history, clinical and biochemical parameters from Fuwai Hospital, National Center for Cardiovascular Diseases of China. The diagnosis was made on the basis of symptoms, laboratory tests, ECG and coronary angiographic results. 190 participants including normal coronary artery (NCA group, n = 49), stable coronary artery disease with the coronary artery stenosis ≥ 70% (sCAD group, n = 93) and acute myocardial infarction (AMI group, n = 48) were recruited between December 2016 and February 2017. The criteria for AMI included: 1) symptoms of chest pain at rest (> 20 min); 2) ischemic electrocardiographic changes: ST-segment changes and/or T-wave inversions; 3) significant increases in myocardial enzyme levels. For sCAD, the criteria included: 1) chest pain symptoms (< 10 min) and electrocardiographic changes only after activity; 2) normal myocardial enzyme level. The coronary angiography was performed on all patients. Plaques or stenosis was not found in age- and sex-matched control subjects. All enrolled participants in the NCA, sCAD and AMI group who were suspected of CAD underwent CAG and had no history of unstable angina, myocardial infarction, stroke, cancers, or coronary revascularization. The angiographic data were confirmed independently by two observers in this study.

5.2 Nuclear Magnetic Resonance (NMR) Sample collection and preparation

Serum before the coronary angiography surgery and urine early morning urinary samples were collected and centrifuged at 278 K at 3,000g for 10 min, the supernatants of samples were stored at -80 °C for metabolic profile establishment and statistical analysis. Faeces samples were stored at -80 °C after homogenate with phosphate buffer (0.2 M NaH2PO4/K2HPO4, pH 7.4). Samples were prepared using the previously reported method[16].

5.3 NMR Spectra Acquisition and Processing

All NMR spectra were recorded at 298 K using a Bruker Avance 500 MHz spectrometer (1H frequency: 500.13 MHz; Bruker, Germany). For quantitative metabolomics profiling of filtered serum, urine, and faeces, spectra were processed with the Chenomx NMR Suite 7.5 software (Chenomx Inc., Edmonton, Canada) using the “targeted profiling” approach[17]. Open database sources, including the KEGG, MetaboAnalyst, Human Metabolome Database, and METLIN, were used to identify metabolic pathways[18, 19].

5.4 NMR Multivariate Data Analysis

Output data were processed with the SIMCA-P+ 14.0 software (Umetrics, Sweden) to elucidate patterns in metabolite concentration shifts. Statistical analysis was also conducted with SPSS19.0 (IBM; USA) using the two-tailed Student’s t-test. P-value of less than 0.05 was considered to be statistically significant between two groups.

5.5 Human faecal sample collection and DNA extraction

Fresh faeces samples were collected from 190 subjects, and then delivered from Fuwai Hospital to the laboratory in an ice bag using insulating polystyrene foam containers. DNA was extracted using an EZNA™ stool DNA isolation kit (Omega Bio-Tek, VWR, Herlev, Denmark). The DNA was then eluted in 50 μL of elution buffer and stored at -80°C.

5.6 DNA library construction and sequencing

DNA library was constructed using the TruSeq Nano DNA LT Library Preparation Kit (FC-121-4001, Illumina, San Diego, CA, USA). The resulting libraries were sequenced on an Illumina HiSeq 4000 sequencer (Illumina, San Diego, CA, USA). The running mode of metagenomics was paired-end of 150 bp and the running mode of 16S rRNA sequencing was paired-end of 300 bp.

5.7 Sequencing data analysis

QIIME 2 was used to process 16S rRNA sequencing data (Figure S1). Sequence quality control, feature table construction and filter chimeric sequences were performed by DADA2 plugins[20]. Features were created by clustering sequences with 100% similarity. Representative sequences for each feature were used to construct a rooted phylogenetic tree by q2-phylogeny plugin. The script will randomly subsample the counts from each sample to the 7327 sequences. Alpha and beta diversity were generated by q2-diversity plugin. Shannon's index, the observed OTUs, and evenness were evaluated. A normalized feature abundance table was used for the constrained principal coordinate analysis (cPCoA). Taxonomic analysis was performed by q2-feature-classifier plugin. Differential abundance taxa were generated by LEfSe (LDA>2)[21].

Quality control for the metagenomics shotgun sequencing data was conducted using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Low quality reads and adapter sequences were removed by Trimmomatic[22]. Taxonomic profiles were generated using MetaPhlAn v2.662[23] and pathways enrichment was done by HUMAnN2[24]. The interactions of pathways and metabolites were integrated using MetaCyc and the relationship networks contained microbiota, metabolites and pathways were visualized by Cytoscape[25, 26].

5.8 Statistical analyses

The Gini coefficients of clinical indexes were generated using R scripts. The wilcox and Fisher's test were used to analyze the differential clinical indexes for continuous and categorical variables, respectively. Spearman correlations between microbiota, metabolites and clinical indexes were calculated using R scripts. The visual presentation of multiple omics correlations was performed using the pheatmap package in R.

5.9 Data availability

The datasets are available in the repository of the Genome Sequence Archive Sequence Database of National Genomics Data Center (https://bigd.big.ac.cn/) under the accession number CRA002142.

List Of Abbreviations

CAD	coronary artery disease
16S rRNA-seq	16S rRNA sequencing
STEMI	ST-segment elevation myocardial infarction
post-MI	post myocardial infarction
WGS	whole genome sequencing
NCA	normal coronary artery
sCAD	stable coronary artery disease
AMI	acute myocardial infarction
NMR	nuclear magnetic resonance
DSS	dimethyl-silapentane-sulfonate
KEGG	kyoto encyclopedia of genes and genomes
PCA	principal component analysis
PLS-DA	partial least squares discriminant analysis
PCR	polymerase chain reaction
rRNA	ribosomal RNA
OTU	operational taxonomic unit
cPCoA	constrained principal coordinate analysis
ECG	electrocardiograms
PN	percentage of neutrophils
ANC	absolute neutrophil count
PL	percentage of lymphocyte
ALC	absolute lymphocyte count
PM	percentage of monocyte
AMC	absolute monocyte count
RBC	red blood cell
PLT	platelet
FDP	fibrin degradation products
BMI	body mass index
TC	total cholesterol

HDL	high density lipoprotein
LDL-C	low density lipoprotein cholesterol
CRP	C-reactive protein
WBC	white blood cell
ALT	alanine aminotransferase
AST	aspartate aminotransferase
CK-MB	creatine kinase myocardic isoenzyme
CK	creatine kinase
LDH	lactate dehydrogenase
TSH	thyroid-stimulating hormone
cTn	cardiac troponin
URL	upper reference limit

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (Beijing, China) in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. Written informed consent was obtained from all participants.

Consent for publication

The authors consent for publication

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the repository of the Genome Sequence Archive Sequence Database of National Genomics Data Center (https://bigd.big.ac.cn/) under the accession number CRA002142.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences [CIFMS 2016-I2M-1–009], Disciplines Construction Project [201920200807], Fundamental Research Funds for the Central Universities [3332019068], The Drug Innovation Major Project [2018ZX09711001-003-011], National Program on Key Basic Research Project of China [973 Program, No. 2015CB554405].

Authors' contributions

HZ, SH and YT conceived the research and designed the experiments. YT recruited patients, collected clinical samples and registered clinical information. YW, CD, FG, SS and QL performed the experiments. CD, ZH, QZ and KL analyzed the data. HZ, CD and ZH wrote the manuscript. JL revised the manuscript. HZ, JL and YT offered funds for the research.

Acknowledgements

We thank all study participants, research staffs, and students who participated in this work.

References

1. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. Nature reviews Genetics. 2017;18(6):331–44.
2. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corra U, Cosyns B, Deaton C, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. Revista espanola de cardiologia. 2016;69(10):939.
3. Ford TJ, Corcoran D, Berry C. Stable coronary syndromes: pathophysiology, diagnostic advances and therapeutic need. Heart. 2018;104(4):284–92.
4. Alpert JS. The Fourth Edition of the Universal Definition of Myocardial Infarction. Am J Med. 2018;131(11):1265–6.
5. Gorenoi V, Schonermark MP, Hagen A. CT coronary angiography vs. invasive coronary angiography in CHD. GMS Health Technol Assess. 2012;8:Doc02.
6. Liu H, Chen X, Hu X, Niu H, Tian R, Wang H, Pang H, Jiang L, Qiu B, Chen X, et al: **Alterations in the gut microbiome and metabolism with coronary artery disease severity**. *Microbiome* 2019, 7(1).
7. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55–60.
8. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for

colorectal cancer. Gut. 2017;66(1):70–8.

9. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, Wu X, Li J, Tang L, Li Y, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nature medicine. 2015;21(8):895–905.
10. Ussar S, Griffin NW, Bezy O, Fujisaka S, Vienberg S, Softic S, Deng L, Bry L, Gordon JI, Kahn CR. Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome. Cell Metabol. 2015;22(3):516–30.
11. Zhong H, Penders J, Shi Z, Ren H, Cai K, Fang C, Ding Q, Thijs C, Blaak EE, Stehouwer CDA, et al. Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. Microbiome. 2019;7(1):2.
12. Aaron C, Ericsson SA, Marina M, Hanson SB, Busi TW, Parker RJ, Schehr MA, Hankins CE, Ahner, Justin W, Davis. Craig L Franklin, James M Amos-Landgraf, Elizabeth C Bryda: **Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota**. *Oncotarget* 2015, 6(32).
13. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun. 2017;8(1):845.
14. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, Wu S, Liu W, Cui Q, Geng B, et al. Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome. 2017;5(1):14.
15. Zhou X, Li J, Guo J, Geng B, Ji W, Zhao Q, Li J, Liu X, Liu J, Guo Z, et al. Gut-dependent microbial translocation induces inflammation and cardiovascular events after ST-elevation myocardial infarction. Microbiome. 2018;6(1):66.
16. Jiang CY, Yang KM, Yang L, Miao ZX, Wang YH, Zhu HB. A (1)H NMR-Based Metabonomic Investigation of Time-Related Metabolic Trajectories of the Plasma, Urine and Liver Extracts of Hyperlipidemic Hamsters. PloS one. 2013;8(6):e66786.
17. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Analytical chemistry. 2006;78(13):4430–42.
18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27–30.
19. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vazquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, et al: **HMDB 4.0**: **the human metabolome database for** 2018. *Nucleic acids research* 2018, **46**(D1):D608-D617.
20. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology. 2019;37(8):852–7.
21. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.
22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

23. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature methods. 2012;9(8):811–4.
24. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962–8.
25. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD: **The MetaCyc database of metabolic pathways and enzymes-a 2019 update**. *Nucleic acids research* 2019.
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498–504.
27. Ndrepepa G, Holdenrieder S, Colleran R, Cassese S, Xhepa E, Fusaro M, Laugwitz KL, Schunkert H, Kastrati A. Inverse association of alanine aminotransferase within normal range with prognosis in patients with coronary artery disease. Clin Chim Acta. 2019;496:55–61.
28. Dhar I, Lysne V, Seifert R, Svingen GFT, Ueland PM, Nygard OK. Plasma methionine and risk of acute myocardial infarction: Effect modification by established risk factors. Atherosclerosis. 2018;272:175–81.
29. Gu H, Chen W, Yin J, Chen S, Zhang J, Gong J. Methionine sulfoxide reductase A rs10903323 G/A polymorphism is associated with increased risk of coronary artery disease in a Chinese population. Clinical biochemistry. 2013;46(16–17):1668–72.
30. Investigators T, Alexander JH, Reynolds HR, Stebbins AL, Dzavik V, Harrington RA, Van de Werf F, Hochman JS. Effect of tilarginine acetate in patients with acute myocardial infarction and cardiogenic shock: the TRIUMPH randomized controlled trial. JAMA. 2007;297(15):1657–66.
31. Yi Wan FW, Yuan J, Li J, Jiang D, Zhang J, Li H, Wang R, Tang J, Huang T, Zheng J, Sinclair AJ. Jim Mann, Duo Li: **Effects of dietary fat on gut microbiota and faecal metabolites, and their relationship with cardiometabolic risk factors: a 6-month randomised controlled-feeding trial**. *Gut* 2019.
32. Liu Z, Li J, Liu H, Tang Y, Zhan Q, Lai W, Ao L, Meng X, Ren H, Xu D, et al. The intestinal microbiota associated with cardiac valve calcification differs from that of coronary artery disease. Atherosclerosis. 2019;284:121–8.
33. Lu H, Chen Y, Li L. Metabolic Pathway Genes Associated with Susceptibility Genes to Coronary Artery Disease. International journal of genomics. 2018;2018:9025841.
34. Bagheri B, Radmard N, Faghani-Makrani A, Rasouli M. Serum Creatinine and Occurrence and Severity of Coronary Artery Disease. Medical archives. 2019;73(3):154–6.

Figures

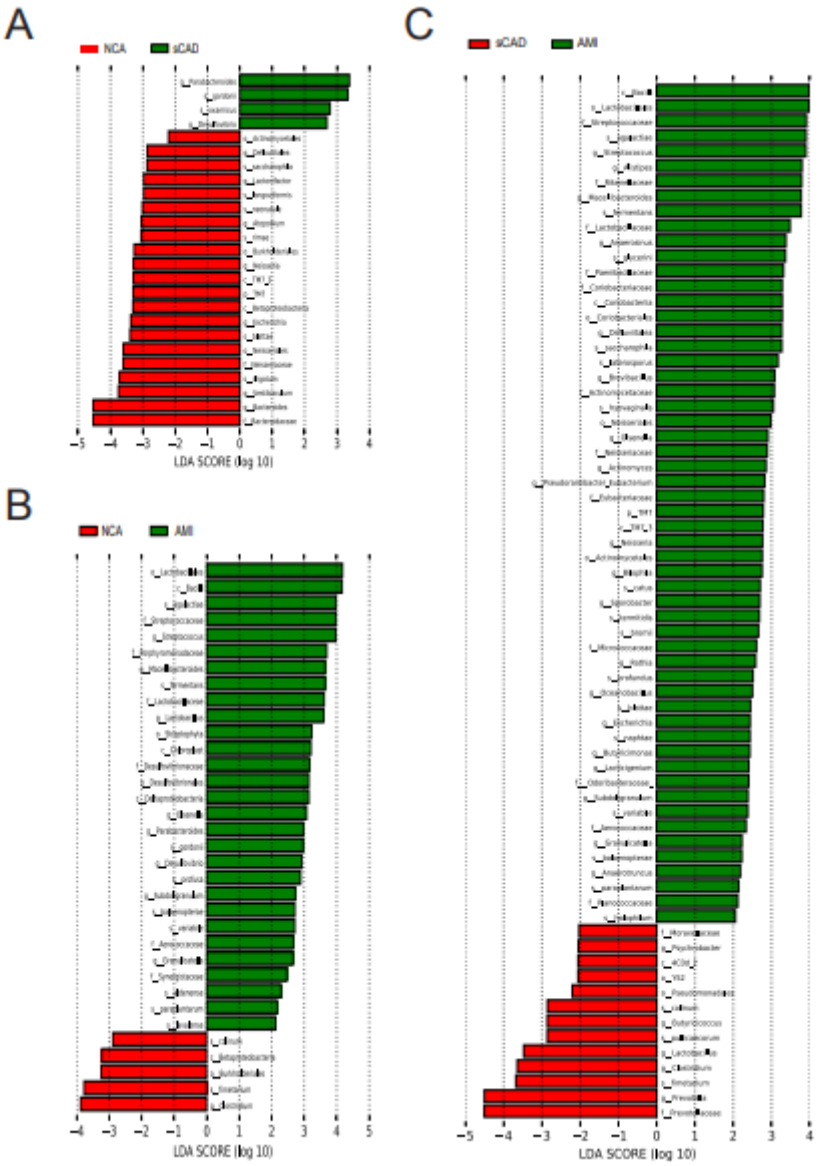

Figure 1

Identification of the enriched bacteria associated with CAD subgroups using LEfSe analysis LEfSe plots showed the most differentially abundant taxa. Only taxa meeting LDA scores > 2 are shown. A: Taxa enriched for NCA in red; taxa enriched for sCAD in green. B: Taxa enriched for sCAD in red; taxa enriched for AMI in green. C: Taxa enriched for NCA in red; taxa enriched for AMI in green.

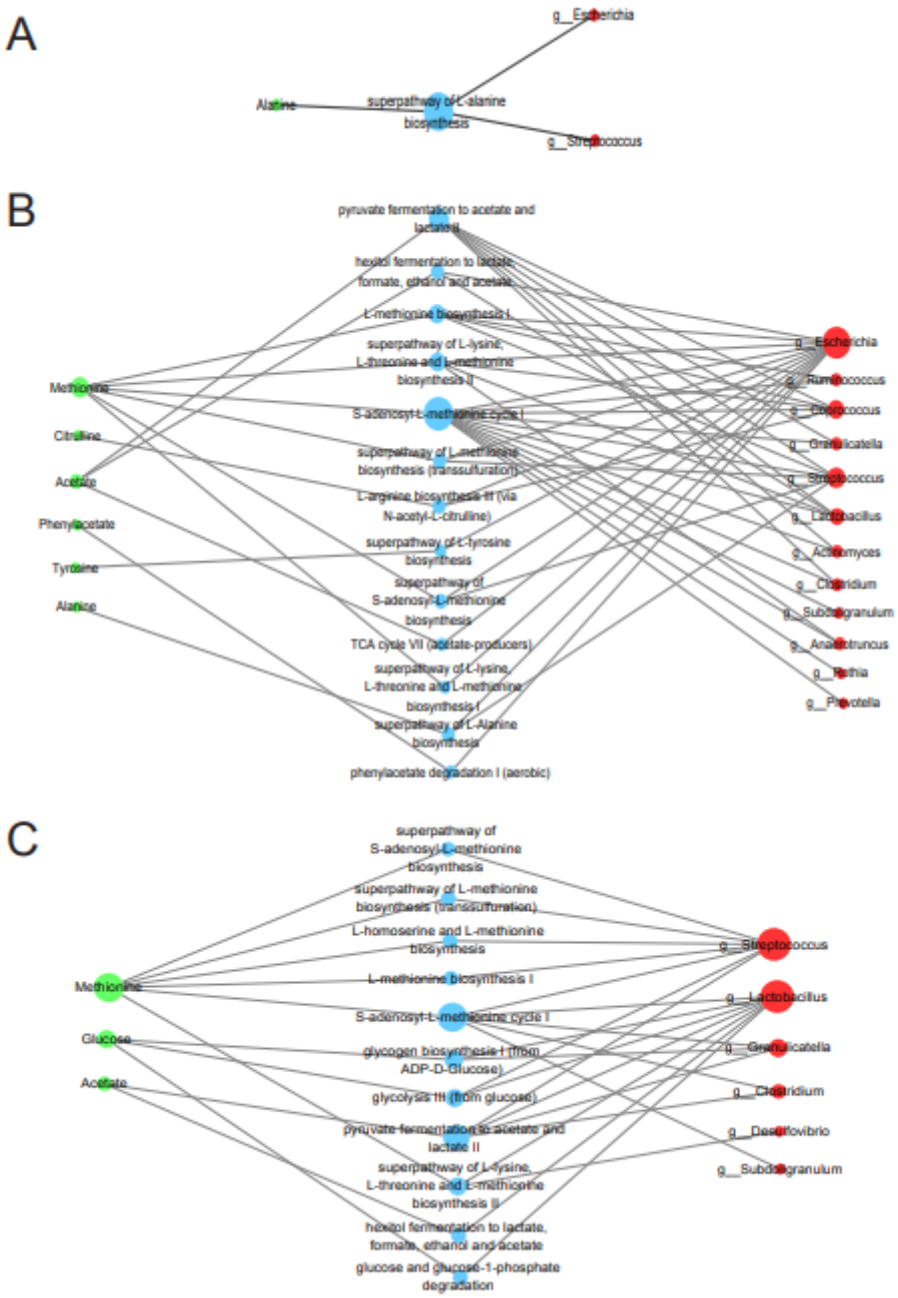

Figure 2

Integrated network of microbiota, metabolites and metabolic pathways among CAD subgroups The direction of enrichment was determined by MetaCyc and HUMAnN2. Sizes of the nodes were in proportion with numbers of links. Green, red and blue nodes represent metabolites, microbiota and related pathways, respectively. A: NCA vs sCAD; B: sCAD vs AMI; C: NCA vs AMI

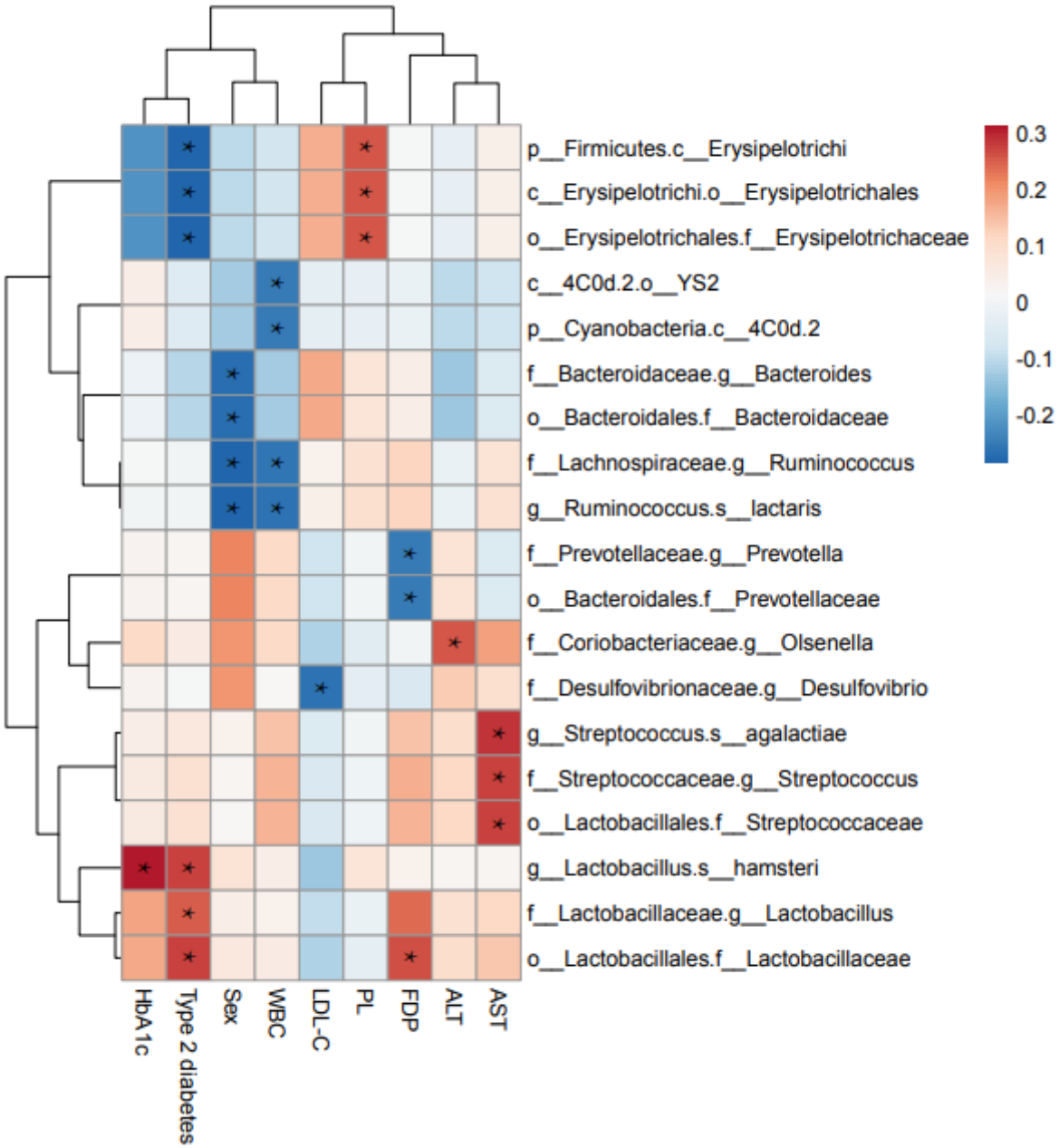

Figure 3

Heatmap of correlation analysis between clinical indexes and microbiota Spearman correlation analysis was performed to assess the correlation of clinical indexes with gut microbiota in the NCA, sCAD and AMI groups. * $p < 0.05$ && rho>0.25

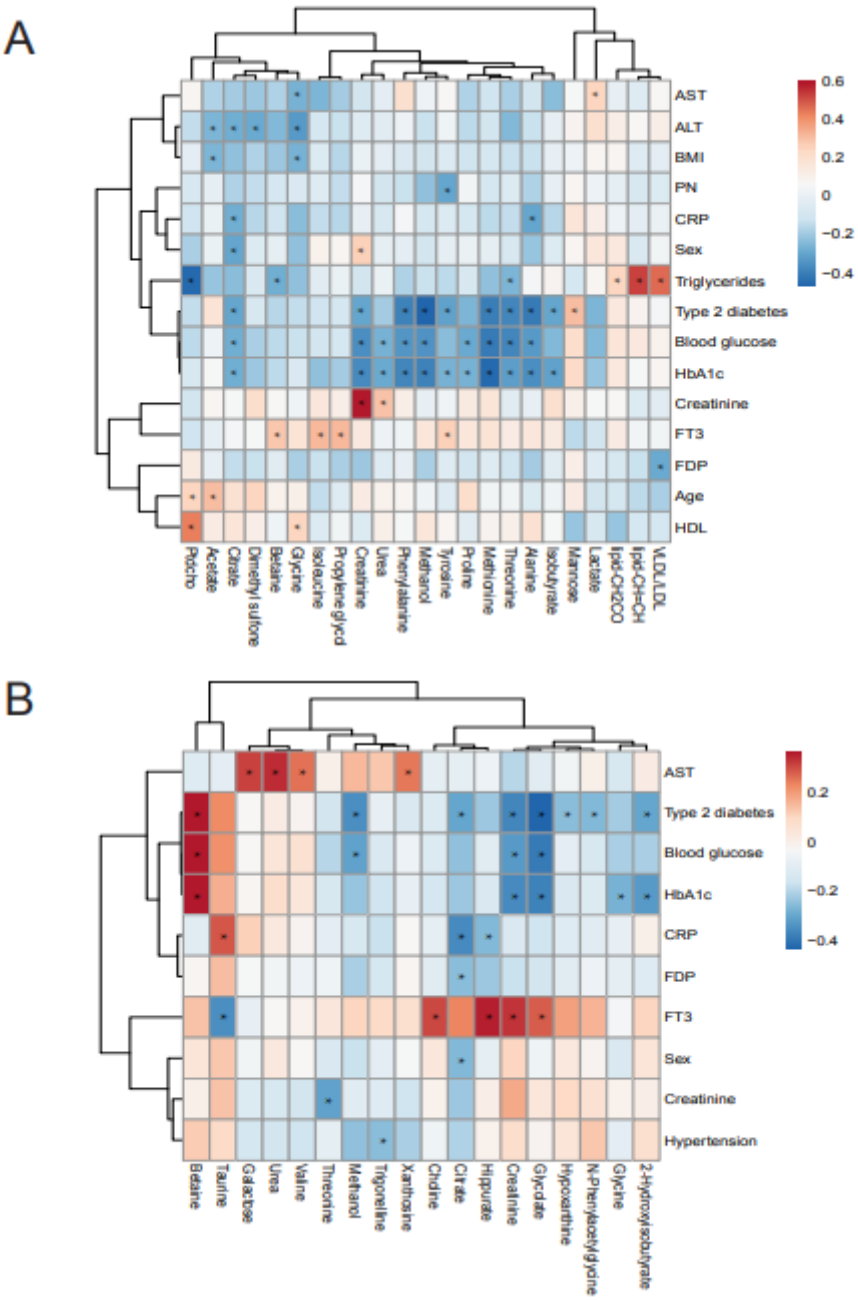

Figure 4

Heatmap of correlation analysis between clinical indexes and three types of metabolites Spearman correlation analysis was performed between serum metabolites and clinical indexes (Figure 4A), urine metabolites and clinical indexes (Figure 4B). Red panel indicated positive correlation, while blue panel suggested negative correlation. * $p < 0.05$ && rho>0.25.

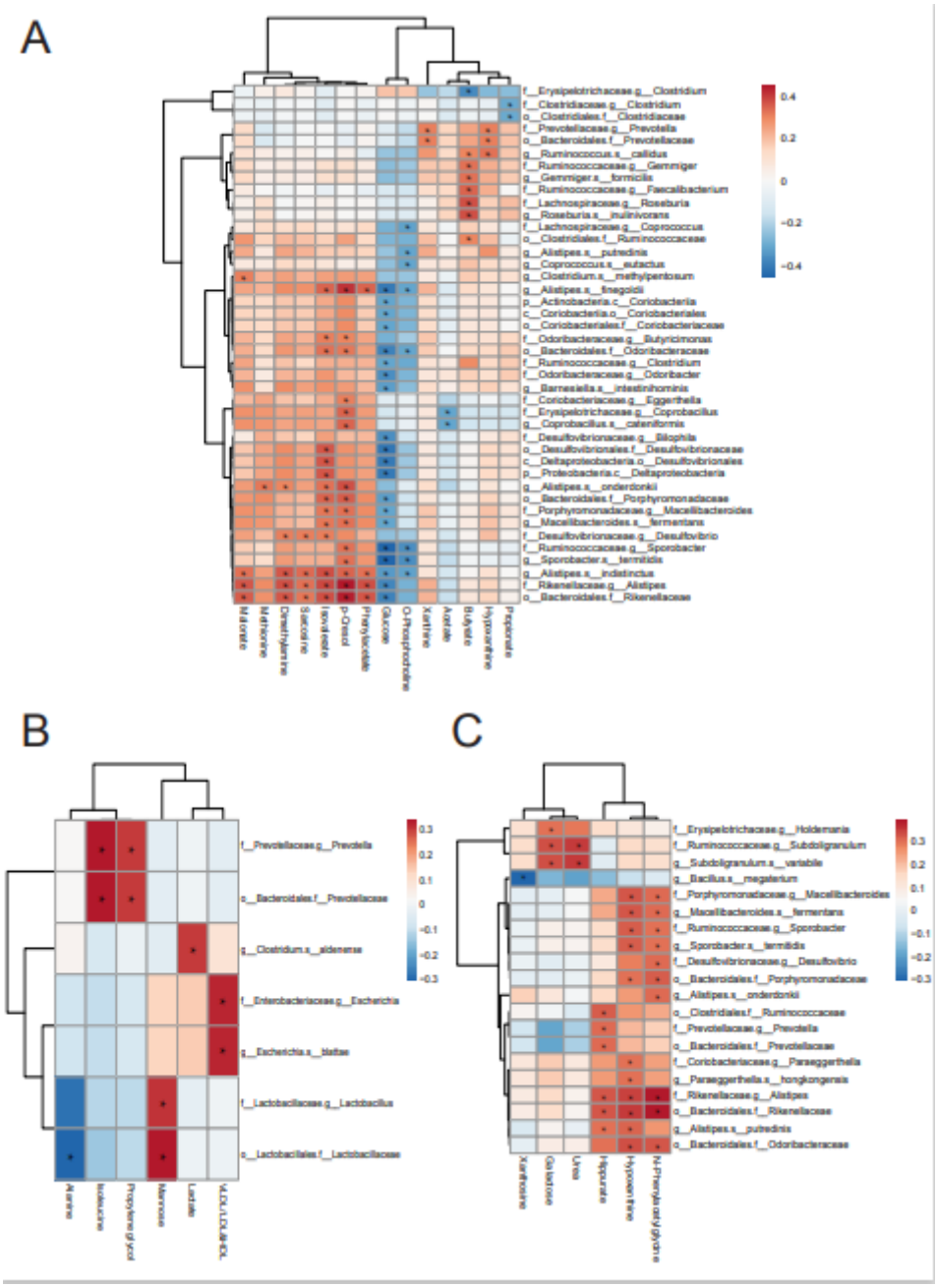

Figure 5

Heatmap of correlation analysis between microbiota and three types of metabolites Spearman's correlation coefficients between the abundance of microbiota and the level of faecal/serum/urine metabolic patterns (Figure 5). Red, negative correlation; blue, positive correlation. * p < 0.05 && rho>0.25.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterial.docx