

MT-ND1 Variants and Protective Haplotypes in Egyptian Chronic Kidney Disease: An Integrated Targeted Mitochondrial Analysis

Burhan M. Nassar

Al-Azhar University

Metwally M. Montaser

Al-Azhar University

Manal Abd El-Salam

Al-Azhar University

Mahmoud A. Khalifa

Mahmoud.khalifa@azhar.edu.eg

Al-Azhar University

Research Article

Keywords: Chronic kidney disease, mitochondrial DNA, MT-ND1, heteroplasmy, protective haplotypes, 16S rRNA, Egyptian population

Posted Date: August 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10464584/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Chronic kidney disease (CKD) is a global health problem increasingly linked to mitochondrial dysfunction. However, mitochondrial DNA (mtDNA) variants remain poorly studied, particularly in understudied populations with renal disease.

Methods

We conducted a case-control study in an Egyptian cohort of 45 CKD patients and 14 controls, using targeted sequencing of selected mitochondrial genes spanning 600–4600 bp.

Results

Single-variant, burden, haplotype, and network analyses revealed two distinct mitochondrial signatures separating CKD patients from controls. CKD showed enrichment of MT-ND1 variants, especially rs397515509 (35.6% cases vs. 0% controls, OR = ∞ , $p = 0.0068$). Controls exhibited higher heteroplasmic variability in 16S rRNA ($\Delta = -28.9\%$), indicating greater mitochondrial tolerance. Protective haplotypes centered on rs28358576 reduced CKD risk, with larger haplotypes showing stronger cumulative protection (OR = 0.08–0.13). CKD patients also had higher total and homoplasmic variant burden.

Conclusion

Network analysis identified rs397515509 as a key variant linked to energy metabolism, oxidative stress, and apoptosis. Despite limited power and conservative bias ($\lambda = 0.355$), major associations showed large effect sizes, supporting mtDNA patterns as potential biomarkers of CKD susceptibility and resilience. The sequence data have been deposited in GenBank under accession numbers PZ760239–PZ760428.

Background

Chronic kidney disease (CKD) is a worldwide health concern, resulting in the gradual worsening of kidney function leading to kidney failure. In addition to known risk factors, recent studies identify bioenergetic failure at the cellular level and oxidative stress at the level of mitochondria as significant contributing factors to the underlying processes of kidney failure [1]. The kidney, and especially the tubular epithelium, which is energy-demanding, and high in mitochondria, is particularly vulnerable to this type of pathology. The oxidative stress caused by the overproduction of reactive oxygen species (ROS), the dysfunction of oxidative phosphorylation, and inflammation is considered an important feature in the development of CKD [2].

Considering the type of metabolic stress, the mitochondrial DNA (mtDNA) is a relevant target and possible point of reference. In comparison to the nuclear genome, mtDNA is found in greater amounts in a cell, does not have protective histones, and does not have a process that allows for the repair of oxidative damage. For this reason, mtDNA is rapidly and irreversibly subject to cumulative oxidative damage [3]. The result is a phenomenon called heteroplasmy (i.e., an intracellular population that consists of both wild-type and altered variant mtDNA). It is known that an increased burden of pathogenic mtDNA heteroplasmy is a detrimental feature of mitochondrial dysfunction, but the available data indicates that the burden of heteroplasmy is very likely to be multifactorial. Some heteroplasmic variants, particularly at low levels or in certain specific genomic configurations, may experience cellular selection and may even, in some instances, be a type of somatic adaptation to metabolic stress [4]. The intricacy illustrates the necessity of assessing both the amount and type of mtDNA heteroplasmy.

Despite the possibility, numerous significant shortcomings are present in researching mtDNA in the context of CKD. First and foremost, the majority of published literature has been on inherited haplogroups or single pathogenic point mutations, which depicts a rather unchanging or static portrayal of the genetics [5]. This methodology typically neglects the somatic mutational landscape, which is the mosaic of acquired variants formed due to chronic oxidative stress (damage). Additionally, a pronounced absence of research focused on the direction and pattern of variants is very notable. The biological consequences of mutations are not uniformly neutral in their impact. For example, certain research data suggests that insertional variants, which may be a direct result of replication slippage,

are disproportionately correlated to the accrual of age-associated heteroplasmy [5, 6]. There is a notable absence of a comprehensive synthesis that sorts variants based on type and juxtaposes them to clinical data syntheses.

Genomic investigations are plagued by a fundamental absence of diversity, particularly with the grossly inadequate representation of populations from the Middle East and North Africa [7]. The distinct genetic framework of the Egyptian populace is still largely uncharted in mitochondrial nephrology, thus, contributing to the restricted applicability of current research..

To address this knowledge gap in an understudied population, we focused on mitochondrial regions with established functional relevance to CKD pathophysiology. The selection of the mitochondrial 12S rRNA, 16S rRNA, ND1, ND2, and tRNA-Ile and tRNA-Val regions was driven by their critical roles in mitochondrial bioenergetics and proteostasis, functions directly relevant to CKD pathophysiology [8, 9]. The kidney, particularly the proximal tubule epithelium, is exquisitely dependent on oxidative phosphorylation (OXPHOS) for ATP generation. Subunits ND1 and ND2, as core Complex I proteins, are central to electron transfer and ATP synthesis [3, 9]. CKD's chronic bioenergetic failure and oxidative stress make variations in these genes prime candidates for disrupting Electron Transport Chain (ETC) efficiency, leading to decreased ATP and increased ROS production [2, 13, 14]. Similarly, 12S and 16S rRNA and tRNAs constitute the mitoribosome, indispensable for synthesizing ETC proteins [9, 10]. Compromised translation would amplify bioenergetic deficits, making this region a key susceptibility locus. These regions are recognized as mutational hotspots with high heteroplasmy propensity [3, 11]. Heteroplasmy introduces a threshold effect where the mutant load must exceed a critical level before biochemical deficits manifest, potentially explaining phenotypic variability in CKD [11].

To address the aforementioned gaps and lay the groundwork for mitochondrial genomics in Egyptian CKD, we undertook the first integrated mtDNA variance analysis in selected mitochondrial genes, incorporating variant type, heteroplasmy levels, haplotype embedding, and clinical correlations. This framework delineates mtDNA alteration architecture, facilitating biomarker discovery in this underserved population.

Materials and Methods

Study Design and Population

This case-control analysis study was performed at the Pediatric Nephrology & Hemodialysis Unit of Al Zahraa University Hospital, Al-Azhar University, Cairo, Egypt with 59 participants. For the End-Stage Renal Disease (ESRD Group (n = 45), the participants were children aged 2 to 18 years with an estimated glomerular filtration rate (eGFR) < 15 mL/min/1.73 m² who were on maintenance hemodialysis (3 times/week, 4 hours/session) for more than 3 months. As for Control Group (n = 14), the participants were children of the same age- and sex-matched with healthy children, with no prior records of illnesses involving the kidneys or chronic diseases, and were recruited during the periodic health evaluations. The sample size (45 cases, 14 controls) was pragmatically determined based on the available resources for this self-funded exploratory study. As previously reported effect sizes for pathogenic mtDNA variants in renal disease are large (OR > 3.0) [3, 11, 22], the cohort was considered adequate for detecting such large-magnitude effects.

Sample Collection

Blood samples were taken from the participants following standard procedures to collect blood from peripheral veins. The samples were collected into EDTA tubes, and they were kept at 4° until the time of extraction.

Ethics declarations

Informed oral consent was obtained from all parents of children involved in the study. Blood samples were collected under protocols approved by the Ethics Committee of the Faculty of Medicine for Girls, Al-Azhar University (FMG-IRB), Cairo, Egypt, under approval number 2026033157. Patients had the right to withdraw from the study at any time without providing reasons.

Assessment of DNA Quality and Extraction

Genomic DNA was extracted from peripheral blood leukocytes using standard commercial extraction kits. DNA concentration and quality were assessed spectrophotometrically by measuring absorbance at 260 nm and 280 nm.

Design of Primers and Selection of Targets for Mitochondrial DNA

Target gene selection was informed by an initial database interrogation using the BioMart tool within Ensembl (www.ensembl.org) to identify mitochondrial genes with known or predicted associations to renal disease phenotypes. This screening prioritized genes involved in oxidative phosphorylation, mitochondrial translation, and regions with documented mutational hotspots. Based on this interrogation, the analysis targeted specific regions of the mitochondrial genome spanning nucleotides 600–4600, encompassing the 12S rRNA, 16S rRNA, ND1, ND2, and tRNA-Ile and tRNA-Val regions. Detailed variant positions and genomic coordinates are provided in Supplementary Table S4. Primers were designed to cover multiple overlapping fragments for the target area (Table 1). The primer pairs flanked the regions of interest, with the tRNA-Phe primers covering the 12S rRNA and ND1 regions, and the tRNA-Leu primers covering the 16S rRNA, ND2, and tRNA-Ile/tRNA-Val regions.

Table 1
PCR primers for targeted mtDNA regions

Primer Name	Sequence (5' → 3')	Annealing Temp. (°C)	Product Size (bp)	mtDNA Coordinates of PCR Fragment	
				start	end
tRNA-Phe-F1	TCCCACTCCCATACTACTAATC	55	864	460	1323
tRNA-Phe-R1	CGTCTTTACGTGGGTACTTG				
tRNA-Phe-F2	GAGGAGCCTGTTCTGTAATC		800	1197	1996
tRNA-Phe-R2	GTTTGTGCGCTCTACCTATAAA				
tRNA-Leu-F1	GTAATCCAGGTCGGTTTCTATC		944	3082	4025
tRNA-Leu-R1	GTAGTGGTGAGGGTGTTTATT				
tRNA-Leu-F2	CTTTAACCTCTCCACCCTTATC		847	3783	4629
tRNA-Leu-R2	CTGTGGAACGAGGGTTTATT				

PCR Amplification

PCR amplification was performed in a total reaction volume of 25 µL, comprising 12.5 µL of GoTaq® Green Master Mix (Promega), 0.5 µM of both forward and reverse primers, and approximately 20–50 ng of target DNA. The thermal cycling profile included: initial denaturation at 95°C for 5 minutes, 30 cycles of denaturation (95°C for 30 seconds), annealing (30 seconds at 55°C), and extension (30 seconds at 72°C), with a final extension at 72°C for 7 minutes. PCR amplification success was confirmed by electrophoresis on a 1.5% agarose gel.

Sanger Sequencing

Purified PCR products were sequenced using the Sanger method and forward primers. The sequencing reactions were subsequently analyzed using capillary electrophoresis to generate chromatograms (AB1 trace files).

Assay Validation and Sequence Analysis

The optimized PCR protocol produced specific amplification products of the expected sizes for the four targeted mitochondrial DNA (mtDNA) fragments (tRNA-Phe1/2 and tRNA-Leu1/2, which flank the 12S rRNA, 16S rRNA, ND1, ND2, and tRNA-Ile and tRNA-Val regions) in all 59 patient and control samples.

Raw AB1 trace files corresponding to the mtDNA regions were processed. Sequence reads were trimmed to remove low-quality bases at both the 5' and 3' ends and manually proofread to ensure accuracy. Using Unipro UGENE v53.0 [12], reads were assembled into consensus sequences. Variant calling was performed by inspecting case and control chromatograms. Peak heights were quantified as

percentages of the total fluorescent signal at each nucleotide position. A complete list of all identified variants with their genomic coordinates and annotations is provided in Supplementary Table S1 and Supplementary Table S4. Heteroplasmy was assigned when the alternative allele peak signal was between 5% and 95% of the total signal and was distinguishable from background noise. Thresholds for background noise were determined from baseline signal between peaks in each chromatogram; only peaks above baseline were considered significant. Minor allele peaks with a signal below 5% and indistinguishable from background were excluded from analysis. For each heteroplasmic position, major and minor allele peak ratios were recorded. To limit subjective analysis, a secondary reviewer independently assessed all chromatograms, and differences were resolved by group agreement. This methodology enabled consistent genotype calls across all samples and established parameters for further comparative analyses between case and control groups. A complete list of all identified mitochondrial variants with their allele frequencies and annotations is provided in Supplementary Table S1.

Statistical Analysis for Association with CKD

Statistical analysis included odds ratios and 95% confidence intervals, comparison of allele frequency, analysis of mitochondrial burden, correlation analysis, haplotype analysis, and network analysis.

Logistic regression was used to test the association of each defined protective haplotype with CKD status while controlling for covariates including age, sex, and principal components of genetic ancestry. The analysis was executed under a categorical coding schema with age, sex, and genetic ancestry components as covariates. The predictor variable was the number of haplotype copies (0, 1, or 2: 0 = homoplasmic reference, 1 = homoplasmic alternative, 2 = heteroplasmic with both alleles present). The effect size for each haplotype was represented by a Beta coefficient with standard error (SE). The Beta coefficient was subsequently used to derive the Odds Ratio (OR) and corresponding 95% confidence interval (CI).

All statistical analyses were performed using Python (version 3.13.0) on Linux (Ubuntu 22.04.3 LTS). The analysis used statistical and genetic packages including PLINK2 and statsmodels (v0.14.5), scipy (v1.16.3), numpy (v2.3.1), pandas (v2.3.0), matplotlib (v3.10.0), and seaborn (v0.13.2). A statistical significance threshold of $p < 0.05$ was applied in all analyses. The relationship between the OR and the protective effect of the haplotype size (number of constituent variants) was represented descriptively to analyze the presence of synergistic interaction.

All sequence data have been deposited in GenBank under accession numbers PZ760239–PZ760428. Complete variant calling results and statistical outputs are available in Supplementary Tables S1–S4.

Results

A sample from a small cohort study along with Sanger sequencing does not allow for all mitochondrial positions to be covered for every sample. However, coverage was found to be enough to call reliable variants along with tight quality control of chromatogram reads (see Methods). A complete list of all identified mitochondrial variants with their allele frequencies, genomic positions, and annotations is provided in Supplementary Table S1. Detailed variant positions and genomic coordinates are provided in Supplementary Table S4. With such a small sample size (45 cases, 14 controls), all samples were kept to increase exploratory capacity, meaning all findings should be interpreted as hypothesis-generating..

An assessment of correlations between all variants was performed to study dependencies prior to the association analyses. As seen in (Fig. 1), the vast majority of variants are independent, suggesting associations are likely not driven by correlated variants. This quality supports the burden and association results. A small number of pairs had correlations, suggesting that this pattern potentially reflects a shared mitochondrial haplotype, and some of the genetic influences are more likely to arise from interaction of multiple variants than a single variant acting alone. Full haplotype analysis results, including all variant combinations tested, are presented in Supplementary Table S2.

As mitochondrial DNA variants potentially impact the relationships with chronic kidney disease (CKD), the variants that were detected were analyzed in reference to each disease state individually. This analysis is depicted in (Fig. 2). In the Manhattan plot, the y-axis is the $-\log_{10}(p\text{-values})$ for the given association. A positive value indicates a statistically significant association, and the higher the positive value, the stronger the association. Each peak in the plot is a different mitochondrial variant with statistically significant association to CKD, and the height of each peak denotes the strength of the association. A number of variants reached a nominal level of statistical significance ($p < 0.05$), indicating that these variants may be contributing factors to the susceptibility to chronic kidney

disease (CKD). The most significant variant is rs397515509 located in the ND1 gene, for which the significance remains under a stricter statistical testing threshold ($p < 0.01$). Therefore, ND1 gene variant rs397515509 is the most likely candidate for a CKD contributing locus. Another ND1 gene variant, rs878853034, is also significantly associated (above the threshold) with CKD. Furthermore, variants rs28358576 and mt1961 located in the 16S rRNA gene were also significant, while variant rs878907222 is almost statistically significant. In conclusion, the ND1 gene and 16S rRNA variants provided a focused set of variants for subsequent haplotype and other integrative analyses, which may influence the risk factor list for CKD.

We conducted a QQ-plot analysis (Fig. 3) to assess the reliability of the association signals elucidated in Fig. 2 and to ascertain that the identified loci candidates in the mitochondria were not a result of spurious statistically driven artifacts. This study showed a conservatively distributed dataset with a genomic inflation factor (λ) of 0.355 well below the accepted false positive range of 0.95–1.05. This conservatively distributed dataset implies that there is a modest control sample size ($n = 14$) that reduces statistical power while minimizing false associations. In the face of excessive conservatism and bias, the controlled association analysis did not necessarily identify the significant variants, loci, and alleles in the association analysis while controlling the testing environment. Furthermore, in compliance with this explanation, the tested observed values remained consistently below the anticipated diagonal, and with the inflation, $1.11\times$ the nominally significant ($p < 0.05$) variants were located in the dataset. This explains the conservatively distributed dataset and supports the identified association signals.

The analysis in (Fig. 4) demonstrates the divergent distribution of selected mitochondrial variants with respect to CKD affected and unaffected individuals and controlling for variants with respect to the CKD cohort and CKD unaffected individuals. Different patterns became apparent in the distribution of variants. The variant rs397515509 is present in CKD cases only and absent in CKD controls. Therefore, it provides a definitive link in the distribution of the variant to the presence of the disease. Variants rs878853034, and rs397515509, exhibit the same trend, with CKD case controls having a comparatively higher distribution. Conversely, the variants mt1961, rs28358576, and rs1569483815, exhibit a higher frequency of distribution in CKD control cases, providing evidence for a disease-associated genetic background in cases, or a potentially protective set of variants in controls. Overall, these patterns exemplify the presence of a set of mitochondrial variants that provide differential explanation for the groups identified and, therefore, warrant attention in further statistical analysis and downstream combinatorial analysis.

Building on previously noted differences relating to certain mitochondrial variants among CKD patients and controls, overall mitochondrial variant profiles were analyzed using box-plot comparisons (Fig. 5). CKD patients demonstrate an elevated overall mitochondrial mutational, exemplified by the higher median overall variants the CKD group carried as compared to controls. Homoplasmic variants were consistent with more mutation fixation among CKD patients. Whereas, controls had a higher median level of heteroplasmic variants. The weighted burden score analysis CKD patients clustered at higher distinctly positive values while controls clustered at lower and predominantly negative values (Table 3). Collectively, these findings demonstrate the extent of divergence of CKD and non-CKD groups is not only individual loci focused, but also at the global mitochondrial mutational burden.

Table 3
Comparison of mitochondrial DNA variant burden between cases and controls.

Burden_Score	Cases		Controls		T-test P-Value	Odds Ratio	OR 95 CI	
	Mean	SD	Mean	SD			Lower	Upper
Total Variant Count	9.2	3.14	8.29	1.94	0.199	1.13	0.9	1.42
Homoplasmic Count	4.2	2.57	2.07	1.27	0.0001	1.63	1.13	2.34
Heteroplasmic Count	5	2.37	6.21	2.36	0.1073	0.82	0.63	1.05
Weighted Burden	2.9	4.57	-4.61	3.46	0	3.2	1.55	6.63

Analysis of mitochondrial haplotypes showed that some variant combinations correlated with significantly lower odds of chronic kidney disease (CKD); thus, there could be a protective effect arising from a higher number of multiple variants present, rather than a single variant affecting the overall outcome. The correlation pattern illustrated in (Fig. 1) showed that most variants in the study were spread single across individuals, but a few were found to co-occur, indicating some individuals may share a common mitochondrial genetic background. This prompted us to explore whether sets of variants rather than single loci could provide a more reasonable explanation to the disease pattern. This strategy allowed the identification of several protective haplotypes, where the most prominent effect was associated with the three-variant haplotype mt1563 - mt4018 - rs28358576, which was linked to a significant decrease in

CKD risk (OR = 0.08, 95% CI: 0.008–0.881, P = 0.0389) (Table 2). This haplotype was detected in a small portion of the cohort (6.8%), but it was the most protective of all combinations that were evaluated, supporting the hypothesis that risk for mitochondrial disorders may be shaped by the overall genetic background in which risk variants are embedded. Full haplotype analysis results, including all variant combinations tested, are presented in Supplementary Table S2.

Table 2
Association analysis of protective mitochondrial variant haplotypes with CKD risk.

Haplotype	Size	Beta	SE	P_Value	Odds_Ratio	OR_CI_Lower	OR_CI_Upper	Frequency
rs28358576	1	-2.051	0.817	0.0121	0.13	0.026	0.638	13.6
mt1563_rs28358576	2	-2.152	0.934	0.0213	0.12	0.019	0.726	10.2
mt4018_rs28358576	2	-2.152	0.934	0.0213	0.12	0.019	0.726	10.2
mt1563_mt4018_rs28358576	3	-2.485	1.203	0.0389	0.08	0.008	0.881	6.8
mt1961_mt4018	2	-1.723	0.841	0.0405	0.18	0.034	0.928	11.9

The composition of specific haplotypes was shown to influence the degree of protection conferred with respect to developing chronic kidney disease (CKD). The single polymorphism rs28358576 exhibited reduced odds of developing CKD (OR = 0.13, P = 0.0121) (Fig. 6). When rs28358576 was combined with mitochondrial variants mt1563 or mt4018 to form two-variant haplotypes, the degree of protection was marginally improved (OR = 0.12 for both pairs). The inclusion of both mitochondrial variants for a three-variant haplotype provided the greatest degree of protection (OR = 0.08), suggesting a cumulative protective effect. This tendency is corroborated by the graphical representation of the size of the haplotype compared to the associated odds ratio, where larger combinations of variants correspond to lower values of the odds ratio.

Additionally, the mt1961_mt4018 haplotype exhibited a significant protective association (OR = 0.18, P = 0.0405), suggesting the possibility of protective interactions within the mitochondrial genome separate of the polymorphism rs28358576. Although the haplotypes with significant results displayed protective associations (OR < 1), they have wide confidence intervals, a consequence of the carrier frequency in the population, which was shown to be low (approximately 6.8–13.6%), indicating the unfavorable nature of these genetic variants.

Discussion

This research shows that the genetic composition of mitochondria relatively differentiates CKD patients from healthy individuals, showing both the presence of CKD-specific variants as well as that control individuals exhibit greater tolerance for heteroplasmy. One of the most relevant examples of this is the complete CKD-specific enrichment of the MT-ND1 variant rs397515509, which was present in CKD cases only and absent in all 14 controls (complete segregation), suggesting that certain mitochondrial mutations may increase the risk of developing renal disease, especially considering the mitochondrial dysfunction and oxidative stress that underlie the progression of renal disease [13–16]. These associations were notable at low levels of heteroplasmy, which suggests that some variants may exert a detrimental effect regardless of the presence; this is consistent with many of the non-linear threshold theories of mitochondrial dysfunction [3, 17–18].

In contrast, the notable and enduring heteroplasmy across several genes in control subjects indicates a biological mechanism whereby healthy individuals may cope with, or possibly benefit from, increased mitochondrial genetic diversity. The pronounced differences in heteroplasmy levels of 16S rRNA and MT-ND1 reinforce the concept of mitochondrial resilience in control subjects. Specifically, controls exhibited a higher median level of heteroplasmic variants, whereas CKD patients carried more homoplasmic variants (Fig. 5). This suggests that mtDNA should be viewed integratively or cumulatively as a marker of metabolic or physiological stress, rather than a binary indicator of disease presence [3, 17–18]. These findings align with reports of somatic mitochondrial variation within the nucleus that occurs without correlating pathology, potentially indicating an adaptive response [3, 17]. Of particular interest to this study are the unique case-control frequency distributions of individual variants, which challenge conventional dualistic frameworks and suggest that these specific mutations engage with the pathological milieu differently from others. This is consistent with previous renal disease studies reporting that somatic mtDNA mutations possess a wide range of pathological relevance [3, 17, 19].

In our study of mitochondrial DNA variants in chronic kidney disease (CKD), we found a number of interesting associations in the mitochondrial region 600-4600bp which contains important functional elements such as the ribosomal RNA genes, tRNA, and parts of the NADH dehydrogenase and cytochrome c oxidase subunits. Two notable variants in the Manhattan plot (Fig. 2) are rs397515509, which was present in CKD cases only and absent in controls (complete segregation; OR not estimable) and rs28358576, which is protective and showed an OR of 0.13 ($p = 0.0121$). Other interesting variants are rs878853034 and mt1961, both of which reached nominal significance, though exact p-values are not reported due to the exploratory nature of the study.

The low genomic inflation factor in this study ($\lambda = 0.355$) reflects conservative statistical behavior rather than systematic bias such as population stratification, as expected under genomic control and mixed-model frameworks [20-21]. This conservatism is largely explained by our modest sample size (45 cases, 14 controls). Single-variant tests are typically underpowered for low-frequency variants, leading to conservative p-value distributions [22]. Moreover, regularized models in high-dimensional frameworks prioritize stability and noise suppression over power, which also yields $\lambda < 1$ [23]. Consistent with this conservative setting, the number of variants reaching nominal significance ($p < 0.05$) was only 1.11 times the number expected by chance alone – indicating that, despite the low power, true associations are still detectable but not inflated.

We implemented an integrated genetic network (Fig. 7) to evaluate functional pathways, burden scores, and identified mitochondrial variants. The network connectivity and pathway data used for the integrated genetic network (Fig. 7) are detailed in Supplementary Table S3. This network provides additional insight into how individual variants and burden measures contribute to CKD pathogenesis. As depicted, it demonstrates that mitochondrial genetic influence in CKD is systems-integrated and driven by multivariant functional interactions that emphasize bioenergetic control.

Although referring to unique mutation-specific impacts, the network's structured topology illustrates that the disease-related variants uniformly align with the interconnected mitochondrial systems that regulate energy production, redox equilibrium, and the signaling pathways of cellular survival [24–26]. The network connectivity and pathway data used for the integrated genetic network (Fig. 7) are detailed in Supplementary Table S3.

The strong position of rs397515509, which links energy metabolism, oxidative stress, and the overall variant burden score, implies a mechanistic involvement in the disruption of oxidative phosphorylation (OXPHOS) efficiency. In our cohort, rs397515509 showed complete segregation (cases only), reinforcing its pathogenic role. An OXPHOS disruption is posited to reduce ATP production, while simultaneously increasing electronic leakage (from the chain) which causes a higher production of reactive oxygen species and a redox imbalance in the mitochondria. This is noteworthy in renal tissue which is extremely reliant on ATP from mitochondria especially for the transport and filtration maintenance of the tubules. Such an energetic failure and oxidative stress intensification is a plausible biological cause of progressive cellular injury. The multilayered connectivity of rs878853034 to apoptosis, mitochondrial complexes, and weighted burden measures reinforces the assumption that a perturbation in the respiratory chain may cause a knock-on effect of apoptotic signaling, which is in agreement with the cleavage of the inner membrane and subsequent cell death seen in the progression of CKD.

The links of mt1961 to energy metabolism and mitochondrial biogenesis broadens this understanding upstream, suggesting possible involvement in regulation of mitochondrial turnover, and compensatory biogenesis responses to bioenergetic deficiency [14, 27–31]. On the other hand, the protective variants have significantly restricted interaction architecture. rs28358576 is mostly restricted to mitochondrial biogenesis and indirect burden links, while rs28358576 has a focus on the regulation of oxidative stress with limited network high centrality. This variant alone reduced CKD odds by 87% (OR = 0.13, $p = 0.0121$; see Fig. 6). When rs28358576 was combined with mt1563 and mt4018 into a three-variant haplotype, the protective effect increased (OR = 0.08, $p = 0.0389$). The mt1961_mt4018 haplotype also showed protection (OR = 0.18, $p = 0.0405$). The additional protective signal focusing on complex III suggests that localized stabilization of specific components of the respiratory chain may keep electron transport uncoupled, thereby limiting ROS and the OXPHOS collapse that is typical of mitochondrial renal dysfunction [24, 26–27].

The emergence of the Variant Burden Score and Weighted Burden Score as central dominant hubs offer additional understanding. Their broad centrality across variants and functional modules suggests that cumulative mitochondrial genomic load may adjust, in a quantitative fashion, the threshold at which loss of oxidative phosphorylation activity engages biologically. In our data, CKD patients exhibited distinctly higher positive weighted burden scores, whereas controls clustered at lower, predominantly negative values (Fig. 5). This hub-centered convergence supports a threshold model of mitochondrial dysfunction, wherein distributed genomic changes cumulatively erode respiratory reserve capacity, ultimately triggering systemic mitochondrial failure [32–34].

Considering the network's hierarchical density, functional convergence, and burden-centered topology, there is strong mechanistic rationale to conclude that the mitochondrial components of CKD are, in essence, cumulative oxidative phosphorylation instability, redox, and bioenergetic insufficiency. This systems-level analysis, in addition to the cumulative plausibility of the biological rationale of the genetic signals, positions cumulative mitochondrial burden coupled with certain high-impact variants as potential integrative biomarkers of renal mitochondrial vulnerability [14, 25, 27]. This was evident in the box-plot comparisons (Fig. 5), where CKD patients carried a higher median overall mutational burden than controls.

The distribution of association signals in the functionally diverse mitochondrial regions suggests a variety of possible pathogenic mechanisms involved in CKD. Concentration of signals in the ribosomal RNA regions suggests probable shortcomings in the translation of mitochondria, while variations in the RNA protein coding regions can disrupt the integrity of the Electron Transport Chain (ETC). Furthermore, complete segregation of groups for rs397515509 suggests the possibility of a mutation that has significant biological impact. On the other hand, the renoprotective effect of the aforementioned variant indicates that certain mitochondrial haplotypes, due to the presence of the variant, can bestow protection against kidney injury. This may arise from an increased antioxidative capability or a more efficient ATP production mechanism. The three-variant haplotype mt1563-mt4018-rs28358576 yielded OR = 0.08 (95% CI: 0.008–0.881, $p = 0.0389$). The mt1961_mt4018 haplotype gave OR = 0.18 ($p = 0.0405$).

Our findings support a dual model of mitochondrial contribution to CKD, wherein certain variants (e.g., rs397515509) act as pathogenic drivers, while others (e.g., rs28358576) confer protection through enhanced heteroplasmy tolerance. However, a descriptive lack of statistical significance for some comparisons should motivate hypothesis-generating research, particularly for understudied populations such as those in the Middle East and North Africa [7,35-36] – a common challenge in emerging fields. The wide confidence intervals (e.g., 95% CI 0.008–0.881 for the three-variant haplotype) and low carrier frequencies (6.8-13.6%) underscore the need for larger cohorts. With only 45 cases and 14 controls, our estimates are hypothesis-generating. We also introduce the concept of prognostic heteroplasmy, whereby the cumulative burden of significant mitochondrial variation may alter disease trajectory and tolerance thresholds. To validate these phenomena, future studies require sufficiently large cohorts paired with functional measures. Such research will clarify whether heteroplasmy patterns can provide positive prognostic information and elucidate the broader potential of mitochondrial genetics in CKD pathogenesis.

Several of the limitations of our study stem from the design itself, and we have attempted to address these as best as possible under the circumstances. The small sample size (45 cases and 14 controls), especially of the controls, will always restrict statistical capacity to the extent that no affirmative conclusion can ever be made for smaller associations. The rs397515509, most likely, due to complete segregation, would most plausibly have a biological reason for displaying an odds ratio that is not estimable (complete segregation artifact of small sample size), which is also an explanation for the necessity of a larger study to be conducted. Further, the study of the 600-4600bp of the mitochondria was a deliberate choice to contain the most significant functional elements, but this has, of course, also omitted the possibility of studying other regions of the mitochondrial DNA. In addition, Sanger sequencing did not cover all mitochondrial positions for every sample, as noted in the Results. This conservative bias ($\lambda = 0.355$) does, however, validate that there is unlikely to be a significant number of false positives to the significant results of the study, which could support the basis of a study with a larger sample size.

To conclude, our integrated analysis, for the first time, with CKD's ascertained analytical rigor identifies credible associations at the level of mitochondrial genomics. It is somewhat paradoxical, but the nature of our analysis strengthens our confidence in rs397515509 as a CKD-specific risk variant, and rs28358576 as a central protective allele (OR = 0.13), both of which merit additional research. As protective haplotypes and the cumulative mutational burden crossed a powerful discriminatory threshold between cases and controls (as shown by the weighted burden score separation – positive vs negative values in Fig. 5; the strongest protective haplotype reduced CKD risk by 92%, OR = 0.08), we supported a threshold model in which, through synergistic combined effect, interactive genetic architecture modulates the risk of diseases. These results further corroborate the role of mitochondrial genomics in the pathophysiology of CKD and elucidate specific individual variants, multi-locus combinations, and protective haplotypes as targets for further research. We recommend that future studies quantify the bioenergetic consequences of the variants through cellular models, and assess both CKD-associated nuclear and genomic mitochondrial variants to provide a basis for targeted therapeutics in CKD based on an individual's mitochondrial genomic profile, and risk and protective haplotypes. Furthermore, we suggest the study of nDNA-mtDNA interactions for further elucidation of the therapeutic potential. Future studies could employ machine learning approaches to further characterize the mitochondrial genetic architecture of CKD, potentially identifying additional predictive patterns.

Abbreviations

Chronic kidney disease
CKD
Reactive oxygen species
ROS
Mitochondrial DNA
mtDNA
End-Stage Renal Disease
ESRD
oxidative phosphorylation
OXPHOS

Declarations

Ethics approval and consent to participate

This study was performed in accordance with the ethical principles of the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of the Faculty of Medicine for Girls, Al-Azhar University (FMG-IRB), Cairo, Egypt (approval number: 2026033157). Informed oral consent was obtained from all parents of children involved in the study. Patients had the right to withdraw from the study at any time without providing reasons.

Consent of publication

Not applicable

Competing interests:

The authors declare no competing interests.

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was self-funded as part of a PhD project.

Author Contribution

B.M.N. performed experimental laboratory work, conducted primary data collection, and assisted in manuscript drafting and editing. M.M.M. served as the primary PhD supervisor and provided final validation of the complete manuscript. M.A.E. supervised clinical sample collection at the university hospital, managed all clinical data acquisition, provided pediatric medical expertise, and contributed to manuscript validation. M.A.K. conceived and designed the study, formulated the research protocol, performed all bioinformatics and statistical analyses (including sequence data processing, variant calling, and generation of all figures), wrote the full manuscript, and directly supervised the student's progress throughout the study. All authors reviewed and approved the final manuscript.

Acknowledgements:

Not applicable.

Data Availability

The datasets generated and/or analysed during the current study are included in this published article and its supplementary information files (Supplementary Tables S1-S4).

References

1. Srivastava A, Rath SK. Mitochondrial dysfunction and oxidative stress: role in chronic kidney disease. *Life Sci.* 2023;319:121432.
2. Ishimoto Y, Inagi R. Mitochondria: a therapeutic target in acute kidney injury. *Nephrol Dial Transpl.* 2016;31:1062–9.
3. Wallace DC. Mitochondrial genetic medicine. *Nat Genet.* 2018;50:1642–9.
4. Stewart JB, Chinnery PF. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nat Rev Genet.* 2021;22:106–18.
5. Govers LP, et al. Mitochondrial DNA mutations in renal disease: an overview. *Pediatr Nephrol.* 2021;36:9–17.
6. Dabravolski SA, et al. The role of mitochondrial DNA mutations in cardiovascular diseases. *Int J Mol Sci.* 2022;23:952.
7. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell.* 2019;177:26–31.
8. Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. *Nature.* 1981;290:457–65.
9. Petropoulou A, Kypraios N, Rizopoulou D, et al. Mitochondrial tRNA-derived diseases. *Int J Mol Sci.* 2025;26:12023.
10. O'Brien TW. Properties of human mitochondrial ribosomes. *Biochem J.* 2003;374:595–607.
11. Gorman GS, Chinnery PF, DiMauro S, et al. Mitochondrial diseases. *Nat Rev Dis Primers.* 2016;2:16080.
12. Rose R, Golosova O, Sukhomlinov D, Tiunov A, Prosperi M. Flexible design of multiple metagenomics classification pipelines with UGENE. *Bioinformatics.* 2019;35:1963–5.
13. Che R, Yuan Y, Huang S, Zhang A. Mitochondrial dysfunction in the pathophysiology of renal diseases. *Am J Physiol Ren Physiol.* 2014;306:F367–378.
14. Fontecha-Barriuso M, et al. Tubular mitochondrial dysfunction, oxidative stress, and progression of chronic kidney disease. *Antioxidants.* 2022;11:1356.
15. Kanazashi Y, et al. Mitochondrial DNA variants at low-level heteroplasmy and decreased copy numbers in chronic kidney disease (CKD) tissues with kidney cancer. *Int J Mol Sci.* 2023;24:17212.
16. Lin X, Zhou Y, Xue L. Mitochondrial complex I subunit MT-ND1 mutations affect disease progression. *Heliyon.* 2024;10:e28808.
17. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet.* 2015;16:530–42.
18. Picard M et al. Progressive increase in mtDNA 3243A > G heteroplasmy causes abrupt transcriptional reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* 111, E4033–4042 (2014).
19. Feng J, Chen Z, Liang W, Wei Z, Ding G. Roles of mitochondrial DNA damage in kidney diseases: a new biomarker. *Int J Mol Sci.* 2022;23:15166.
20. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999;55:997–1004.
21. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet.* 2014;46:100–6.
22. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95:5–23.
23. Morris JS. Functional regression. *Annu Rev Stat Appl.* 2015;2:321–59.
24. Ho HJ, Shirakawa H. Oxidative stress and mitochondrial dysfunction in chronic kidney disease. *Cells.* 2023;12:88.
25. Chen J, Zhou Q, Su L, Ni L. Mitochondrial dysfunction: the hidden catalyst in chronic kidney disease progression. *Ren Fail.* 2025;47:2506812.
26. Jia X, Zhu L, Zhu Q, Zhang J. The role of mitochondrial dysfunction in kidney injury and disease. *Autoimmun Rev.* 2024;23:103576.
27. Tirichen H, et al. Mitochondrial reactive oxygen species and their contribution in chronic kidney disease progression through oxidative stress. *Front Physiol.* 2021;12:627837.
28. Hoogstraten CA, Hoenderop JG, de Baaij JH. F. Mitochondrial dysfunction in kidney tubulopathies. *Annu Rev Physiol.* 2024;86:379–403.
29. Bernardi P, et al. Identity, structure, and function of the mitochondrial permeability transition pore: controversies, consensus, recent advances, and future directions. *Cell Death Differ.* 2023;30:1869–85.

30. Chang LY, Chao YL, Chiu CC, Chen PL, Lin HYH. Mitochondrial signaling, the mechanisms of AKI-to-CKD transition and potential treatment targets. *Int J Mol Sci.* 2024;25:1518.
31. Takasu M, et al. The role of mitochondria in diabetic kidney disease and potential therapeutic targets. *Kidney Int Rep.* 2025;10:328–42.
32. Rossignol R, et al. Mitochondrial threshold effects. *Biochem J.* 2003;370:751–62.
33. Tuppen HAL, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta Bioenerg.* 2010;1797:113–28.
34. Aryaman J, Johnston IG, Jones NS. Mitochondrial DNA density homeostasis accounts for a threshold effect in a cybrid model of a human mitochondrial disease. *Biochem J.* 2017;474:4019–34.
35. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology.* 2008;19:640–8.
36. Button KS, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* 2013;14:365–76.

Figures

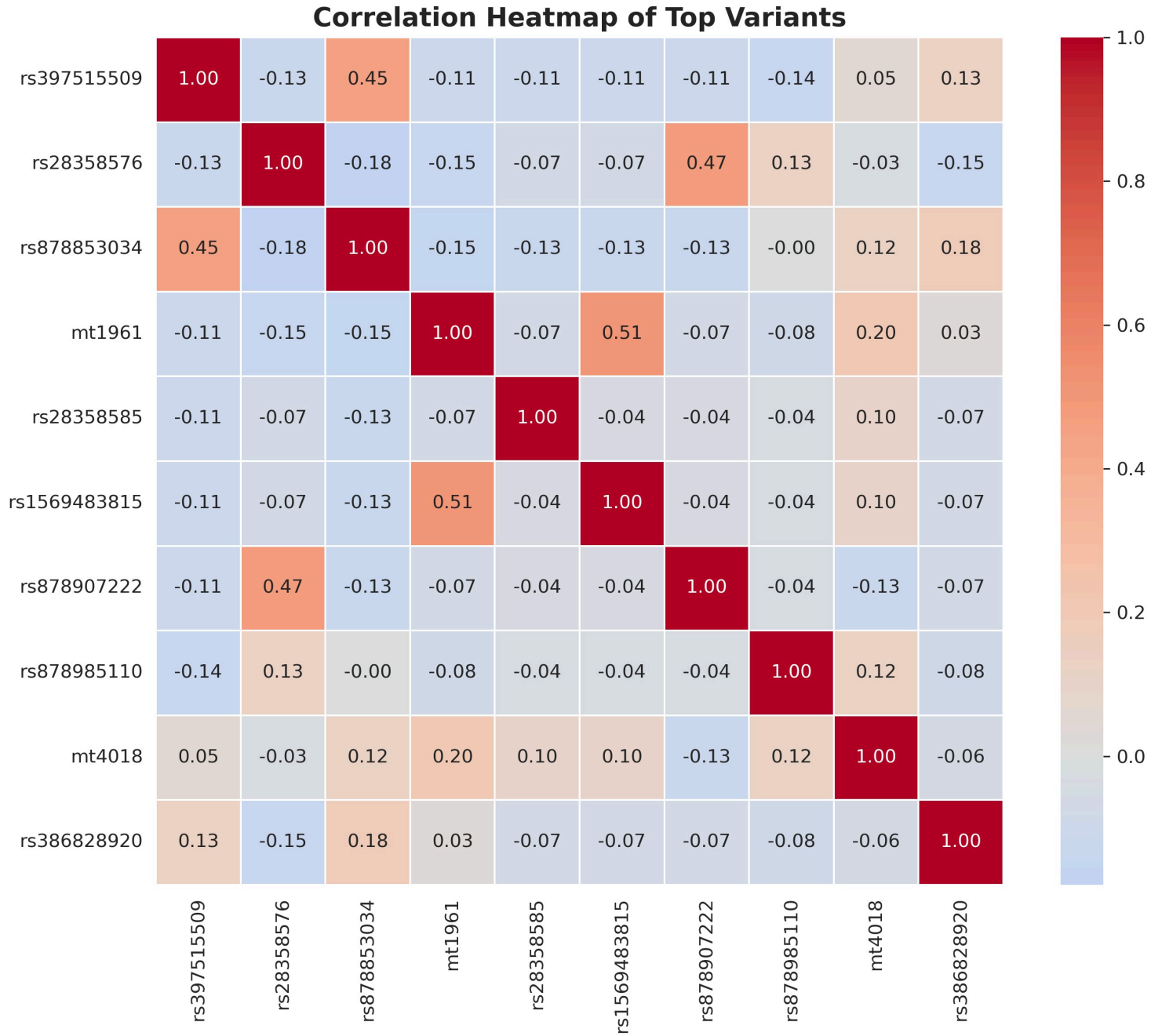


Figure 1

Correlation analysis of mitochondrial DNA variants. Correlation heatmap of top variants. Pairwise correlation analysis of all identified mtDNA variants in the study cohort. The color scale represents the correlation coefficient (r), with positive correlations in red and light blue indicating negative correlations. The vast majority of variants showed independent segregation, with a few small clusters suggesting shared mitochondrial haplotypes.

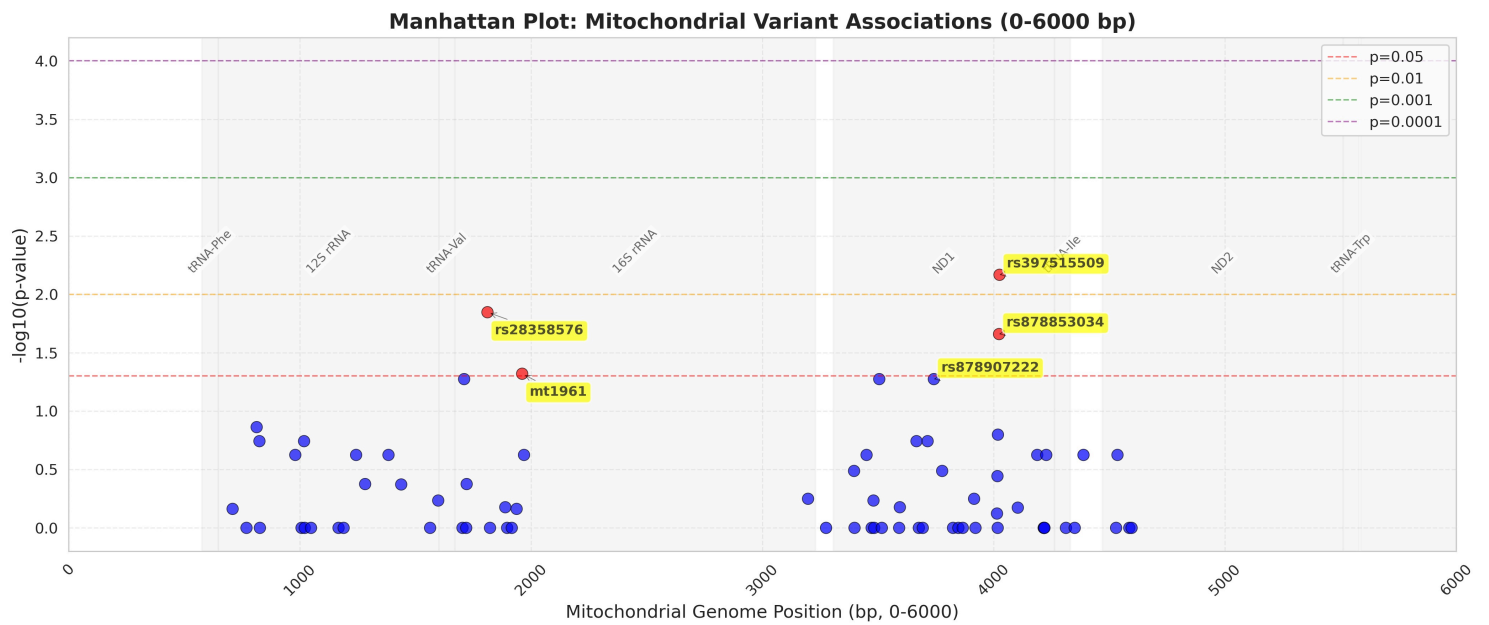


Figure 2

Manhattan plot of mitochondrial variant associations (600–4600 bp). The y-axis represents $-\log_{10}(p\text{-values})$, while the x-axis indicates mitochondrial genome position. Horizontal dashed lines denote significance thresholds at $p = 0.05$, $p = 0.01$, $p = 0.001$, and $p = 0.0001$. Highlighted points indicate variants exceeding nominal significance levels, including rs397515509 (MT-ND1, $p < 0.01$), rs878853034 (MT-ND1), rs28358576 (16S rRNA), and mt1961 (16S rRNA). The top of the plot shows a statement: "Manhattan Plot: Mitochondrial Variants Associations (600–4600 bp)".

QQ Plot: Single Variant Association Analysis

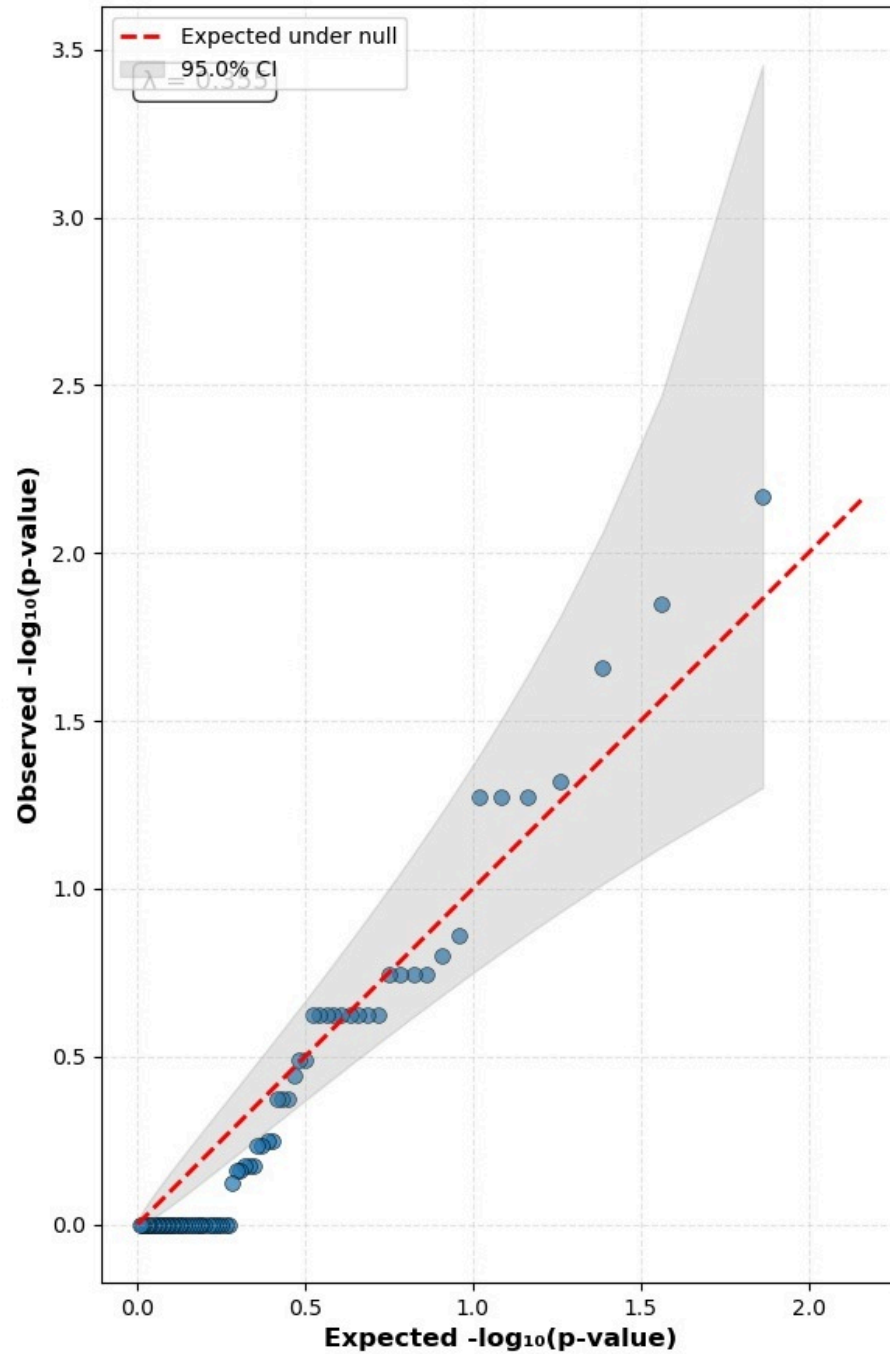


Figure 3

Quantile-quantile (QQ) plot of observed vs. expected p-values ($-\log_{10}(\text{p-values})$). The QQ plot assesses the distribution of association test statistics. The diagonal line represents the expected distribution under the null hypothesis. The observed points (blue circles) fall below the diagonal, indicating conservative statistical behavior (genomic inflation factor $\lambda = 0.355$), consistent with the modest sample size and absence of population stratification bias.

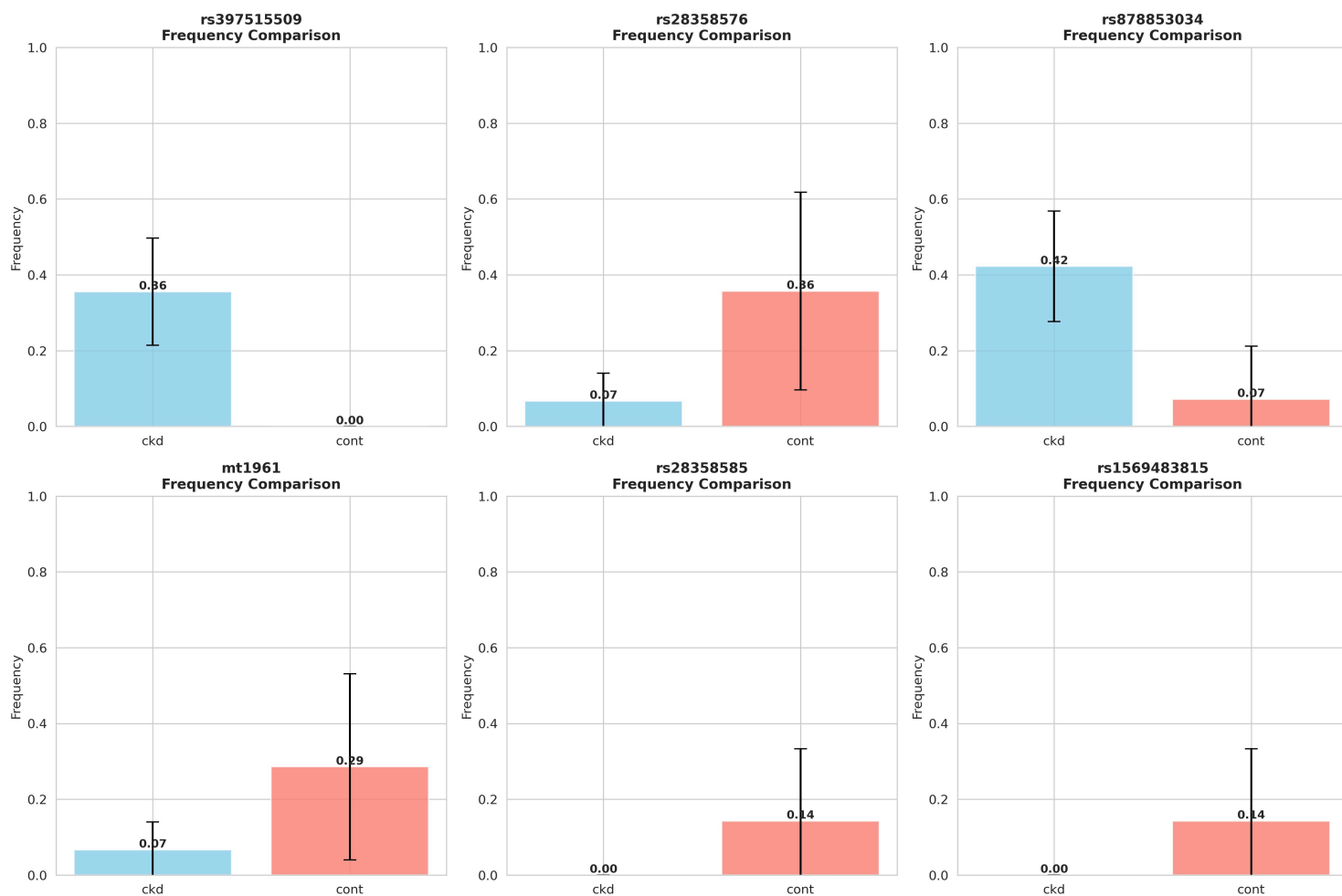


Figure 4

Distribution of selected mitochondrial variants by disease status. The bar chart shows the frequency of each variant in CKD patients (blue) and controls (orange). Variants rs397515509 and rs878853034 were enriched in CKD cases, while mt1961 and rs28358576 showed higher frequency in controls, suggesting potential protective effects.

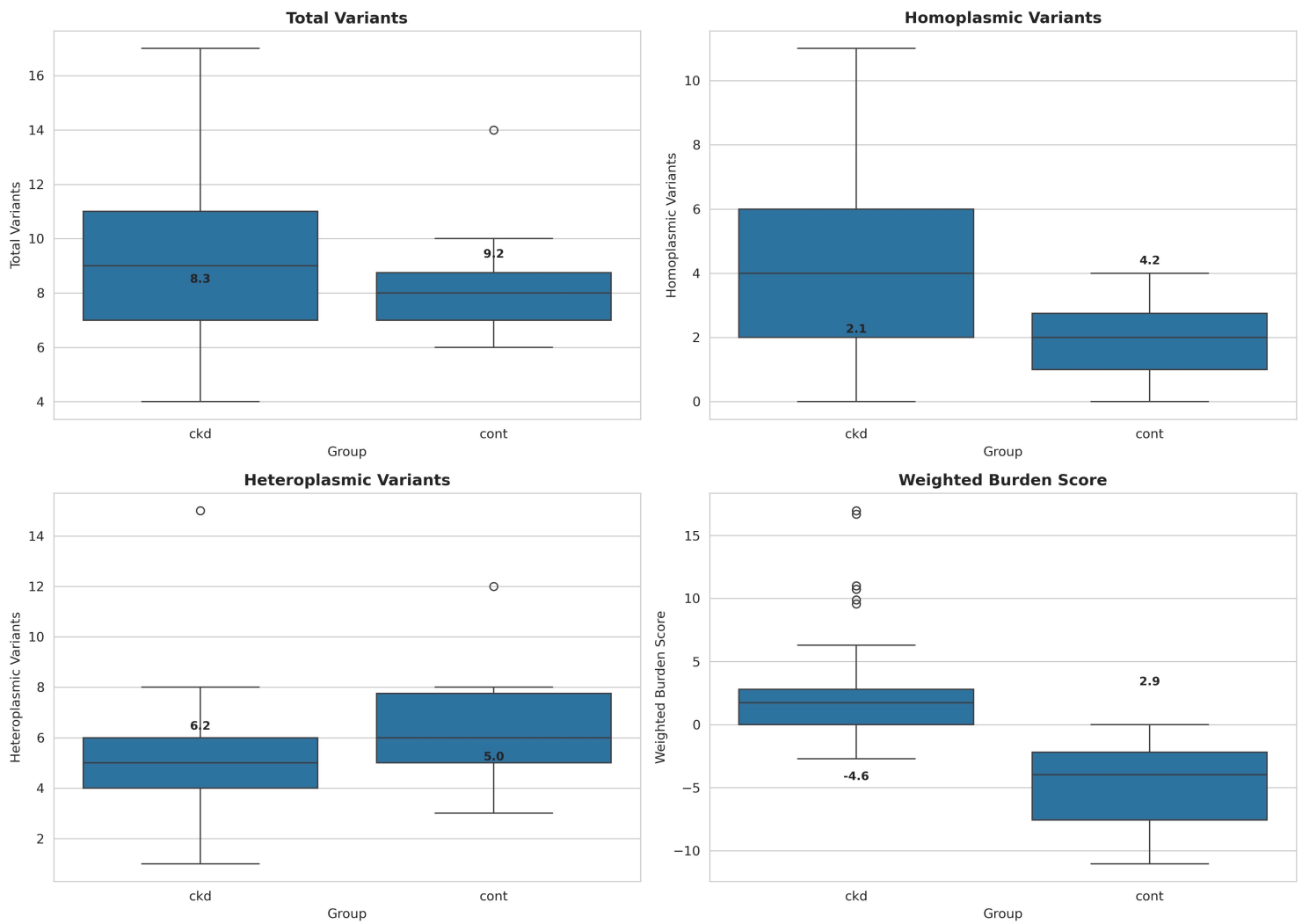


Figure 5

Mitochondrial single nucleotide variant burden analysis. Box plots comparing total variant count, homoplasmic variant count, and heteroplasmic variant count between CKD patients and controls. CKD patients showed significantly higher total and homoplasmic variant burden, while controls exhibited higher heteroplasmic variability. Weighted burden scores were distinctly positive in CKD patients and negative in controls.

Protective Haplotype Analysis in CKD **Synergistic Effects of Mitochondrial Variant Combinations**

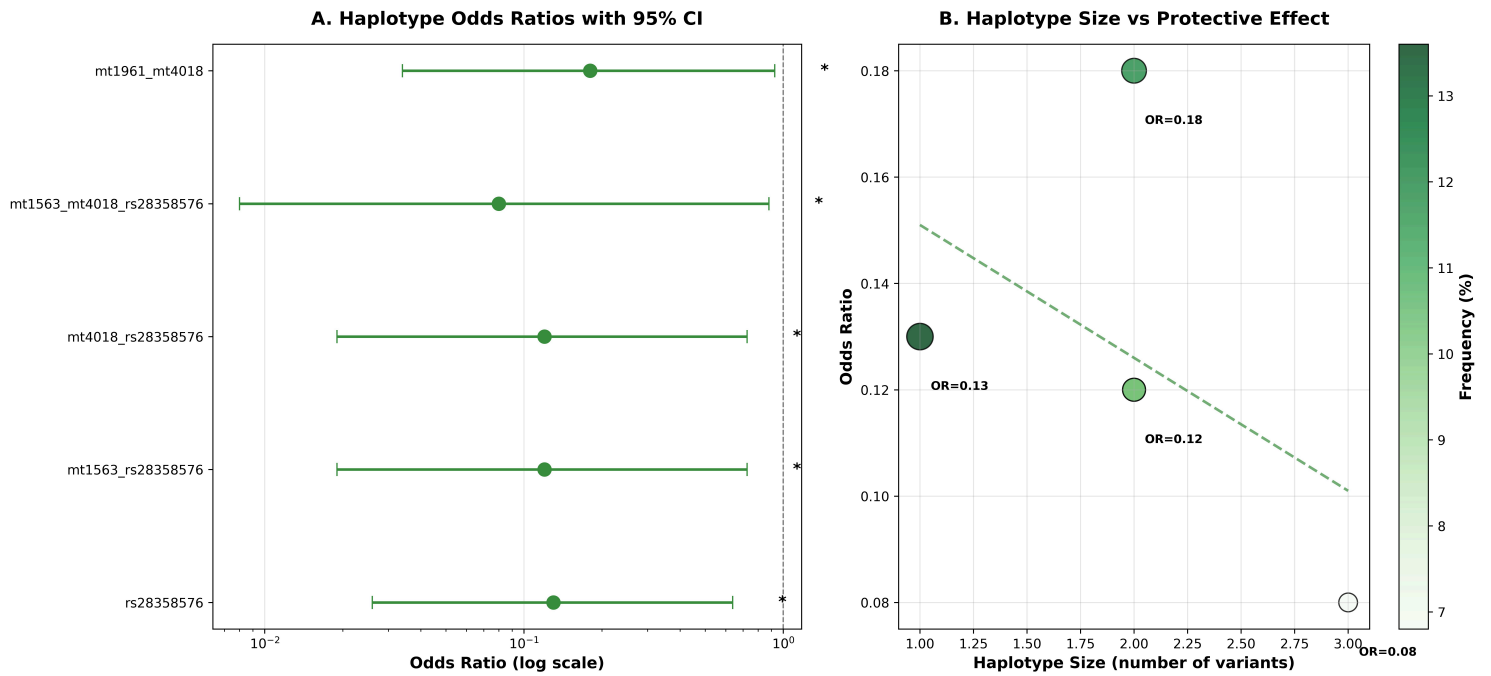


Figure 6

Protective haplotype analysis. Forest plot showing odds ratios (OR) with 95% confidence intervals for haplotypes associated with reduced CKD risk. The single variant rs28358576 (OR = 0.13, $p = 0.0121$), two-variant haplotypes (OR = 0.12), and the three-variant haplotype mt1563-mt4018-rs28358576 (OR = 0.08, $p = 0.0389$) demonstrate cumulative protective effects with increasing haplotype size.

Chronic Kidney Disease Genetic Associations Genetic Association Network (59 samples: 45 CKD, 14 Controls)

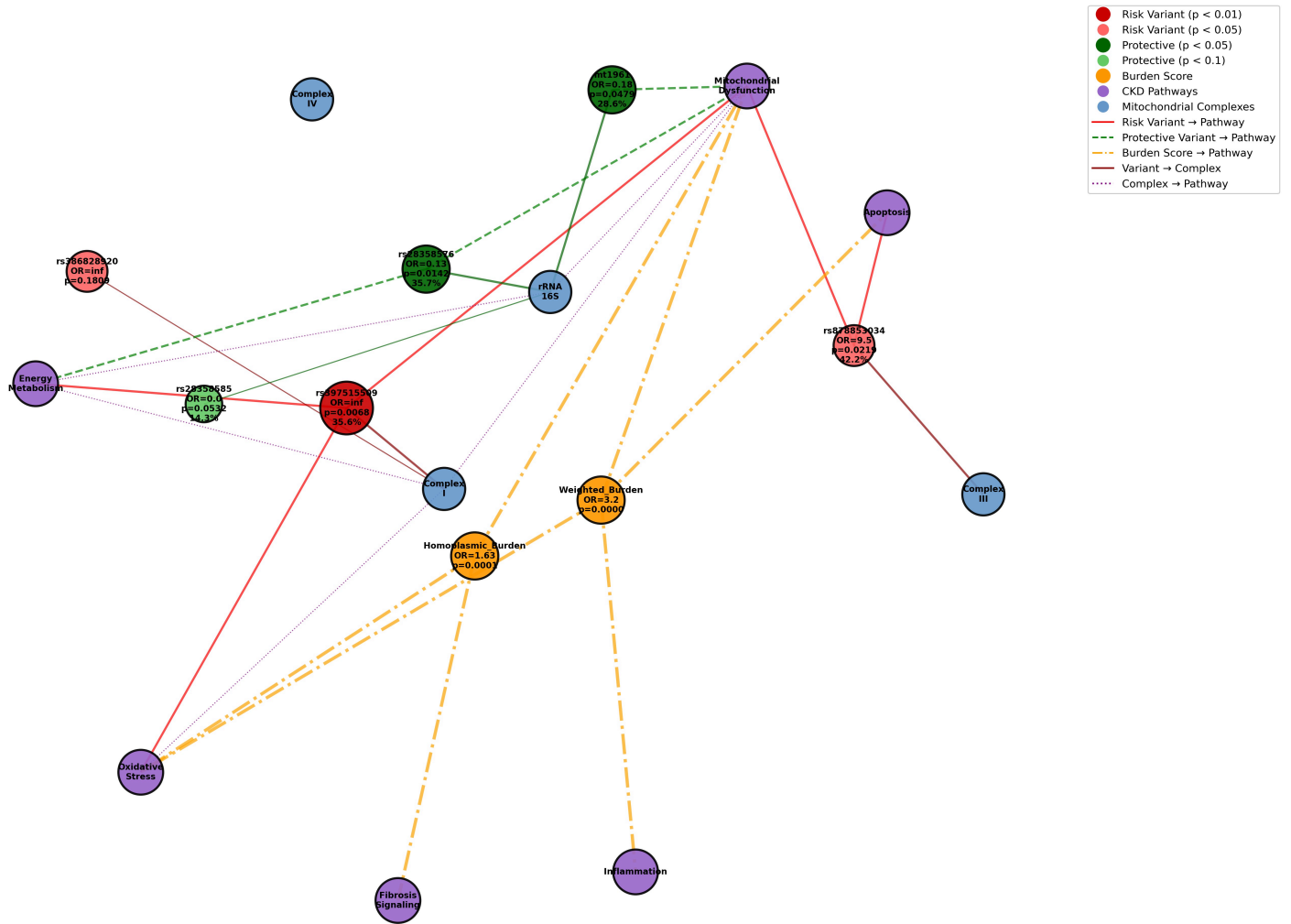


Figure 7

Network-based mitochondrial genetic variants association analysis in CKD. It illustrates a genetic association network integrating specific genetic variants, mitochondrial functional pathways, and burden-related measures in CKD. Solid and dashed edges represent different types of associations, including variant–pathway, variant–burden, and burden–pathway relationships, while node colors indicate variant categories and functional groups.

Supplementary Files

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

- [TableS3pathwayforFigure7.xlsx](#)
- [TableS4variantpositions.csv](#)
- [TableS2haplotypeassociationresults.csv](#)
- [TableS1processedmitochondrialvariants.csv](#)