

Received: 2020.04.24

Accepted: 2020.08.04

Available online: 2020.09.10

Published: 2020.11.02

Detecting Rare Variants and Heteroplasmy of Mitochondrial DNA from High-Throughput Sequencing in Patients with Coronary Artery Disease

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABE 1,2 **Qian Jia***
CE 1 **Lu Xu***
CD 3 **Juan Shen**
CD 3 **Yanping Wei**
B 3 **Huaiqian Xu**
D 2,4 **Jionglong Shi**
D 2,4 **Zhilong Jia**
D 1,2 **Xiaojing Zhao**
D 1,2 **Chunlei Liu**
F 1,2 **Qin Zhong**
AD 1 **Yaping Tian**
A 1,2 **Kunlun He**

1 Core Laboratory of Translational Medicine, Chinese People's Liberation Army (PLA) General Hospital, Beijing, P.R. China
1 Beijing Key laboratory of Chronic Heart Failure Precision Medicine, Chinese People's Liberation Army (PLA) General Hospital, Beijing, P.R. China
3 BGI Genomics, Shenzhen, Guangdong, P.R. China
4 Key Laboratory of Biomedical Engineering and Translational Medicine, Ministry of Industry and Information Technology, Chinese People's Liberation Army (PLA) General Hospital, Beijing, P.R. China

* Qian Jia and Lu Xu are co-first authors

Corresponding Authors: Kunlun He, e-mail: kunlunhe@plagh.org, Yaping Tian, e-mail: tianyp@301hospital.com.cn
Source of support: Departmental sources

Background: Although mutations and dysfunction of mitochondrial DNA (mtDNA) are related to a variety of diseases, few studies have focused on the relationship between mtDNA and coronary artery disease (CAD), especially the relationship between rare variants and CAD.


Material/Methods: Two-stage high-throughput sequencing was performed to detect mtDNA variants or heteroplasmy and the relationship between them and CAD phenotypes. In the discovery stage, mtDNA was analyzed by high-throughput sequencing of long-range PCR products generated from the peripheral blood of 85 CAD patients and 80 demographically matched controls. In the validation stage, high-throughput sequencing for mtDNA target regions captured by GenCap Kit was performed on 100 CAD samples and 100 controls. Finally, tRNA fine mapping was performed between our study and the reported Chinese CAD study.

Results: Among the tRNA genes, we confirmed a highly conserved rare variant, A5592G, previously reported in the Chinese CAD study, and 2 novel rare mutations that reached Bonferroni's correction significance in the combined analysis were found ($P=7.39 \times 10^{-4}$ for T5628C in tRNA^{Ala} and $P=1.01 \times 10^{-5}$ for T681C in 12S rRNA) in the CAD study. Both of them were predicted to be pathological, with T5628C disrupting an extremely conservative base-pairing at the AC stem of tRNA^{Ala}. Furthermore, we confirmed the controversial issue that the number of non-synonymous heteroplasmic sites per sample was significantly higher in CAD patients.

Conclusions: In conclusion, our study confirmed the contribution of rare variants in CAD and showed that CAD patients had more non-synonymous heterogeneity mutations, which may be helpful in identifying the genetic and molecular basis of CAD.

MeSH Keywords: **Coronary Artery Disease • DNA, Mitochondrial • Genetic Variation**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/925401v>

 3190

 6

 4

 40



Background

Coronary artery disease (CAD) has become one of the most common and serious forms of cardiovascular disease affecting national health in recent decades. Although CAD-related death rates have decreased dramatically in the US since the 1960s [1,2], research by Zhu and colleagues showed that morbidity due to CAD had increased in developing countries and decreased in developed countries [3]. Despite the differences between developed countries and developing countries, CAD remains one of the most significant global public health issues.

Currently, CAD is considered a multifactorial disease caused by assorted environmental and hereditary factors. Many risk factors, such as obesity, smoking, hyperlipidemia, hypertension, and diabetes, stimulate the progression of CAD. In addition, genetic factors, primarily the effects of nuclear and mitochondrial genes, also stimulate the progression of CAD. In recent years, genome-wide association study GWAS and meta-analyses of large-scale studies have identified numerous CAD-related polymorphisms in the nuclear genome [4,5]. Furthermore, as a new driving force of cardiovascular disease, somatic mutations in blood cells have become a popular topic in the last 2 years, and the contribution of somatic mutations in the hematopoietic system to CAD has also been confirmed [6,7]. To date, these somatic mutations and germlines in the nuclear genome only account for a small percentage of CAD cases. At the same time, mutations in the mtDNA, which are similar to the somatic mutations of the nuclear genome, have been neglected for many years, especially in heterogeneity studies. Thus, the question that we aimed to address in this study is whether mitochondrial DNA (mtDNA) sequencing can identify new mutation sites associated with cardiovascular diseases such as CAD.

Mutations in and dysfunction of the mtDNA are related to a variety of diseases, including cardiomyopathies [8], essential hypertension [9], diabetes [10], and Leber's hereditary optic neuropathy [11]. ATP is produced in the mitochondria through oxidative phosphorylation within the respiratory chain. Electron transport is coupled with the production of ATP from ADP. However, reactive oxygen species (ROS) are byproducts of the respiratory chain, causing the mitochondria to become the primary source of reactive oxygen species in cells. Cell death and oxidative stress, inflammation, and plaque formation and development are closely related to ROS. Many studies have examined the association between CAD and mtDNA [12–14]; however, to our knowledge, few studies have focused on the heteroplasmy levels of mtDNA. The purpose of this study was to investigate the relationship between mtDNA and CAD, especially rare variants and heteroplasmy levels, using extremely high-depth sequencing.

Material and Methods

Patient population

This study was conducted at the Chinese PLA General Hospital. A total of 85 patients who underwent coronary angiography (CAG) between April 2017 and September 2018 and met the criteria for CAD were consecutively enrolled in this study [15]. Additionally, demographically matched healthy controls with normal computed tomography coronary angiography (CTCA) and without other associated cardiovascular diseases were recruited. The exclusion criteria were malignant tumors, immunological diseases, pregnancy, lactation, and other conditions deemed unsuitable for this study. To verify the previous results, another 100 patients with CAD and 100 controls meeting the same criteria were enrolled. The study complied with the Declaration of Helsinki and was approved by the Ethics Committee of the Chinese PLA General Hospital, and the document number is S.2017-035-01. Informed consent was provided by all of the patients before participation in the study.

Additionally, a published Chinese CAD study on mitochondrial tRNA genes with 80 patients and 512 controls was integrated into our analysis [16]. A total of 2704 sequences containing 1865 complete sequences and 839 coding region sequences were downloaded from the Human Mitochondrial Genome Database (mtDB) [17] (<http://www.mtDB.igp.uu.se/>) for analysis in this study.

High-coverage sequencing of mtDNA

Discovery stage

A pair of primers that targeted the sequence of the MT-CYTB gene on the mitochondrial DNA but did not show sequence similarity with the gene on nuclear DNA was designed in our discovery study (forward primer: 5'TGAGGCCAAATATCATTCTGAGGGGC3'; and reverse primer: 5'TTTCATCATGCGGAGATGTTGGATGG3') [18,19]. Of the mitochondrial genome, 97.99% can be covered by this PCR product. The long-range PCR products were then sheared to approximately 300- to 500-bp fragments and used to construct libraries for sequencing using Illumina HiSeq4000 (San Diego, CA, USA). The libraries' quality was identified using an Agilent 2100 Bioanalyzer. Sequencing reads obtained from this process only contained mtDNA data.

Validation stage

The library was prepared using the Library Preparation Kit (MyGenostics Co., Ltd.), according to the Illumina platform requirements. Briefly, genomic DNA of 1 to 5 µg was fragmented with specific enzymes. The ends of the DNA segments were

repaired and linked with adapters. The length of the mature library was approximately 350–400 bp. The library was amplified, and its quality was identified using an Agilent 2100 Bioanalyzer. The target regions were captured using the GenCap Kit® (Mygenostics Co., Ltd.). Biotin-labeled probes of 60 bp were designed and hybridized to the target region. The target genes were captured by a covalent combination between magnetic beads and modified with streptavidin and these biotin-labeled probes. Subsequently, magnetic beads with target genes were fixed on the magnetic rack. The target genes were enriched after washing and purification. The library with the target genes was loaded into the sequencing Flowcell of the Illumina NextSeq 500 sequencer. A unique “bridge” amplified reaction occurred, and the fluorescent images were collected automatically.

Determination of background sequencing errors

In the analysis of mtDNA heterogeneity, how to distinguish heterogeneity levels from background errors has always been a controversial topic. However, high-coverage sequencing has permitted the identification of heteroplasmic variants even when they are relatively rare (>0.2%) [19,20]. To better distinguish false-positive and false-negative heteroplasmic sites, we estimate the background error among the sequencing samples.

The reads for each sample were aligned to the revised Cambridge reference sequence (GenBank: NC_012920.1) with no more than 4 mismatches using BWA (bwa_mem_v1.0) and were stored in a BAM file. Low-quality reads, unmapped reads and duplication reads were removed. We used GATK's unified genotype to identify SNPs and counted the number of reads supporting the non-reference allele to determine heteroplasmic sites. The read depths for each allele were calculated from mpileup files, which were generated using Samtools software (version 0.1.18), with the reads of a base quality score less than 20 (Phred-scaled) for bases and mapping quality scores less than 50 (Phred-scaled) not counted [18].

Based on the above processing, the background error rate was determined. The maximum background error rate was 0.45%. We doubled it (~1%) and used 1% as the heteroplasmic threshold. In addition, only sites with 5% of the samples with mtDNA heterogeneity mutations were considered to perform single-association analysis. Additionally, we focused our efforts on rare functional mutations. Functional mutations were defined as follows: 1) the conservative score of mutation sites was greater than 0.75, or the mutation generated the classic Watson-Crick base pair for tRNA and 12S/16S rRNA; 2) missense mutations were annotated as “deleterious” by at least 1 of the 2 protein prediction algorithms (SIFT or Polyphen), and their conservative score was also greater than 0.75.

Structural analysis

Published website tools (tRNAscan-SE 2.0) were used to identify the secondary structures of the tRNAs [21] and the classic Watson-Crick (WC) base pair was detected according to these structures.

Conservative analysis

We used Muscle software (a popular multiple alignment software) to calculate the conservation scores of mutation sites in 17 species by multiple sequence alignment [16]. The score is the percentage of the human nucleotide variations in 16 other species with wild-type nucleotides at this site [16].

Statistical analysis

We used the means±standard deviations and percentages to summarize the results of continuous and categorical variables, respectively. The *t* test or one-way analysis of variance was used to test the continuous variables under the assumption of normal distribution. Otherwise, the Mann-Whitney U test was used. For categorical variables, chi-square analysis was performed. All of the statistical analyses were performed using SPSS software (version 20.0 for Windows), and *P*<0.05 was considered statistically significant. We also performed power analysis. Based on the previous studies on CAD [3,22], we assumed that the prevalence was 0.05 and selected the Bonferroni's correction significance (0.05/40=0.00125) as the significance level [23]. We set the case-control ratio to be 185 cases vs. 2884 controls (our study and the public database [17]). For rare SNVs, we calculated the power based on different rare allele frequencies and genotype relative risk levels using Power Calculator (http://csg.sph.umich.edu/abecasis/gas_power_calculator/).

Results

Baseline clinical characteristics

Eighty healthy controls and 85 patients with CAD were enrolled in the discovery stage, while 100 healthy controls and 100 patients with CAD were included in the validation stage. The baseline characteristics of all the participants in the discovery stage are presented in Table 1. The distributions of healthy controls and patients with CAD were similar in age, sex distribution, blood pressure, and body mass index. Regarding the biochemical tests, glucose (Glu) was much higher in patients with CAD in the validation stage (*P*<0.001), while superoxide dismutase (SOD) levels were much higher in the CAD group in both stages.

Table 1. Baseline clinical characteristics of all participants.

	Discovery stage			Validation stage		
	Control (n=80)	CHD (n=85)	P	Control (n=100)	CHD (n=100)	P
Age	63.10±10.69	63.07±11.42	0.763	62.27±11.22	65.11±12.26	0.524
Male n (%)	55 (68.75)	63 (74.12)	0.445	61 (61.00)	72 (72.00)	0.134
SBP(mmHg)	123.39±14.36	129.56±13.54	0.079	126.61±15.33	131.12±14.12	0.145
DBP(mmHg)	71.31±9.54	74.62±9.21	0.316	72.45±10.21	76.01±9.94	0.202
BMI(kg/m ²)	24.72±3.73	25.07±3.27	0.522	23.17±2.17	24.88±3.15	0.674
Biochemical						
ALT (U/L)	18.85±7.33	24.74±11.66	0.085	17.53±9.42	23.57±14.57	0.087
AS T(U/L)	19.16±9.84	21.05±12.62	0.161	18.64±8.27	22.32±11.39	0.090
TBIL (ummol/L)	11.71±5.62	11.42±4.91	0.657	10.79±5.42	12.68±4.53	0.532
DBIL (ummol/L)	3.77±1.86	3.35±1.96	0.474	3.55±1.63	3.82±1.38	0.573
Cr (umol/L)	70.14±18.05	89.96±35.37	0.181	73.26±18.31	86.56±23.53	0.287
BUN (mmo/L)	5.51±2.40	6.44±2.60	0.069	5.69±2.15	6.24±2.93	0.182
UA (mmo/L)	374.27±28.53	397.48±32.86	0.672	356.35±26.61	403.93±30.70	0.378
TG (mmol/L)	1.52±0.63	1.61±0.96	0.091	1.59±0.41	1.66±0.85	0.176
T-CH (mmol/L)	3.93±0.92	4.22±1.11	0.884	3.47±0.78	4.52±1.36	0.640
HDL (mmol/L)	1.13±0.33	1.05±0.34	0.406	1.16±0.37	1.05±0.43	0.346
LDL (mmol/L)	2.40±0.81	2.63±0.85	0.862	2.42±0.87	2.72±0.90	0.603
GLU (mmol/L)	5.22±1.36	5.73±1.84	0.257	5.07±0.93	6.17±2.15	<0.001*
SOD (U/ml)	155.29±17.48	136.54±20.62	0.010*	156.38±21.58	137.31±19.24	0.026*

Data are expressed as means±standard deviation or percentages as appropriate. ALT – alanine aminotransferase; AST – aspartate aminotransferase; TBIL – total bilirubin; DBIL – direct bilirubin; Cr – creatinine; BUN – blood urea nitrogen; UA – uric acid; TG – triglyceride; T-CH – total cholesterol; HDL – high-density lipoprotein; LDL – low-density lipoprotein; GLU – glucose; SOD – superoxide dismutase. * $P<0.05$.

Analysis of mtDNA heterogeneity

In the discovery stage, the average depth of coverage for 165 individuals sequenced by Illumina Hiseq4000 was approximately 7705×, which was sufficient power to detect low-frequency heteroplasmies. This study identified 191 heteroplasmic sites, 46 of which were non-synonymous mutations. The same analysis process was performed for the validation data, and an average depth of coverage of 2670× was obtained (Figure 1A, 1B). The average depth of each site and the heteroplasmic levels for A, C, G, and T among the 165 mtDNA samples in the discovery stage and 200 mtDNA samples in the validation stage are shown in Figure 1C and 1D. Finally, 477 heteroplasmic sites were identified, 120 of which were non-synonymous mutations, as shown in Table 2.

We further studied the site frequency spectrum of heterogeneity and found that the derived allele frequency (DAF) distributions

of these sites were significantly shifted toward lower frequencies, consistent with recent reports [20,24]. In the validation stage, we confirmed that there were many low-level heteroplasmic sites in the mitochondria genome, and the frequency spectral characteristics of heteroplasmy in these 2 studies were very similar (Figure 2). Perhaps due to the insufficient sample size, we did not find significant heterogeneity *loci* associated with CAD.

Subsequently, we investigated whether the total number of heteroplasmic sites differed between the CAD patients and the healthy controls, and we computed heteroplasmy counts for each sample, using Wilcoxon's statistic to test the significance [18]. In the discovery stage, we found that the non-synonymous heteroplasmy counts were significantly different between the CAD patients and controls (mean fold increase=1.50; and $P=1.51\times 10^{-4}$, Figure 3A1–A3). The same results were produced during the validation stage since the

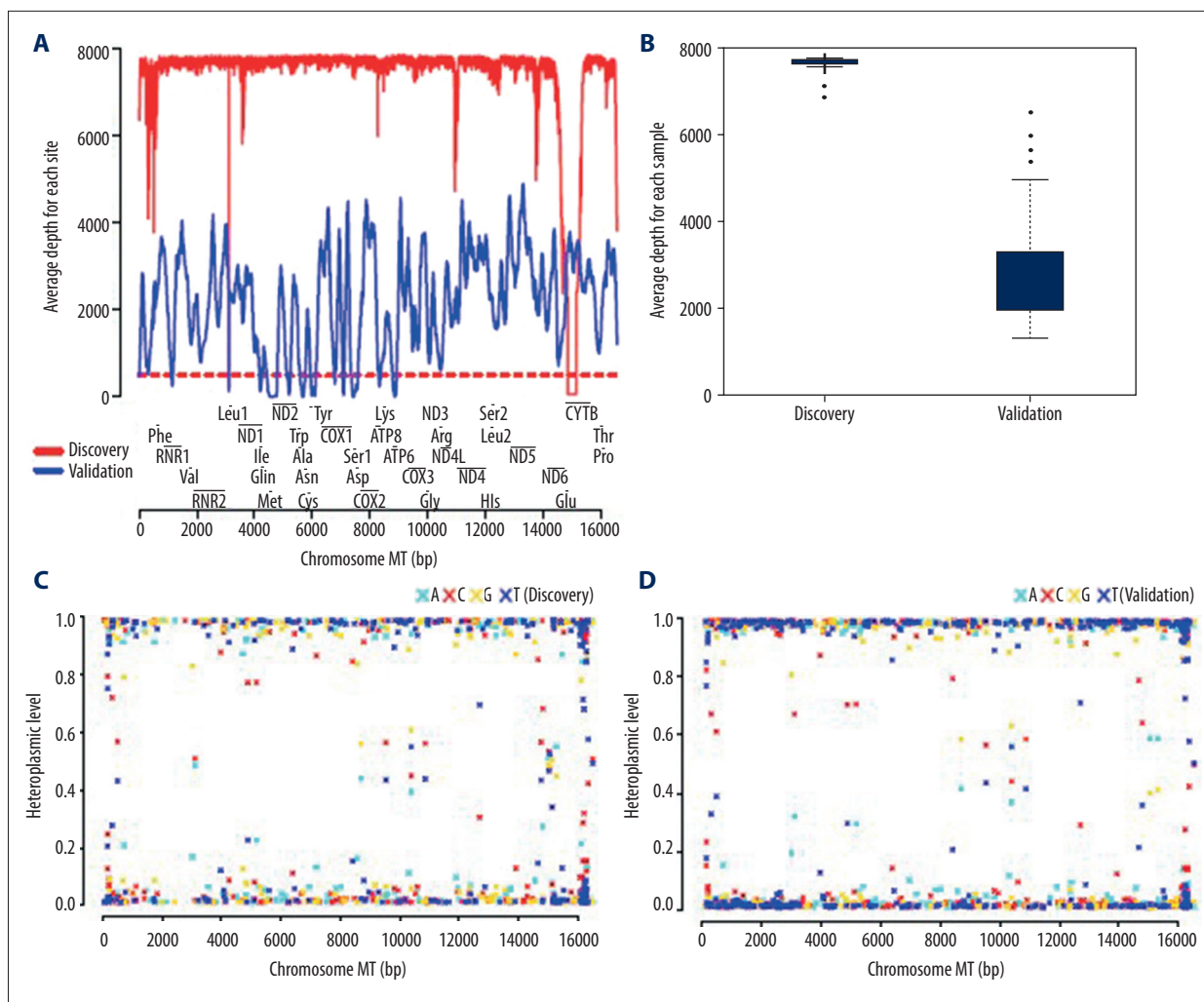


Figure 1. Summary of the distribution of read depth (A, B) and heteroplasmic levels of A, C, G, and T for each site (C, D).

Table 2. Summary of heteroplasmy detection in our study.

Sample	No. of samples (case/control)	No. of samples carrying at least one heteroplasmy	No. of heteroplasms* (case/control)	No. of non-synonymous heteroplasms* (case/control)
Discovery	165 (85/80)	165	191 (72/36)	46 (18/9)
Validation	6200 (100/100)	200	477 (187/69)	120 (39/13)

* The heteroplasmic sites were split into 3 groups according to the percentage of samples carrying 1 heteroplasmic site (e.g., the percentage >0%, >2%, and 5%). For each group, we counted the number of heteroplasmic sites.

non-synonymous heteroplasmy counts also showed a significant difference between cases and controls (mean fold increase=1.19 and $P=2.65 \times 10^{-2}$, Figure 3B1–B3). Furthermore, we obtained a stronger signal under common heteroplasmic sites ($P_{combined}=2.73 \times 10^{-5}$ and mean fold increase=1.31 for all heteroplasmic sites vs. $P_{combined}=8.77 \times 10^{-7}$ and mean fold increase=1.44 for common heteroplasmic sites with frequency >2% or $P_{combined}=4.10 \times 10^{-6}$ and mean fold

increase=1.45 for common heteroplasmic sites with frequency >5%, Figure 3C1–C3). The effect size of this difference in the validation stage was very similar to that in the discovery stage.

Evaluation of SNP variants

We identified 758 variants in the discovery stage, including 42 variants in tRNA genes, 60 variants in rRNA genes, 134 variants

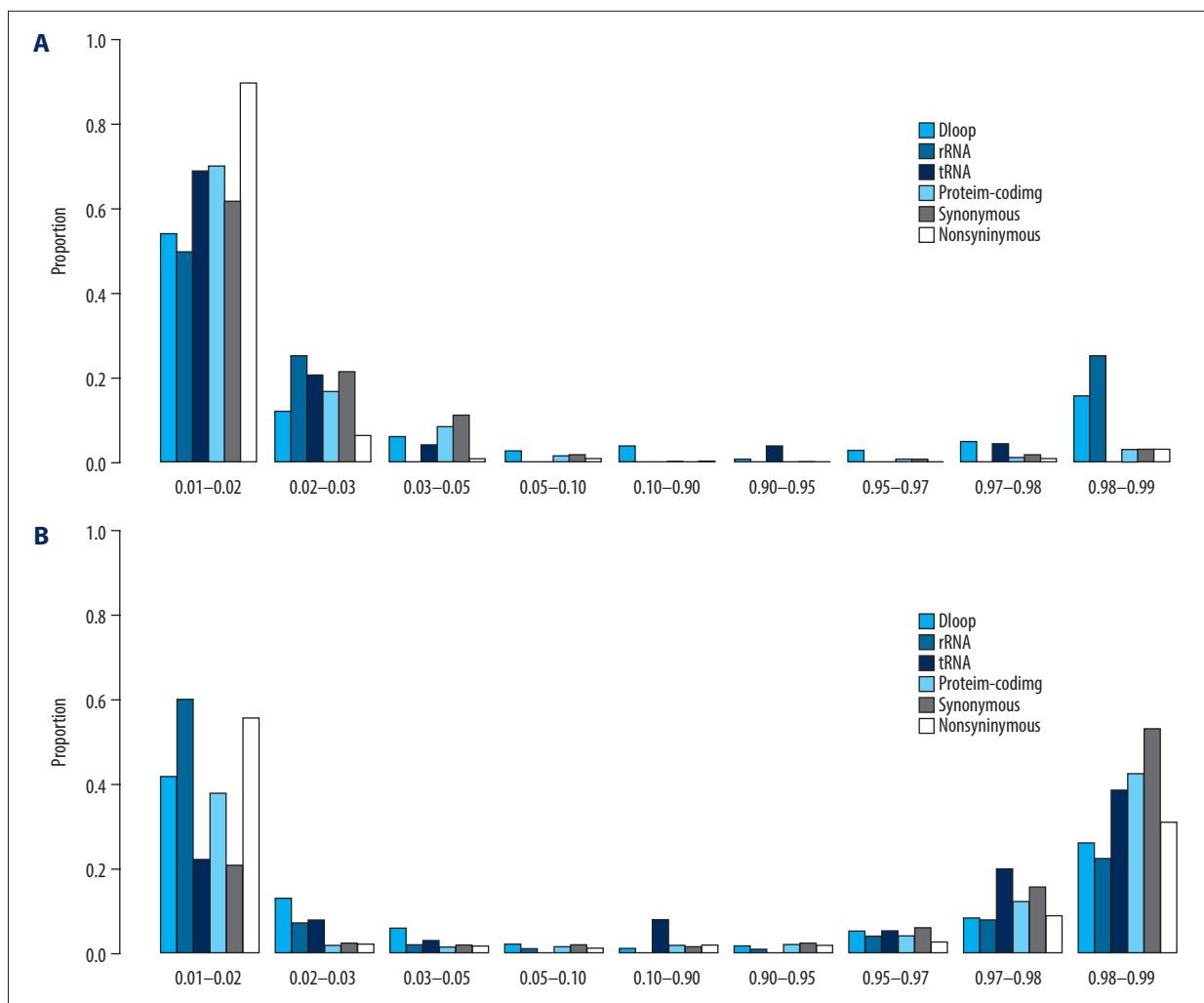


Figure 2. The DAF distribution of heteroplasmy in different mtDNA genomic regions for discovery samples (A) and validation samples (B), respectively.

in the D-Loop functional regions, and 522 variants in protein-coding genes, and 851 variants in the validation stage, including 60 variants in tRNA genes, 51 variants in rRNA genes, 169 variants in the D-Loop functional regions, and 571 variants in protein-coding genes. Of these variants, 54.61% were present in the whole-genome sequencing data from 1000 Genomes Phase 3, and 55.93% were of low frequency ($MAF < 0.01$), consistent with the results from high-throughput sequencing of autosomes. In total, 40 rare functional mutations (including 17 tRNA variants, 12 rRNA variants, and 11 damaging missense variants) in the discovery stage were selected for association analysis (Bonferroni's correction significance $P = 0.00125$, $0.05/40$, Tables 3 and 4).

Based on an integrative analysis of 8 mutations detected in previous tRNAs studies of CAD in the Chinese population (Table 3), we confirmed the association signal A5592G; however, another signal, G15927A, was not validated in our study.

We also detected other mutation sites on tRNAs. As shown in Figure 4, structural variations in tRNAs changed the classical Watson-Crick base pair. Of these variants, only T5628C had a significant signal ($P = 7.39 \times 10^{-4}$), and it disrupted an extremely conservative base-pairing at the AC stem of tRNA^{Ala} (Table 4). In addition, we found that T681C also reached Bonferroni's correction significance ($P = 1.01 \times 10^{-5}$), occurring in the stems of 12S rRNA, and it changed the classical Watson-Crick base pair (Table 5).

Discussion

Mitochondria are often regarded as cellular power stations because of their ability to generate ATP. Recently, studies have found that mitochondrial dysfunction promotes oxidative stress, inflammation, cell apoptosis, and cholesterol accumulation, which are key processes in atherosclerosis. Several

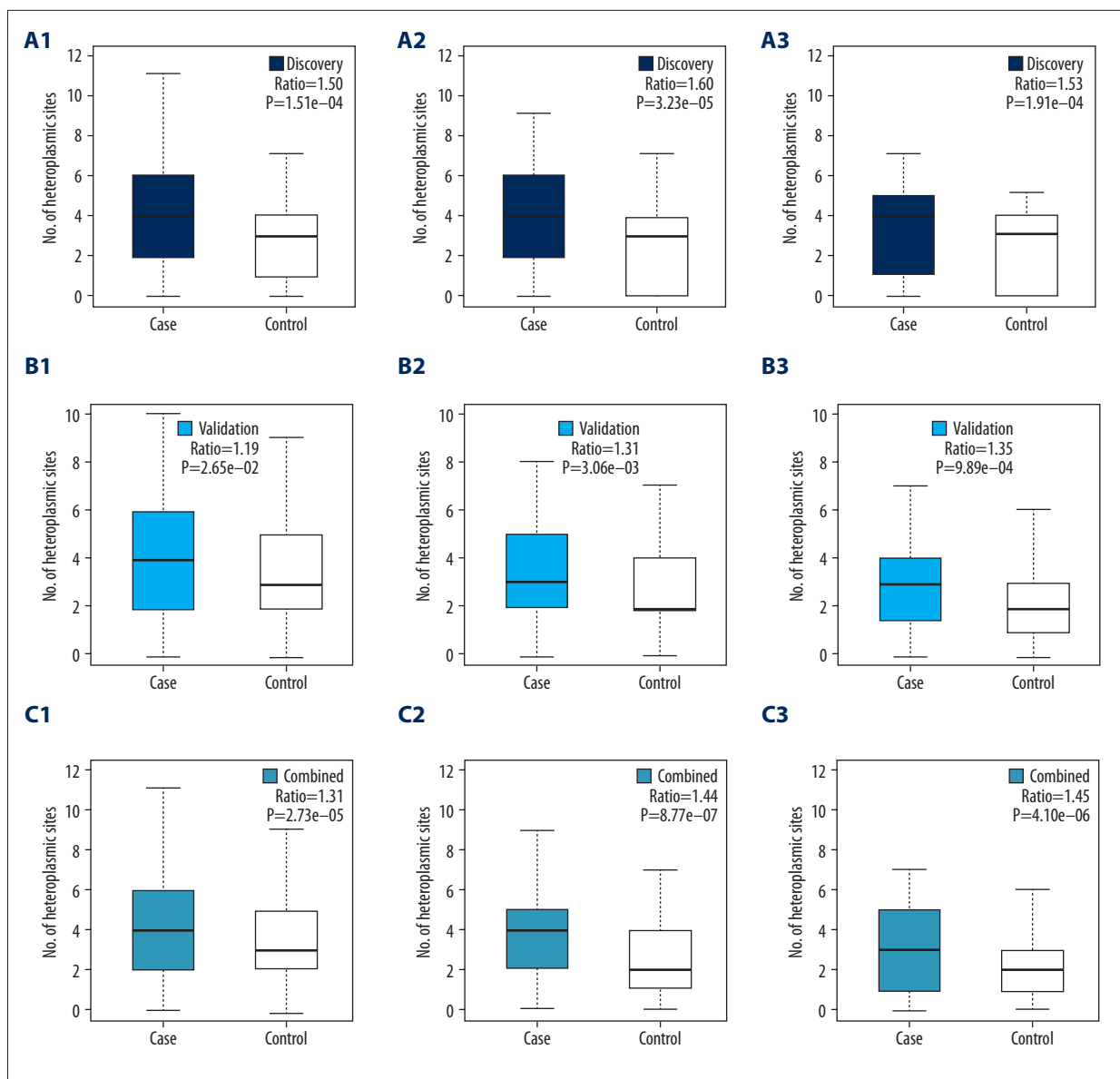


Figure 3. The non-synonymous heteroplasmy counts differed between cases and controls in both stages. We split heterogeneity sites into 3 groups according to the percentage of samples carrying one heterogeneity sites: the percentage >0% (**A1, B1, C1**), >2% (**A2, B2, C2**), and >5% (**A3, B3, C3**). For each group, the Wilcoxon statistic was used to test the heteroplasmy counts difference. The ratio was calculated as the mean count of heteroplasmic sites in cases divided by the mean count of heteroplasmic sites in controls.

pieces of evidence have suggested that numerous clonally expanded mutations of mtDNA exist in human tissues [16,21,24], but the main reason to choose the whole blood of participants to study CAD is that somatic mutations in blood cells have become a new driving force in cardiovascular research in recent years. Jaiswal et al. detected the presence of clonal hematopoiesis of indeterminate potential (CHIP) in peripheral blood cells using whole-exome sequencing and associated this presence with early-onset acute coronary syndrome and CAD [6]. Since intra-cellular mutations of mtDNA are an

important source of mitochondrial genotype heterogeneity [25], the widespread heterogeneity of the mitochondria in CAD patients has raised concerns that they may expand to high frequencies in a small proportion of cells in the later stages of life, exceeding the critical phenotype threshold and causing CAD. More research is necessary to reveal the mechanism of the cloning and amplification of pathogenic heterogeneity to further clarify the role of mitochondrial heteroplasmy in complex diseases. Controlling the expansion of mtDNA mutations could be a promising method to prevent disease progression.

Table 3. Integrated analysis of variants in tRNA genes with the previously reported samples. These variants came from a published Chinese CAD study in mitochondrial tRNA gene with 80 patients and 512 controls [16].

Variants	Refer/Alt	Cons	WC Base Pairs	Our study		Reported study [16]		Public database [17]	
				185 cases	180 controls	80 cases	512 controls	2704 controls	
4386	T/C	0.75		6	6	1	4	51	
5592	A/G	1.00	G-C ↑	1	0	1	0	3	
5821	G/A	0.63	G-C ↓	2	4	1	12	14	
14693	A/G	1.00		4	2	2	7	10	
15889	T/C	0.19	U-A ↓	1	0	2	1	3	
15927	G/A	0.69	G-C ↓	2	5	4	7	44	
15928	G/A	0.69	G-C ↓	3	0	1	2	132	
15951	A/G	0.56	A-U ↓	1	5	1	2	22	

Variants	Refer/Alt	Cons	WC Base Pairs	Our study + reported study				Our study + reported study + public database			
				265 cases	692 controls	P	OR	265 cases	3396 control	P	OR
4386	T/C	0.75		7	10	0.2740	1.82	7	61	0.3385	1.48
5592	A/G	1.00	G-C ↑	2	0	0.0782	inf	2	3	0.0451	8.59
5821	G/A	0.63	G-C ↓	3	16	0.3065	0.48	3	30	0.7297	1.28
14693	A/G	1.00		6	9	0.3829	1.73	6	19	0.0074	4.11
15889	T/C	0.19	U-A ↓	3	1	0.0690	7.77	3	4	0.0105	9.70
15927	G/A	0.69	G-C ↓	6	12	0.6018	1.29	6	56	0.4528	1.38
15928	G/A	0.69	G-C ↓	4	2	0.0552	5.19	4	134	0.0434	0.37
15951	A/G	0.56	A-U ↓	2	7	1.0000	0.73	2	29	1	0.88

Cons – conservative; WC Base Pairs – Watson-Crick base pairs.

Table 4. Frequency comparison of tRNA variants with an additional 2704 controls from the mtDB database [17]. These tRNA variants were absent from the published paper [16] but were detected in our study.

Variants	Refer/Alt	Cons	WC Base Pairs	Discovery		Validation		Combined		P	Our study + public database [17] (185 cases vs. 2704+180 controls)			
				85 cases	80 controls	100 cases	100 controls	185 cases	180 controls		2884 controls	P	OR	Gene
14696	A/C,G	0.81	G-C ↑	1	0	0	0	1	0	1	6	0.3532	2.61	tRNA-Glu
8334	G/A	1.00		1	0	0	0	1	0	1	1	0.1169	15.63	tRNA-Lys
5788	T/C	1.00		1	0	0	0	1	0	1	6	0.3532	2.61	tRNA-Cys
5783	G/A	0.94	G-C ↓	1	0	0	0	1	0	1	1	0.1169	15.63	tRNA-Cys
5628	T/C	0.94	U-A ↓	3	0	1	0	4	0	0.1231	4	0.0007*	15.86	tRNA-Ala
5587	T/C	0.94		1	0	0	2	1	2	0.6189	1	0.2203	5.21	tRNA-Ala
5539	A/G	1.00	G-C ↓	1	0	0	1	1	1	1	3	0.2673	3.91	tRNA-Trp
5514	A/G	0.88	U-A ↓	1	2	0	0	1	2	0.6189	0	0.1702	7.82	tRNA-Trp
606	A/G	0.56	U-A ↓	1	0	0	0	1	0	1	1	0.1169	15.63	tRNA-Phe

Cons – conservative; WC Base Pairs – Watson-Crick base pairs.

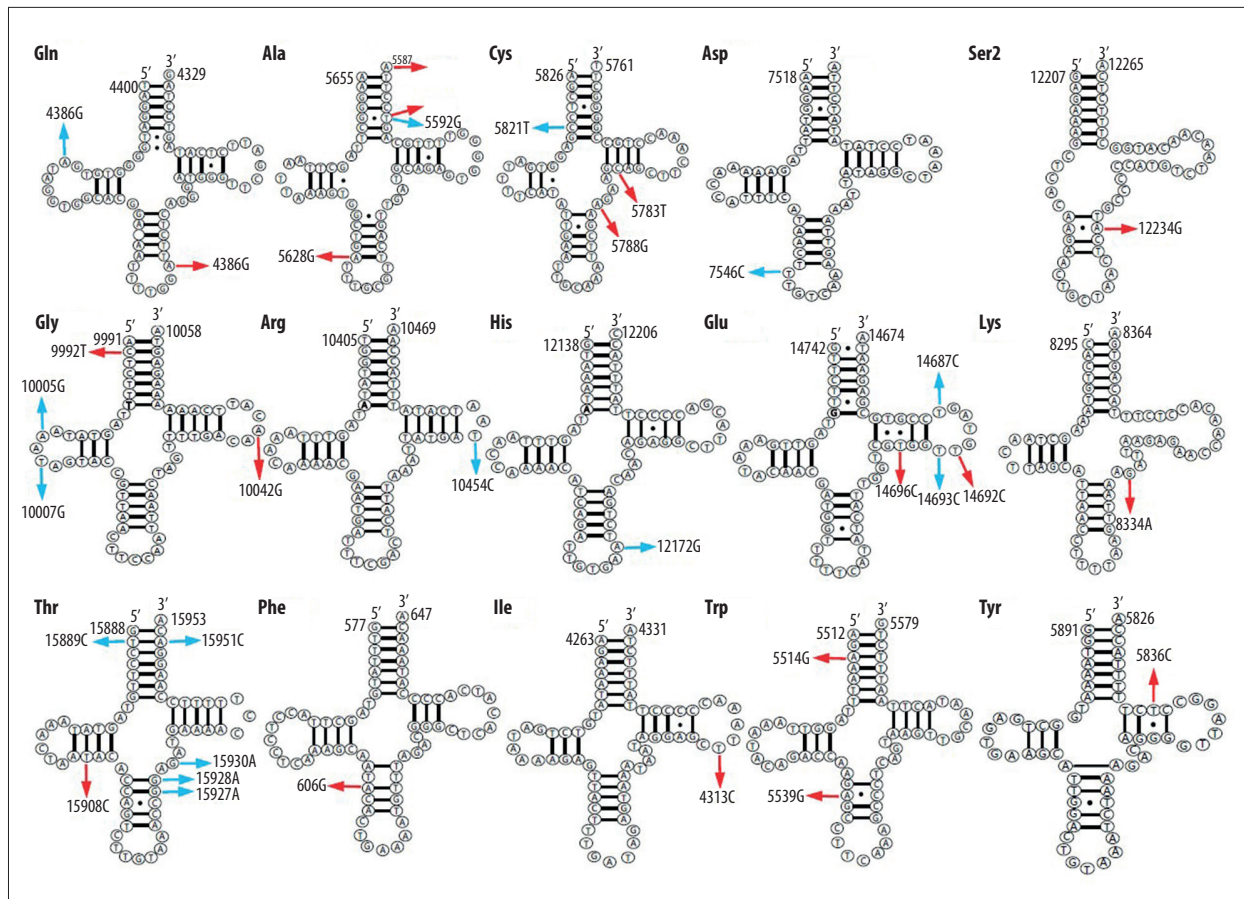


Figure 4. Summary of tRNA variations in 2 studies of coronary heart disease in the Chinese population. Fifteen published tRNA structures are displayed. The blue arrows represent the mutation sites found in a previous Chinese CAD study, and the red arrows represent the novel variants after strict filtering in our study.

In our study, we identified a set of mtDNA mutations in CAD patients and healthy controls using high-throughput sequencing. These results provided a better understanding of the mtDNA in CAD patients and showed a large number of rare variants in CAD. High-throughput sequencing of long-range PCR products now allows for the simultaneous screening of single-nucleotide polymorphisms (SNPs) and low-level heteroplasmy. Among tRNA genes, a novel rare mutation in CAD research, T5628C (combined $P=7.39E \times 10^{-4}$), disrupted highly conservative base-pairing at the AC stem of tRNA^{Ala}. This mutation altered the structure of tRNA^{Ala}, which could lead to abnormal metabolism of tRNA and then cause mitochondrial dysfunction [26], while mutations with pathological potential in the occurrence and development of CAD are biologically possible due to the necessity of the amino acylation defect of tRNAs for protein synthesis caused by Watson-Crick mispairing [27]. Previous studies have shown that T5628C is associated with hypertension [28], gout [29], dysphagia [30] and hearing loss or deafness [26]. Among these related diseases, hypertension and hyperuricemia are closely related to CAD. It is universally known that hypertension is one of the risk factors for CAD [31].

In the present study, the SBP and DBP of the CAD group were higher than those of the controls, especially in the validation stage. We believe that if the sample size is further expanded, the results could be more significant. Uric acid has drawn attention in the study of risk factors for CAD in recent years. Whether uric acid is an independent predictor of CAD has been controversial, with some studies suggesting that uric acid is an independent risk factor for CAD [32,33], while others have drawn negative conclusions [34]. In our study, although the difference in uric acid levels between the CAD group and the healthy control group did not reach statistical significance, the concentration of uric acid in the CAD group was higher than that in the control group both in the discovery and validation stages. Recent studies suggest that uric acid is biologically active and can stimulate atherosclerosis through oxidative stress [35], endothelial dysfunction [36], and inflammation [37]. Therefore, we speculated that the pathogenic site could play a crucial role in the development of CAD by influencing blood uric acid concentrations. However, its actual mechanism still requires further study.

Furthermore, our research provides strong confirmation of the results of previous CAD studies in the Chinese population. We confirmed a highly conserved, rare variant, A5592G, previously reported in a Chinese CAD study [16]. This study found a 5592 mutation in a family. Based on this and previous studies, the mutation was considered to be a risk factor for CAD. The mitochondrial A5592G altered the structure and function of tRNA^{Ala}, leading to oxidative stress and mitochondrial dysfunction [16]. The sample size of Qin’s study was small and had not been confirmed in sporadic CAD. Our study targeted the population of sporadic CAD, which is an important complement to the pathogenicity of this mutation.

In addition to the discovery in the tRNA region, our study also yielded some interesting results in the rRNA and protein-coding regions. As shown in Table 5, 5 mutations were found in the stems of rRNAs that changed the classical Watson-Crick base pair. Three of them had the consistent effect direction and were statistically significant ($P=1.01 \times 10^{-5}$ for T681C, $P=3.2 \times 10^{-2}$ for A735G, $P=3.63 \times 10^{-3}$ for A2412G). Among them, A735G has been reported in hearing loss [38], while T681C in RNR1 and A2412G in RNR2 are novel variants that had not been previously reported. For the protein-coding region, we found that C15402T had a consistent effect direction and was statistically significant ($P=1.97 \times 10^{-3}$).

Table 5. The association signal for mitochondrial rRNA and protein-coding variants. Joint analysis of variants in our study with an additional 2704 controls [17].

Variants	Refer/Alt	WC Base Pairs	Cons	Discovery		Validation		Combined			
				85 cases	80 controls	100 cases	100 controls	185 cases	180 controls	P	OR
681	T/C	U-A ↓	0.81	2	1	6	0	8	1	0.0369	8.1
735	A/G	G-C ↑	0.88	2	0	0	0	2	0	0.4987	Inf
752	C/T		0.88	3	5	3	2	6	7	0.7842	0.8
789	T/C		0.81	2	0	0	0	2	0	0.4987	Inf
951	G/A		0.81	1	0	0	2	1	2	0.6189	0.5
1095	T/C	U-A ↓	0.94	3	0	0	6	3	6	0.3318	0.5
1762	A/G		0.88	1	0	0	0	1	0	1	Inf
1914	A/G		0.94	1	0	0	0	1	0	1	Inf
2281	A/C, G		0.81	1	0	1	0	2	0	0.4987	Inf
2363	A/G	U-A ↓	0.88	1	0	0	0	1	0	1	Inf
2412	A/G	U-A ↓	0.81	2	0	0	0	2	0	0.4987	Inf
2757	A/G		0.81	1	0	0	0	1	0	1	Inf
6340	C/T		0.81	1	0	0	0	1	0	1.00	Inf
8459	A/G		0.81	1	0	1	2	2	2	1.00	1.0
8572	G/A		0.88	0	2	1	1	1	3	0.37	0.3
8854	G/A		0.88	1	0	0	0	1	0	1.00	Inf
8921	G/A		1.00	1	0	0	0	1	0	1.00	Inf
10086	A/G		0.94	1	0	0	0	1	0	1.00	Inf
11253	T/C		0.81	1	2	0	0	1	2	0.62	0.5
14225	C/T		0.75	1	0	0	0	1	0	1.00	Inf
14337	C/T		1.00	0	1	1	2	1	3	0.37	0.3
15402	C/T		0.94	2	0	1	0	3	0	0.25	Inf
15617	G/A		1.00	1	0	0	0	1	0	1.00	Inf

Table 5 continued. The association signal for mitochondrial rRNA and protein-coding variants. Joint analysis of variants in our study with an additional 2704 controls [17].

Variants	Refer/Alt	WC Base Pairs	Cons	Our study + public database [17] (185 cases vs. 2704+180 controls)			PolyPhen2	SIFT	Gene	Amino acid	Associated disease
				2704 controls	P	OR					
				681	T/C	U-A ↓					
735	A/G	G-C ↑	0.88	3	3.20E-02	10.5	–	–	RNR1	12S rRNA	-
752	C/T		0.88	20	1.25E-02	3.5	–	–	RNR1	12S rRNA	-
789	T/C		0.81	1	1.04E-02	31.4	–	–	RNR1	12S rRNA	-
951	G/A		0.81	8	4.96E-01	1.6	–	–	RNR1	12S rRNA	-
1095	T/C	U-A ↓	0.94	5	4.79E-02	4.3	–	–	RNR1	12S rRNA	-
1762	A/G		0.88	0	6.03E-02	Inf	–	–	RNR2		-
1914	A/G		0.94	0	6.03E-02	Inf	–	–	RNR2		-
2281	A/C, G		0.81	2	2.00E-02	15.7	–	–	RNR2	16S rRNA	-
2363	A/G	U-A ↓	0.88	0	6.03E-02	Inf	–	–	RNR2		-
2412	A/G	U-A ↓	0.81	0	3.62E-03	Inf	–	–	RNR2		-
2757	A/G		0.81	3	2.20E-01	5.2	–	–	RNR2	16S rRNA	-
6340	C/T		0.81	3	2.20E-01	5.2	N	D	COX1	p.T146I	Prostate Cancer
8459	A/G		0.81	0	2.00E-02	15.7	P	N	ATP8	p.N32D	-
8572	G/A		0.88	11	6.07E-01	1.1	P	N	ATP6	p.G16S	-
8854	G/A		0.88	3	2.20E-01	5.2	N	D	ATP6	p.A110T	-
8921	G/A		1.00	1	1.17E-01	15.6	P	D	ATP6	p.G132D	-
10086	A/G		0.94	26	1.00E+00	0.6	P	D	ND3	p.N10D	Hypertensive end-stage renal disease
11253	T/C		0.81	10	5.55E-01	1.3	N	D	ND4	p.I165T	LHON; PD
14225	C/T		0.75	0	6.03E-02	Inf	P	N	ND6	p.R150H	-
14337	C/T		1.00	3	3.53E-01	2.6	P	N	ND6	p.V113M	-
15402	C/T		0.94	2	1.97E-03	23.7	P	N	CYTB	p.T219I	-
15617	G/A		1.00	0	6.03E-02	Inf	P	N	CYTB	p.V291I	-

Cons – conservative; WC Base Pairs – Watson-Crick base pairs.

For the most critical findings in our study, the mutation frequencies of A5592G, T5628C and T681C are comparable with those of previous studies [15]. Nevertheless, for rare variants, we had ~80% power to detect a SNV of 0.5% frequency and effect size of 10 at 0.00125 (Table 6). A larger sample size is needed to detect additional variants with smaller effect size, lower frequency, or lower penetrance.

Conclusions

In our study, independent samples were used in the discovery and validation stages, indicating that the results are reliable. At the same time, 2 independent high-depth sequencing technologies were applied in these 2 stages, providing results that are more convincing. Although we found several statistically significant *loci*, multi-center, large population-based studies and functional studies must be performed to verify the pathogenicity of the *loci*.

Table 6. The power analysis result calculated by the Power Calculator based on the study design. The population disease prevalence was 0.05 and the significance level was 0.00125.

	Genotype relative risk	Power (185 cases vs. 2704+180 controls)
MAF=0.002	2.00	0.005
	3.00	0.015
	3.50	0.022
	4.00	0.032
	4.50	0.043
	5.00	0.056
	6.00	0.086
	7.00	0.123
	8.00	0.166
	9.00	0.212
MAF=0.005	2.00	0.01
	3.00	0.06
	3.50	0.10
	4.00	0.14
	4.50	0.19
	5.00	0.25
	6.00	0.38
	7.00	0.50
	8.00	0.61
	9.00	0.71
	10.00	0.79

	Genotype relative risk	Power (185 cases vs. 2704+180 controls)
MAF=0.01	2.00	0.04
	3.00	0.18
	3.50	0.29
	4.00	0.41
	4.50	0.52
	5.00	0.63
	6.00	0.79
	7.00	0.89
	8.00	0.95
	9.00	0.98
	10.00	0.99

In conclusion, we confirmed the increased rates of mtDNA heteroplasmy in the CAD group, and most of them were present at low frequencies. As previously reported, heteroplasmic variation in the mitochondrial genome varies with the degree of oxidative stress [39,40], and since the level of oxidative stress is a changeable risk factor, we speculate that it can be used for the prevention of CAD in the future. Our study increases the understanding of the pathogenesis of CAD and provides a basis for further study of mtDNA. Further studies are needed to verify the findings of this research.

Acknowledgements

We are grateful to the staff at the Department of Core Laboratory of Translational Medicine and Cardiology, Chinese People’s Liberation Army (PLA) General Hospital.

Conflict of interests

None.

References:

1. Galis ZS, Black JB, Skarlatos SI: National Heart, Lung, and Blood Institute and the translation of cardiovascular discoveries into therapeutic approaches. *Circ Res*, 2013; 112: 1212–18
2. Writing Group Members, Mozaffarian D, Benjamin EJ et al: Heart Disease and Stroke Statistics-2016 Update: A report from the American Heart Association. *Circulation*, 2016; 133: e38–360
3. Zhu KF, Wang YM, Zhu JZ et al: National prevalence of coronary heart disease and its relationship with human development index: A systematic review. *Eur J Prev Cardiol*, 2016; 23: 530–43
4. Zeng L, Talukdar HA, Koplev S et al: Contribution of gene regulatory networks to heritability of coronary artery disease. *J Am Coll Cardiol*, 2019; 73: 2946–57
5. Chen G, Levy D: Contributions of the Framingham Heart Study to the epidemiology of coronary heart disease. *JAMA Cardiol*, 2016; 1: 825–30
6. Jaiswal S, Natarajan P, Silver AJ et al: Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med*, 2017; 377: 111–21
7. Fuster JJ, Walsh K: Somatic mutations and clonal hematopoiesis. Unexpected potential new drivers of age-related cardiovascular disease. *Circ Res*, 2018; 122: 523–32
8. Brunel-Guitton C, Levtova A, Sasarman F: Mitochondrial diseases and cardiomyopathies. *Can J Cardiol*, 2015; 31: 1360–76
9. Chen X, He XY, Zhu C et al: Interaction between mitochondrial NADH dehydrogenase subunit-2 5178 C > A and clinical risk factors on the susceptibility of essential hypertension in Chinese population. *BMC Med Genet*, 2019; 20: 121
10. Maassen JA, t Hart LM, Ouwens DM: Lessons that can be learned from patients with diabetogenic mutations in mitochondrial DNA: Implications for common type 2 diabetes. *Curr Opin Clin Nutr Metab Care*, 2007; 10: 693–97
11. Saikia BB, Dubey SK, Shanmugam MK, Sundaresan P: Whole mitochondrial genome analysis in South Indian patients with Leber's hereditary optic neuropathy. *Mitochondrion*, 2017; 36: 21–28
12. Sobenin IA, Sazonova MA, Ivanova MM et al: Mutation C3256T of mitochondrial genome in white blood cells: Novel genetic marker of atherosclerosis and coronary heart disease. *PLoS One*, 2012; 7: e46573
13. Yu E, Calvert PA, Mercer JR et al: Mitochondrial DNA damage can promote atherosclerosis independently of reactive oxygen species through effects on smooth muscle cells and monocytes and correlates with higher-risk plaques in humans. *Circulation*, 2013; 128: 702–12
14. Sobenin IA, Sazonova MA, Postnov AY et al: Mitochondrial mutations are associated with atherosclerotic lesions in the human aorta. *Clin Dev Immunol*, 2012; 2012: 832464
15. Patel MR, Dehmer GJ, Hirshfeld JW et al., American College of Cardiology Foundation Appropriateness Criteria Task Force; Society for Cardiovascular Angiography and Interventions; Society of Thoracic Surgeons; American Association for Thoracic Surgery; American Heart Association, and the American Society of Nuclear Cardiology Endorsed by the American Society of Echocardiography; Heart Failure Society of America; Society of Cardiovascular Computed Tomography: ACCF/SCAI/STS/AATS/AHA/ASNC 2009 Appropriateness Criteria for Coronary Revascularization: A report by the American College of Cardiology Foundation Appropriateness Criteria Task Force, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons, American Association for Thoracic Surgery, American Heart Association, and the American Society of Nuclear Cardiology Endorsed by the American Society of Echocardiography, the Heart Failure Society of America, and the Society of Cardiovascular Computed Tomography. *J Am Coll Cardiol*, 2009; 53: 530–53
16. Qin Y, Xue L, Jiang P et al: Mitochondrial tRNA variants in Chinese subjects with coronary heart disease. *J Am Heart Assoc*, 2014; 3: e000437
17. Ingman M, Gyllenstein U: mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res*, 2006; 34: D749–51
18. Cai N, Li Y, Chang S et al: Genetic control over mtDNA and its relationship to major depressive disorder. *Curr Biol*, 2015; 25: 3170–77
19. Cai N, Chang S, Li Y et al: Molecular signatures of major depression. *Curr Biol*, 2015; 25: 1146–56
20. Ye K, Lu J, Ma F et al: Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci USA*, 2014; 111: 10654–59
21. Lowe TM, Chan PP: tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*, 2016; 44: W54–57
22. Ma LY, Chen WW, Gao RL et al: China cardiovascular diseases report 2018: An updated summary. *J Geriatr Cardiol*, 2020; 17: 1–8
23. Do R, Stitzel NO, Won HH et al: Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, 2015; 518: 102–6
24. Payne BA, Wilson IJ, Yu-Wai-Man P et al: Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*, 2013; 22: 384–90
25. Aryaman J, Johnston IG, Jones NS: Mitochondrial Heterogeneity. *Front Genet*, 2018; 9: 718
26. Han D, Dai P, Zhu Q et al: The mitochondrial tRNA(Ala) T5628C variant may have a modifying role in the phenotypic manifestation of the 12S rRNA C1494T mutation in a large Chinese family with hearing loss. *Biochem Biophys Res Commun*, 2007; 357: 554–60
27. Kelley SO, Steinberg SV, Schimmel P: Functional defects of pathogenic human mitochondrial tRNAs related to structural fragility. *Nat Struct Biol*, 2000; 7: 862–65
28. Liu Y, Li Y, Wang X, Ma Q et al: Mitochondrial tRNA mutations in Chinese hypertensive individuals. *Mitochondrion*, 2016; 28: 1–7
29. Tseng CC, Chen CJ, Yen JH et al: Next-generation sequencing profiling of mitochondrial genomes in gout. *Arthritis Res Ther*, 2018; 20: 137
30. Spagnolo M, Tomelleri G, Vattemi G et al: A new mutation in the mitochondrial tRNA(Ala) gene in a patient with ophthalmoplegia and dysphagia. *Neuromuscul Disord*, 2001; 11: 481–84
31. Joosten MM, Gansevoort RT, Mukamal KJ et al: Sodium excretion and risk of developing coronary heart disease. *Circulation*, 2014; 129: 1121–28
32. Jun JE, Lee YB, Lee SE et al: Elevated serum uric acid predicts the development of moderate coronary artery calcification independent of conventional cardiovascular risk factors. *Atherosclerosis*, 2018; 272: 233–39
33. Braga F, Pasqualetti S, Ferraro S, Panteghini M: Hyperuricemia as risk factor for coronary heart disease incidence and mortality in the general population: A systematic review and meta-analysis. *Clin Chem Lab Med*, 2016; 54: 7–15
34. Battaggia A, Scalisi A, Puccetti L: Hyperuricemia does not seem to be an independent risk factor for coronary heart disease. *Clin Chem Lab Med*, 2018; 56: e59–e62
35. Cortese F, Giordano P, Scicchitano P et al: Uric acid: From a biological advantage to a potential danger. A focus on cardiovascular effects. *Vasc Pharmacol*, 2019; 120: 106565
36. Cai W, Duan XM, Liu Y et al: Uric acid induces endothelial dysfunction by activating the HMGB1/RAGE signaling pathway. *Biomed Res Int*, 2017; 2017: 4391920
37. Spiga R, Marini MA, Mancuso E et al: Uric acid is associated with inflammatory biomarkers and induces inflammation via activating the NF-kappaB signaling pathway in HepG2 cells. *Arterioscler Thromb Vasc Biol*, 2017; 37: 1241–49
38. Alves RM, da Silva Costa SM, do Amor Divino Miranda PM et al: Analysis of mitochondrial alterations in Brazilian patients with sensorineural hearing loss using MALDI-TOF mass spectrometry. *BMC Med Genet*, 2016; 17: 41
39. Canter JA, Eshaghian A, Fessel J et al: Degree of heteroplasmy reflects oxidant damage in a large family with the mitochondrial DNA A8344G mutation. *Free Radic Biol Med*, 2005; 38: 678–83
40. Lee SR, Kim N, Noh YH et al: Mitochondrial DNA, mitochondrial dysfunction, and cardiac manifestations. *Front Biosci (Landmark Ed)*, 2017; 22: 1177–94