The role of control region mitochondrial DNA mutations in cardiovascular disease: stroke and myocardial infarction

Miriam Umbria¹, Amanda Ramos²,³,⁴, Maria Pilar Aluja¹* & Cristina Santos¹*

Recent studies associated certain type of cardiovascular disease (CVD) with specific mitochondrial DNA (mtDNA) defects, mainly driven by the central role of mitochondria in cellular metabolism. Considering the importance of the control region (CR) on the regulation of the mtDNA gene expression, the aim of the present study was to investigate the role of mtDNA CR mutations in two CVDs: stroke and myocardial infarction (MI). MtDNA CR mutations (both fixed and in heteroplasmy) were analysed in two demographically-matched case-control samples, using 154 stroke cases, 211 MI cases and their corresponding control individuals. Significant differences were found, reporting mutations m.16145 G > A and m.16311 T > C as potential genetic risk factors for stroke (conditional logistic regression: p = 0.038 and p = 0.018, respectively), whereas the m.72 T > C, m.73 A > G and m.16356 T > C mutations could act as possible beneficial genetic factors for MI (conditional logistic regression: p = 0.001, p = 0.009 and p = 0.016, respectively). Furthermore, our findings also showed a high percentage of point heteroplasmy in MI controls (logistic regression: p = 0.046; OR = 0.209, 95% CI [0.045–0.972]). These results demonstrate the possible role of mtDNA mutations in the CR on the pathogenesis of stroke and MI, and show the importance of including this regulatory region in genetic association studies.

Cardiovascular disease (CVD) is one of the most widespread and common causes of death in the world. The onset and severity of these diseases are influenced by both genetic and environmental factors. Recent evidences associate mitochondrial dysfunction with several cardiovascular manifestations, mainly driven by the central role of mitochondria in cellular metabolism, particularly in energetically demanding tissues such as brain and heart

Human mitochondrial DNA (mtDNA) is a 16.6-kb double-stranded circular DNA molecule that encodes for 13 electron transport chain (ETC) proteins, 2 ribosomal RNAs (rRNAs) and 22 transports RNAs (tRNAs). The control region (CR) encompasses the light and heavy strand promoters, the heavy strand origin of replication (OH), three conserved sequence blocks and the termination associated sequences (TAS)

MtDNA is more susceptible than nuclear DNA to oxidative damage, probably due to the lack histone complex and an inefficient DNA repair mechanisms, which may serve as a protective barrier against external and internal noxious agents as reactive oxygen species (ROS)

However, the hypothesis of direct damage by ROS is increasingly criticized and it is suggested that errors in mtDNA replication and repair may be the main cause of its high mutation rate (~10-fold greater than in nDNA)

Recent evidence have linked certain CVDs with specific mtDNA mutations including base substitutions, deletions, duplications and point or length heteroplasmy both in coding and noncoding region of mtDNA. In particular, mtDNA mutations located in CR have a potential importance since they may influence on the regulation of the mtDNA gene expression. In fact, several studies detected association of a great range of mtDNA variants (with negative or beneficial effect, both fixed or in heteroplasmy) and different diseases.

¹Unitat d'Antropologia Biológica, Departament de Biologia Animal, Biologia Vegetal i Ecologia. Facultat de Biociències, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, GREAB – Research Group in Biological Anthropology, Barcelona, Spain. ²Faculdade de Ciências e Tecnologia, Universidade dos Açores (UAc), Ponta Delgada, Portugal. ³Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Porto, Portugal. ⁴Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal. *email: mariapilar.aluja@uab.cat; cristina.santos@uab.cat
In general, these disparities could occur because mtDNA mutations in the CR may not be directly tied to any form of pathology, but could be capable of influencing mitochondrial function through changes in the number of copies, inducing profound effects on the expression of mitochondrial-encoded gene transcripts and related enzymatic activities (complexes I, III, and IV)\(^\text{18,19}\).

The ability of mtDNA mutation to influence in the development of CVDs is directly related to its prevalence and the severity of its impact on mitochondrial function. In addition, several studies have demonstrated that due to the differences in the prevalence of the main etiological factors between intra- and extracranial arteries, the effect of mtDNA mutations in individuals with stroke or MI could be different\(^\text{20–23}\). The main aim of the present study was to investigate the role of CR mtDNA mutations (fixed or in heteroplasmy) in two CVDs; stroke and MI.

### Results

#### Analysis of fixed and heteroplasmic mtDNA mutations with stroke.

A detailed matrix of all mtDNA positions analyzed in stroke cases and controls are reported in Supplementary Table S1 and the frequencies of fixed mutations found are shown in Table 1. The percentages of m.16145 G > A and m.16311 T > C were overrepresented in stroke cases (5.2% and 18.2%, respectively) comparatively to controls (1.9% and 9.7%, respectively). After correction for the effect of CV risk factors with significant differences between stroke cases and controls (hypercholesterolemia\(^\text{24}\) ), significant association was still observed in m.16145 G > A (conditional logistic regression: \(p = 0.038; \text{OR} = 4.407, 95\% \text{CI} [1.086–17.883]\) ) and m.16311 T > C (conditional logistic regression: \(p = 0.018; \text{OR} = 2.417, 95\% \text{CI} [1.165–5.016]\) ), emerging as possible genetic risks factors for stroke (Table 1).

Stability analyses were performed to predict the impact of these mutations. Several measures as the number of hits in the mtDNA phylogeny, the probability of mutation, the frequency in the population database and the conservation index (CI) at nucleotide level, were calculated, and results are showed in Table 1. The results obtained revealed m.16145 G > A and m.16311 T > C as non-stable positions since they present a minimum of 37 hits in the phylogeny, a high probability of mutation, a high frequency of the variant in the population database (here denoted by minor allele frequency [MAF] > 5%) detected on m.16311 T > C or low-frequency (MAF 1–5%) detected on m.16145 G > A and a maximum nucleotide CI of 58% (Table 1). To infer about the impact of m.16145 G > A and m.16311 T > C on the stability of secondary structures of the mtDNA, a prediction of different structures with the Revised Cambridge Reference Sequence (rCRS) and mutant variant was performed. It seems that m.16311 T > C implies a conformational rearrangement, resulting in structure of Fig. 1 as the new predicted minimum free energy solution (~0.40 kcal/mol), causing a stability reduction of the region. No structural or thermodynamic differences were found for m.16145 G > A.

The distribution of the heteroplasmic positions between stroke cases and controls are reported in Table 2. Eighty-eight stroke cases (57.1%) and eighty-six controls (55.8%) presented point and/or length heteroplasmy (PH and LH, respectively), and no significant differences were obtained between groups. The most prevalent variant detected was a length heteroplasmy located in the poly-C tract of the HVRII (between positions 303–315 of the mtDNA), which was present in a 52% of stroke cases and in 46.7% of controls. Point heteroplasmy were found in six stroke cases and six controls, involving eight different positions of the mtDNA: 146, 150, 152, 185, 204, 16092, 16129 and 16399.

The analysis of stability performed to predict the impact of these heteroplasmic positions is presented in Table 3. In general, these mutations have a minimum of 16 hits in the phylogeny, were located in hotspots positions, have a high frequency of the minor variant in the population database and a low conservation index, indicating that these heteroplasmies have typical characteristics of non-stable positions. Moreover, the frequency of the minor variant is relatively low (Table 3).

#### Analysis of fixed and heteroplasmic mtDNA mutations with MI.

MtDNA positions studied for MI cases and controls are available in Supplementary Table S1 and frequencies of fixed mutations are reported in Table 4. The m.72 T > C, m.73 A > G and m.16356 T > C were more frequent in MI controls (12.3%, 49.3% and 3.3%, respectively) than cases (7.6%, 38.9% and 1.4%, respectively). When corrected for the effect of CV risk factors with significant differences between MI cases and controls (hypertension and hypercholesterolemia\(^\text{24}\)),

| Position 16145 G > A | Position 16311 T > C |
|----------------------|-----------------------|
| Controls: n = 154 (%)| Cases: n = 154 (%)     |
| 3 (1.9)/8 (5.2)      | 15 (9.7)/28 (18.2)    |

### Table 1. Complete results of stroke fixed mtDNA mutation analysed.

*Conditional regression model was performed considering significant covariates for stroke (hypercholesterolemia and mtDNA mutations).
A significant association was observed in these tree mutations (m.72 T > C: conditional logistic regression: \( p = 0.001 \); OR = 0.041, 95% CI [0.006–0.290], m.73 A > G: conditional logistic regression: \( p = 0.009 \); OR = 0.009, 95% CI [0.307–0.843] and m.16356 T > C: conditional logistic regression: \( p = 0.016 \); OR = 0.091, 95% CI [0.013–0.639]), emerging as possible protective genetic factors for MI (Table 3).

In order to predict the impact of these mutations, several measures were calculated to analyze the stability of each position, and results are showed in Table 4. The results obtained revealed that m.72 T > C, m.73 A > G and m.16356 T > C are non-stable positions since they present a minimum of 9 hits in the phylogeny, a high probability of mutation, a high frequency of the minor variant in the population database (MAF > 5%) for m.73 A > G and low-frequency (MAF 1–5%) for m.72 T > C and m.16356 T > C, and a maximum nucleotide CI of 79% (Table 4).

Using the proposed previously method to predict the impact of these three mutations on the stability of secondary structure of the mtDNA, it seems that m.72 T > C, m.73 A > G and m.16356 T > C led to a folded structure with the same minimum free energy as the rCRS structure, which means that these mutations do not condition the stability of the region.

Table 2. Classification of the analyzed stroke and myocardial infarction (MI) individuals depending on the type(s) of heteroplasmy they present. PH: point heteroplasmy; LH: length heteroplasmy. Logistic regression model was performed considering significant covariates for stroke (hypercholesterolemia and mtDNA heteroplasmy) and MI (hypertension, hypercholesterolemia and mtDNA heteroplasmy).

| Stroke Controls | Patients with stroke | MI Controls | Patients with MI |
|----------------|---------------------|-------------|-----------------|
| n = 154 (%)    | n = 154 (%)         | n = 211 (%) | n = 211 (%)     |
| Homoplasmy     | 66 (42.86)          | 68 (44.16)  | 91 (43.13)      |
| Heteroplasmy   | 88 (57.14)          | 86 (55.84)  | 120 (56.87)     |
| 1PH            | 1 (0.65)            | 1 (0.65)    | 1 (0.47)        |
| 1LH            | 71 (46.10)          | 70 (45.45)  | 104 (49.29)     |
| >1PH           | 0 (0.0)             | 0 (0.0)     | 0 (0.0)         |
| >1LH           | 15 (9.74)           | 15 (9.74)   | 15 (7.11)       |
| PH + LH        | 5 (3.25)            | 5 (3.25)    | 10 (4.74)       |
| Total PH       | 6 (3.90)            | 6 (3.90)    | 11 (5.21)       |
| Total LH       | 86 (55.84)          | 85 (55.19)  | 119 (56.40)     |

Table 2. Classification of the analyzed stroke and myocardial infarction (MI) individuals depending on the type(s) of heteroplasmy they present. PH: point heteroplasmy; LH: length heteroplasmy. Logistic regression model was performed considering significant covariates for stroke (hypercholesterolemia and mtDNA heteroplasmy) and MI (hypertension, hypercholesterolemia and mtDNA heteroplasmy).
heteroplasmy, being the length heteroplasmy located in the poly-C tract of the HVRII the most prevalent variant in both MI cases (54.03%) and controls (48.34%). In this analysis, point heteroplasmy was significantly more frequent in MI cases (54.03%) and controls (48.34%). In this analysis, point heteroplasmy was significantly more frequent in MI controls (n = 11; 5.21%) than cases (n = 24. In Fig. 2 are shown the distribution of m.16145 G > T in stroke and MI. The stability analysis to identify the impact of these point heteroplasmic positions is presented in Table 3. All of them were considered non-stable positions. As previously stated, these positions presented a minimum of 12 hits in the phylogeny, were located in hotspots positions, have a high frequency of the minor variant in the population database and low conservation index at nucleotide level. No different trends were observed between stroke and MI cases and controls samples (position, sample name, heteroplasmy type, heteroplasmy origin, distribution in population database, number of hits in mtDNA phylogeny [PhyloTree.org] and by Soares et al., probability of mutation and nucleotide Conservation Index).

Table 3. Complete results of heteroplasmic positions analysed in stroke and myocardial infarction cases and controls samples (position, sample name, heteroplasmity type, heteroplasmy origin, distribution in population database, number of hits in mtDNA phylogeny [PhyloTree.org] and by Soares et al., probability of mutation and nucleotide Conservation Index).

| Position | Sample name | Het | CRS | Distribution in population database | Mean proportion height peaks | No. Hits phylogeny (PhyloTree.org) | No. Hits Soares et al.24 | Probability of mutation | Nucleotide Conservation Index (%) |
|----------|-------------|-----|-----|------------------------------------|-----------------------------|-----------------------------------|--------------------------|--------------------------|----------------------------------|
| 146      | Ca10_Stroke | T/c | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 83.33 T 16.67 C            | 121                              | 0.01020331               | 0.00589722               | 0.054875               |
| 150      | Ca111_Stroke| T/c | C   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 68.72 T 32 C              | 74                               | 0.01020331               | 0.00589722               | 0.054875               |
| 152      | Ca69_Stroke | C/t | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 52.94 C 47.06 T           | 196                              | 0.01469625               | 0.01469625               | 0.014696               |
| 185      | Ca115_Stroke| G/a | G   | G:99.6; A:0.4; T:0.0; C:0.0; GAP:0.0 | 66.67 G 33.33 A          | 64                               | 0.01469625               | 0.01469625               | 0.014696               |
| 204      | Ca72_Stroke | T/c | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 54.55 T 45.45 C          | 44                               | 0.0402509                | 0.0402509                | 0.040250               |
| 16129    | Ca13_Stroke | G/a | A   | G:99.6; A:0.4; T:0.0; C:0.0; GAP:0.0 | 73.33 G 26.67 A          | 93                               | 0.00805017               | 0.00805017               | 0.008050              |
| 146      | Co07_Stroke | T/c | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 69.23 T 30.77 C           | 121                              | 0.01020331               | 0.00589722               | 0.054875               |
| 146      | Co68_Stroke | C/t | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 80.20 C 19.80 T          | 121                              | 0.01020331               | 0.00589722               | 0.054875               |
| 16092    | Co40_Stroke | C/t | T   | T:99.8; C:0.2; G:0.0; A:0.0; GAP:0.0 | 82.35 C 17.65 C          | 16                               | 0.00159131               | 0.00159131               | 0.001591              |

Distributions of mtDNA mutations between haplogroups. Haplogroup assignment was previously performed by Umbria et al.24. In Fig. 2 are shown the distribution of m.16145 G > A and m.16356 T > C in stroke group and m.72 T > C, m.73 A > G and m.16356 T > C in MI group taking into account the different mtDNA haplogroups. The search of the mentioned positions in the updated mtDNA phylogeny – mt. Tree build 1725 shows that with the exception of m.72 T > C and m.73 A > G, which can be observed in several European haplogroups such as HV

...
This finding would support our original hypothesis about the consequences that can be associated across multiple diseases, and that the same variant could have opposite effects (increase or decrease the risk) as possible beneficial genetic factors for MI. It has been previously described that the CR mutations can be associated with secondary structures, which act as recognition sites to molecules involved in the premature arrest of H strand elongation.

Recent studies demonstrated that the 5'end of the D-loop is capable of forming secondary structures. The CR variants do not act directly on the ETC affecting mitochondria bioenergetics or ROS generation; they may impact mtDNA transcription because contains the main regulatory sequences for replication initiation and transcription.

An association of several mtDNA alterations (fixed and in heteroplasmy) in the two diseases has been detected in the present study. As regards fixed mtDNA mutations, the set of mutations in stroke and MI cases was compared to controls and significant differences were found in the two diseases, reporting the m.16145 G > A and m.16311 T > C as potential genetic risk factors for stroke, and m.72 T > C, m.73 A > G and m.16356 T > C as possible beneficial genetic factors for MI. It has been previously described that the CR mutations can be associated across multiple diseases, and that the same variant could have opposite effects (increase or decrease the risk) for two different diseases.

The CR variants do not act directly on the ETC affecting mitochondria bioenergetics or ROS generation; they may impact mtDNA transcription because contains the main regulatory sequences for replication initiation and transcription.

Transitions 16145 G > A and 16311 T > C seem to have a pathogenic role in stroke. The analysis of distribution of these mutations clearly showed that they are located in many different haplogroups and consequently these mutations act as haplogroup-independent risk factors. The m.16145 G > A is located between MT-TAS sequence (nt. 16157–16172) and MT-TAS2 sequence (nt. 16081–16138). According to the classic strand-asynchronous mechanism, recent studies demonstrated that the 5' end of the D-loop is capable of forming secondary structures, which act as a recognition site to molecules involved in the premature arrest of H strand elongation.

The biological importance of this region was confirmed by Brandon et al. who also observed multiple tumor specific mutations in the pre-TAS region. These observations suggest that mutations arising near to this conserved motive might be responsible of the alterations in mtDNA replication and transcription. In the same line, m.16311 T > C has been found to be significantly associated with certain types of cancer. This mutation was previously described by Chen et al. in patients with prostate cancer and also has been reported in colorectal cancer and more recently in acute myeloid leukemia. This mutation is located between the control elements M15 sequence (nt. 16194–16208) and the Mt3l sequence (nt. 16499–16506). In this case, our results showed that m.16311 T > C may imply a reduction in the stability of secondary structure of this region, which would affect in the binding grade to mtDNA transcription factors, ultimately affecting on the intensity of transcription regulation. In both cases, these findings strongly suggest that mtDNA CR dysfunction may cause a decrease on the mtDNA copy number, which could affect the efficiency of ETC, lowering the ATP:ADP ratio and increasing ROS production, contributing in stroke development.

### Table 4

| Position 72 T > C | Position 73 A > G | Position 16356 T > C |
|------------------|------------------|----------------------|
| Controls: n = 211 (%) | Cases: n = 211 (%) |
| 26 (12.3)/16 (7.6) | 104 (49.3)/82 (38.9) | 7 (3.32)/3 (1.42) |

Logistic Regression

- **p-value**: 0.001
- **OR [95% CI]**: 0.041 [0.006–0.290] 0.091 [0.013–0.639] 0.001 [0.006–0.290]

**Stability analysis**

| Distribution in population database |
|------------------------------------|
| T: 97.4; C: 2.4; G: 0.1; GAP: 0.1 | T: 80.8; A: 19.1; GAP: 0.1; C: 0.02 |
| No. Hits phylogeny (PhyloTree.org) | 9 | 12 |
| No. Hits Soares et al. | 6 | 11 | 19 |
| Probability of mutation | 0.00056164 | 0.00102967 | 0.00177853 |
| Nucleotide Conservation Index (%) | T: 77.1; GAP: 8.3; A: 6.3; C: 4.2; G: 4.2 | A: 41.7; C: 35.4; G: 22.9 | T: 79.2; C: 12.5; A: 6.3; GAP: 2.1 |

Discussion

In western countries, where the burden of CVD is growing due to effect of CV risk factors, several studies have already shown the strongly relation of the genetic factors. However, little is known about the role of mtDNA CR mutations in development of stroke and MI.

An association of several mtDNA alterations (fixed and in heteroplasmy) in the two diseases have been detected in the present study. As regards fixed mtDNA mutations, the set of mutations in stroke and MI cases was compared to controls and significant differences were found in the two diseases, reporting the m.16145 G > A and m.16311 T > C as potential genetic risk factors for stroke, and m.72 T > C, m.73 A > G and m.16356 T > C as possible beneficial genetic factors for MI. It has been previously described that the CR mutations can be associated across multiple diseases, and that the same variant could have opposite effects (increase or decrease the risk) for two different diseases. This finding would support our original hypothesis about the consequences that can affect the mtDNA mutations in the CR depending on the disease.

The CR variants do not act directly on the ETC affecting mitochondria bioenergetics or ROS generation; they may impact mtDNA transcription because contains the main regulatory sequences for replication initiation and transcription.

Transitions 16145 G > A and 16311 T > C seem to have a pathogenic role in stroke. The analysis of distribution of these mutations clearly showed that they are located in many different haplogroups and consequently these mutations act as haplogroup-independent risk factors. The m.16145 G > A is located between MT-TAS sequence (nt. 16157–16172) and MT-TAS2 sequence (nt. 16081–16138). According to the classic strand-asynchronous mechanism, recent studies demonstrated that the 5' end of the D-loop is capable of forming secondary structures, which act as a recognition site to molecules involved in the premature arrest of H strand elongation.

The biological importance of this region was confirmed by Brandon et al. who also observed multiple tumor specific mutations in the pre-TAS region. These observations suggest that mutations arising near to this conserved motive might be responsible of the alterations in mtDNA replication and transcription. In the same line, m.16311 T > C has been found to be significantly associated with certain types of cancer. This mutation was previously described by Chen et al. in patients with prostate cancer and also has been reported in colorectal cancer and more recently in acute myeloid leukemia. This mutation is located between the control elements M15 sequence (nt. 16194–16208) and the Mt3l sequence (nt. 16499–16506). In this case, our results showed that m.16311 T > C may imply a reduction in the stability of secondary structure of this region, which would affect in the binding grade to mtDNA transcription factors, ultimately affecting on the intensity of transcription regulation. In both cases, these findings strongly suggest that mtDNA CR dysfunction may cause a decrease on the mtDNA copy number, which could affect the efficiency of ETC, lowering the ATP:ADP ratio and increasing ROS production, contributing in stroke development.
Concerning MI, our results showed that m.72 T>C, m.73 A>G and m.16356 T>C act as a beneficial factor for MI. Although a high percentage of individuals with these mutations belonged to the haplogroups HV0, H or U, which have been shown to may have higher oxidative damage\cite{9,24,33,34}, the distribution of mutations in positions 72, 73 and 16356 in our samples was independent of these haplogroups. Since the role of the mitochondrial genome in CVD susceptibility remains uncertain, it is difficult to explain how these mutations can decrease or counteract the progression of MI. Although m.72 T>C, m.73 A>G and m.16356 T>C have been previously related to certain types of cancer\cite{35}, many studies consider that they are recurrent variants common in humans\cite{36}.

Even though the most deleterious mutations are removed by natural selection, a wide range of milder bioenergetic alterations are introduced in certain populations\cite{37}. Some of these variants as m.72 T>C, m.73 A>G and m.16356 T>C could be advantageous and seen as way to facilitate survival in specific environments. In contrary, other mutations as m.16145 G>A and m.16311 T>C escape of intraovarian selection and could cause significant mitochondrial defects and stroke development. Much of the progress in linkage disequilibrium mapping of complex diseases has been made using the major assumptions of the CDCV hypothesis, that is, that common alleles cause common diseases. After found positive associations with common alleles (e.g., those found by Umbria \textit{et al.}\cite{24}), it was necessary to replicate the results and then look for rarer variants, with potentially greater penetrance. However, all the mutations analysed in the present study had a minor allele variant >5% (common variant) or between 1–5% (low-frequency). Although common variants are often associated with OR between 1.2 and 1.5\cite{38}, our results showed higher effect size for pathogenic variants, with OR values of 2.4–4.4, and similarly, higher protective effect size of variants found at higher frequency in controls (ORs < 0.5) relative to cases.

Our findings also showed a significant increase of point heteroplasmy in MI controls in comparison to cases. This result is contrary to expectations, because the presence of heteroplasmy has been commonly associated with aging and degenerative diseases, due to a decline in mitochondrial function in both these processes\cite{39}. Our registered heteroplasmic positions (16399, 16129, 16093, 16092, 73, 146, 150, 152, and 204) were located in hotspots positions of the hypervariable segments. Recent evidences demonstrate that an important fraction of mutations detected in heteroplasmy are germinal or originated in very early stages of the development\cite{40}. Moreover, it is probable that heteroplasmic variants without apparent genetic or cellular functional consequences are observed in apparently healthy individuals\cite{41}. In the present study, the frequency of MI controls with point heteroplasmy in the CR (5.2%; 95% CI [0.045–0.972]) are slightly higher than those reported by Santos \textit{et al.}\cite{44} (3.81%; 95% CI [0.166–0.737]), but less than described by Ramos \textit{et al.}\cite{45} (7.9%; 95% CI [0.041–0.149]), demonstrating that heteroplasmy occurs with appreciable frequency in the general population\cite{42}. This idea is even more reinforced in front of the present data, since the higher representation of point heteroplasies detected were located in non-stable positions and by the fact that no significant differences were found between frequencies of point heteroplasmy in stroke cases and controls.

Many studies have shown that heteroplasmic variants without apparent genetic or cellular functional consequences are observed in apparently healthy individuals\cite{41}. In the present study, the frequency of MI controls with point heteroplasmy in the CR (5.2%; 95% CI [0.045–0.972]) are slightly higher than those reported by Santos \textit{et al.}\cite{44} (3.81%; 95% CI [0.166–0.737]), but less than described by Ramos \textit{et al.}\cite{45} (7.9%; 95% CI [0.041–0.149]), demonstrating that heteroplasmy occurs with appreciable frequency in the general population\cite{42}. This idea is even more reinforced in front of the present data, since the higher representation of point heteroplasies detected were located in non-stable positions and by the fact that no significant differences were found between frequencies of point heteroplasmy in stroke cases and controls.

**Figure 2.** Distribution of fixed mutations between the mtDNA haplogroups. (A) m.16145 G > A and (B) m.16311 T > C for stroke group (1) and (C) m.72 T > C, (D) m.73 A > G and (E) m.16356 T > C for MI group (2).
It should be noted that this study has some limitations and further functional analysis may be required in independent large samples to clarify the contribution of mitochondrial control region variants to stroke and MI development. At the moment there are no experimental models that prove that alterations in these variants really have consequences. In fact, there are mtDNA regulatory sequences that have been assigned to control region although they remain practically unproved. In addition, it must also be taken into account that measurement of mtDNA genetic variation in blood may differ from that in cardiovascular tissues. Another important point is that, although the population specificity of mtDNA exists, it has been observed that positions reported in this work are worldwide distributed (m.16145 G > A, m.16311 T > C and m.16356 T > C) or at least, in several Eurasian haplogroups such as HV and H (m.72 T > C and m.73 A > G) (http://www.phylotree.org/). For this reason, it would be interesting to explore if our results are extrapolable to other populations, since in a different genetic context, these positions might not present the same effects.

Conclusions

In conclusion, the statistically significant differences in the frequencies of variants in the mitochondrial control region sequence between stroke and MI cases and controls at five positions provide new evidence and better understanding of the cellular mechanism by which mtDNA variants contribute to CVD, and endorse the importance of including this regulatory region of the mtDNA in genetic association studies.

Material and Methods

Patients and samples. In this study, data from 730 subjects (154 individuals with stroke history, 211 individuals with MI history and their corresponding control individuals - matched for age (categories ≤ 44, 45–49, 50–54, 55–59, 60–64, ≥ 65 years), sex and geographic origin (North, Central and South regions of Castile and Leon) were used. Samples come from the Cardiovascular Disease Risk Study of Castile and Leon24 in a cross-sectional, observational and descriptive study performed in Castile and Leon (centre-north region of Spain), whose design and analysis have already been described by Umbria et al.24.

For each individual, we also obtained information about history of hypertension (≥ 140/90 mmHg), history of diabetes, history of hypercholesterolemia (≥ 200 mg/dl), cigarette consumption (smokers, former smokers and non-smokers), presence of overweight or obesity (body mass index ≥ 25 kg/m²), presence of high abdominal perimeter in risk range (risk: ≥ 80 cm for women and ≥ 94 cm for men) and presence of high levels of triglycerides (≥ 170 mg/dl). The study was approved by the ethical committee from Universitat Autònoma de Barcelona and appropriate informed consent was obtained from all the individuals.

MtDNA sequence analysis and heteroplasmy authentication. The mtDNA sequences used in the present study were previously obtained by Umbria et al.24. In brief, the control region of the mtDNA was amplified and sequenced between positions 15907 and 580 using primers and conditions described by Santos et al.44. Moreover, coding region phylogenetic mtDNA informative polymorphisms 7028, 11719, 12308, 12705, 13708 and 14766 were further analyzed as previously detailed by Santos et al.47 to improve haplogroup classification.

In the present study, sequences were reassessed and analysed at the nucleotide level to identify not only fixed mutations but also mutations in heteroplasmy. The alignment in relation to the the revised Cambridge Reference Sequence (rCRS)48 and the heteroplasmy detection were performed using the SeqScape 2.5 software (Applied Biosystems, Foster City, USA) following the manufacturer’s protocol.

The authentication of mtDNA heteroplasm was performed following a similar strategy to that used by Santos et al.44,49

1. PCR amplification and sequencing of the control region of the mtDNA.
2. To authenticate the results for samples presenting heteroplasmy in step 1, a second PCR amplification and sequencing were performed.
3. In addition, to exclude a possible contamination of the samples, an analysis of Short Tandem Repeat (STR) DNA profiling was carried out employing AmpFISTR® Identifier® PCR Amplification Kit (Applied Bio-systems, Foster City, USA) following the manufacturer’s protocol.

Thus, point heteroplasmic positions were accepted if they appeared in all the validation steps and no evidence of sample contamination was detected.

Levels of heteroplasmy were determined using the height of peaks in the electropherograms44. To calculate the average heteroplastic levels, the results obtained for at least two sequence reads of each heteroplastic position were used.

Data analysis. Statistical analyses. To compare differences in the CR profile between cases and controls in both stroke and MI, all fixed and heteroplasmic mtDNA mutations were compiled into a matrix considering cases and controls analysed for each disease.

All fixed mtDNA mutations detected in cases and controls (present in a minimum of 10 individuals) were tested together by using a conditional logistic regression analysis (forward stepwise model), adjusting the association analysis for the potential confounding effect of CV risk factors previously detected by Umbria et al.24.

The authors used McNemar’s test or marginal homogeneity test to compare the frequency of sociodemographic,
Table 5. Sociodemographic, biochemical and clinical characteristics of stroke and myocardial infarction cases and controls. MI: Myocardial infarction, SD: Standard deviation, N: North, C: Central, S: South P-value of a paired McNemar test for dichotomous variables samples and Marginal Homogeneity test when a category of the samples is more than two. *p-value of McNemar or marginal homogeneity test used to compare stroke and MI cases and controls.

|                | Stroke Controls n = 154 (%) | Patients with stroke n = 154 (%) | p-value* | MI Controls n = 211 (%) | Patients with MI n = 211 (%) | p-value* |
|----------------|-----------------------------|----------------------------------|----------|--------------------------|-------------------------------|----------|
| Age (mean ± SD, years) | 71.4 ± 8.6                  | 73.9 ± 10.5                      | Matched  | 68.9 ± 11.1              | 71.2 ± 12.1                   | Matched  |
| Male/female     | 91 (59.1)/ 63 (40.9)        | Matched                          |          | 128 (60.7)/ 83 (39.3)    | Matched                       |          |
| Geographic origin (N/C/S) | 40 (26)/75 (48.7)/ 39 (25.3) | Matched                          |          | 54 (25.6)/88 (41.7)/69 (32.7) | Matched                       |          |
| Ever smoking (n/N, %) | 18/154 (11.7)               | 16/154 (10.4)                   | 0.448    | 33/211 (15.6)            | 24/211 (11.4)                 | 0.566    |
| Hypertension (≥ 140/90 mmHg) | 120 (77.9)                 | 134 (87)                         | 0.055    | 146 (69.2)               | 191 (90.5)                   | < 0.001  |
| Diabetes        | 22 (14.3)                   | 32 (20.8)                        | 0.144    | 40 (19)                  | 54 (25.6)                    | 0.099    |
| Hypercholesterolemia (≥ 200 mg/dl) | 64 (41.6)               | 92 (59.7)                        | 0.002    | 100 (47.4)               | 151 (71.6)                   | < 0.001  |
| Overweight or obesity (≥ 25 kg/m2) | 118 (76.6)               | 103 (66.9)                       | 0.092    | 162 (76.8)               | 158 (74.9)                   | 0.734    |
| High abdominal perimeter (≥ 80 or ≥ 94 cm) | 120 (77.9)               | 108 (70.1)                       | 0.169    | 163 (77.3)               | 159 (75.4)                   | 0.731    |
| Triglycerides (≥ 170 mg/dl) | 14 (9.1)                   | 15 (9.7)                         | 1.000    | 28 (13.3)                | 27 (12.8)                    | 1.000    |

biochemical and clinical characteristics above mentioned between stroke and MI cases and controls (Table 5). Hypercholesterolemia was considered a CV risk factor with a potential confounding effect for stroke, while both hypertension and hypercholesterolemia were considered for MI samples. Therefore, Odds Ratios (ORs) and their 95% Confidence Intervals (CIs) were calculated adjusting for the effect of these risk factors in each disease. To compare the presence or absence of point and length heteroplasmy, a logistic regression analysis was used to correct for the effect of CV risk factors above mentioned.

Finally, mtDNA mutations were revised to infer if they were located in haplogroup-defining positions (previously examined in Umbria et al., or acts as an independent genetic factor. Statistical analyses were performed using IBM SPSS ver. 22.0 (SPSS Inc.). All differences were considered significant at p < 0.05.

**Hits in the phylogeny, population database and Conservation Index (CI).** The stability of fixed mtDNA mutations and point heteroplasmic position were analysed as previously detailed by Ramos et al. The number of hits in the phylogeny for each position was compiled from the updated mtDNA phylogeny – mit. Tree build 17 and from Soares et al. From these data, it has been possible to calculate the probability of mutation as the ratio between the observed and the total number of hits. An mtDNA position was considered a hotspot if the mutation probability was ten times higher than the expected mean value. In order to calculate the frequency of each variant between the observed and the total number of hits. An mtDNA position was considered a hotspot if the mutation probability was ten times higher than the expected mean value.

**Structure prediction.** Secondary structures were performed to understand the structural impact of the different variants found. Secondary structures for each position were generated from sequences (A–M) identified by Pereira et al. All sequences were submitted to the RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) using default parameters for DNA secondary structures calculations. The minimum free energy prediction and base pair probabilities were used to estimate the implication in the molecule.

**Data availability**
The data that supports the findings of this study are within the paper and its Supplementary Material File.

Received: 24 January 2019; Accepted: 30 January 2020;
Published online: 17 February 2020

**References**
1. Taverne, Y., Bogers, A., Duncker, D. & Merkus, D. Reactive oxygen species and the cardiovascular system. *Oxid. Med. Cell Longev.* 2013, 862423, [https://doi.org/10.1155/2013/862423](https://doi.org/10.1155/2013/862423) (2013).
2. Panth, N., Paudel., K. & Parajuli, K. Reactive Oxygen Species: A Key Hallmark of cardiovascular disease. *Adv. Med.* 2016, 9152732, [https://doi.org/10.1155/2016/9152732](https://doi.org/10.1155/2016/9152732) (2016).
3. Bonawitz, N., Clayton, D. & Shadel, G. Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol. Cell.* 2017, 61, 813–825, [https://doi.org/10.1016/j.molcel.2016.11.024](https://doi.org/10.1016/j.molcel.2016.11.024) (2006).
4. Alexeyev, M., Shokolenko, I., Wilson, G. & LeDoux, S. The maintenance of mitochondrial DNA integrity–critical analysis and update. *Cold Spring Harb. Perspect. Biol.* 5, a012641, [https://doi.org/10.1101/cshperspect.a012641](https://doi.org/10.1101/cshperspect.a012641) (2013).
5. Melvin, R. & Ballard, J. Cellular and population level processes influence the rate, accumulation and observed frequency of inherited and somatic mtDNA mutations. *Mutagenesis* 32, 323–334, [https://doi.org/10.1093/mutage/gex004](https://doi.org/10.1093/mutage/gex004) (2017).
11. Gaweł, B., Głogowska, J. & Mazurek, U. The G16319A substitution frequency in a hemorrhagic stroke.

12. Levitsky, S.

13. Cardena, M., Mansur, A., Pereira, C. & Fridman, C. A new duplication in the mitochondrially encoded tRNA proline gene in a patient with dilated cardiomyopathy. *Mitochondrial DNA* 24, 46–49, https://doi.org/10.3109/19401736.2012.717933 (2013).

14. Mitrofanov, K. et al. Analysis of mitochondrial DNA heteroplasmic mutations A1555G, C1325T, T3336C, C5178A, G13151A, G14459A, G14846A and G15059A in CHD patients with the history of myocardial infarction. *Exp. Mol. Pathol.* 100, 87–91, https://doi.org/10.1016/j.xmjp.2015.12.003 (2016).

15. Sobenin, I. et al. Association of mitochondrial genetic variants with carotid atherosclerosis. *PLoS One* 3, e68070, https://doi.org/10.1371/journal.pone.0068070 (2013).

16. Golubenko, M. et al. Analysis of heteroplasmy in the major noncoding region of mitochondrial DNA in the blood and atherosclerotic plaques of carotid arteries. *Genetika* 52, 497–502, https://doi.org/10.1134/S1022795416040409 (2016).

17. Sazonova, M. et al. Mitochondrial genome sequencing in atherosclerosis: what's next? *Curr. Pharm. Des.* 22, 390–396, https://doi.org/10.2174/138161282266615112152335 (2016).

18. Tsutsui, H., Kinugawa, S. & Matsushima, S. Oxidative stress and mitochondrial DNA damage in heart failure. *Circ. J.* 72, 31–37, https://doi.org/10.1253/circj.cj.18-0048 (2018).

19. Huang, J. et al. Decreased peripheral mitochondrial DNA copy number is associated with the risk of heart failure and long-term outcomes. *Medicine (Baltimore)* 95, e3323, https://doi.org/10.1097/MD.0000000000003326 (2016).

20. Bang, O. et al. Inflammatory markers, rather than conventional risk factors, are different between carotid and MCA atherosclerosis. *Neurology* 85, 1128–1134, https://doi.org/10.1212/wnl.0000000000004033 (2015).

21. DE'Arrigo, E. et al. Age-related effects on atherogenesis and scavenger enzyme activities in intracranial and extracranial arteries in men without classic risk factors for atherosclerosis. *Stroke* 32, 2472–2479, https://doi.org/10.1161/110010.098520 (2001).

22. Li, D. et al. Atherosclerosis in stroke-related vascular beds and stroke risk: A 3-D MR vessel wall imaging study. *Ann. Clin. Transl. Neurol.* 5, 1599–1610, https://doi.org/10.1002/acn3.673 (2018).

23. Wei, J. et al. Relationship between coronary atherosclerotic stenosis and cerebral atherosclerotic stenosis. *J. Geriatr. Cardiol.* 5, 227–229 (2008).

24. Umbria, M. et al. Involvement of mitochondrial haplogroups in myocardial infarction and stroke: A case-control study in Castile and Leon (Spain) population. *Mitochondrion* 44, 1–6, https://doi.org/10.1016/j.mito.2017.12.004 (2019).

25. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, 386–394, https://doi.org/10.1002/humu.20921 (2009).

26. Pereira, F. et al. Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol. Biol. Evol.* 25, 2759–2770, https://doi.org/10.1093/molbev/msx225 (2008).

27. Nicholls, T. & Minczuk, M. In D-loop: 40 years of mitochondrial 7S DNA. *Mitochondrion* 227–229 (2008).

28. Brandon, M., Baidi, F. & Wallace, D. Mitochondrial mutations in cancer. *Oncogene* 25, 4647–4662, https://doi.org/10.1038/sj.onc.1208907 (2006).

29. Chen, J., Golden, N., Greene, G., Mukunyadzi, P. & Kadlubar, F. Extensive somatic mitochondrial mutations in primary prostate cancer using laser capture microdissection. *Cancer Res.* 62, 6470–6474 (2002).

30. Akouchekian, M., Househ, M., Hemati, S., Ansarpour, M. & Shafa, M. High rate of mutation in mitochondrial DNA displacement loop region in human colorectal cancer. *Dis. Colon Rectum* 52, 526–530, https://doi.org/10.1097/01.DCR.0000381992.699 (2009).

31. Silka, T. et al. Characterization and prognostic significance of mitochondrial DNA variations in acute myeloid leukemia. *Eur. J. Haematol.* 90, 385–396, https://doi.org/10.1111/ejh.12090 (2013).

32. Murakami, H. et al. Polymorphisms in control region of mtDNA relates to individual differences in endurance capacity or trainability. *Jpn. J. Physiol.* 52, 247–256, https://doi.org/10.2170/jjphysiol.52.247 (2002).

33. Fernández, M. et al. Mitochondrial haplogroups H and J: risk and protective factors for ischemic cardiomyopathy. *PLoS One* 7, e44128, https://doi.org/10.1371/journal.pone.0044128 (2012).

34. Hagen, C. et al. Mitochondrial haplogroups modify the risk of developing hypertrophic cardiomyopathy in a Danish population. *PLoS One* 8, e71904, https://doi.org/10.1371/journal.pone.0071904 (2013).

35. Kirches, E. et al. High frequency of mitochondrial DNA mutations in glioblastoma multiforme identified by direct sequence comparison to blood samples. *Int. J. Cancer* 93, 534–538, https://doi.org/10.1002/ijc.1357 (2001).

36. Samanta, D. et al. Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS Genet.* 9, e1003929, https://doi.org/10.1371/journal.pgen.1003929 (2013).

37. Fan, W. et al. A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* 319, 958–962, https://doi.org/10.1126/science.117786 (2008).

38. Bodiner, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 40, 695–701, https://doi.org/10.1038/ng.1326 (2008).

39. Ye, K., Lu, J., Ma, F., Keinan, A. & Gu, Z. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc. Natl. Acad. Sci. USA* 111, 10654–10659, https://doi.org/10.1073/pnas.1403521111 (2014).

40. Ramos, A., Aluja, M. P., Lima, M. & Santos, C. Somatic vs Germinal Mutations in Mitochondrial DNA: Is There Any Relation with Human Health and Aging? In *Anthropology: Current and Future Developments: Genomics in Biological Anthropology: New Challenges, New Opportunities* (eds. Ramos, A., Lima, M. & Santos, C.) 50–64 (Bentham Science Publishers, 2016).

41. Rose, G. et al. The mitochondrial DNA control region shows genetically correlated levels of heteroplasmy in leukocytes of centenarians and their offspring. *BMJ Genomics* 8, 293, https://doi.org/10.1186/1471-2164-8-293 (2007).

42. Irwin, J. et al. Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *J. Mol. Evol.* 68, 516–527, https://doi.org/10.1007/s00239-009-9227-4 (2009).

43. He, Y. et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 464, 610–614, https://doi.org/10.1038/nature08802 (2010).

44. Santos, C. et al. Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Nat. Rev. Mol. Biol.* 22, 1490–1505, https://doi.org/10.1038/molbevms141 (2005).
45. Ramos, A. et al. Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. PLoS One 8, e74636, https://doi.org/10.1371/journal.pone.0074636 (2013).
46. Vega, T. et al. Design of a population-based study of cardiovascular risk in Castile and Leon (Spain) through primary care teams. Gac. Sanit. 21, 84–87, https://doi.org/10.1016/s0213-9111(07)71976-3 (2007).
47. Santos, C. et al. Determination of human caucasian mitochondrial DNA haplogroups by means of a hierarchical approach. Hum. Biol. 76, 431–453, https://doi.org/10.1353/hub.2004.0049 (2004).
48. Andrews, R. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147, https://doi.org/10.1038/13779 (1999).
49. Santos, C. et al. Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. J. Mol. Evol. 67, 191–200, https://doi.org/10.1007/s00239-008-9138-9 (2008).
50. Soares, P. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am. J. Hum. Genet. 84, 740–759, https://doi.org/10.1016/j.ajhg.2009.05.001 (2009).

Acknowledgements
This work was supported by MINECO (project: CGL2014-53781-r) and by Generalitat de Catalunya (Ref. 2017 SGR 1630). Postdoctoral fellowship SFRH/BPD/105660/2015 (AR) was supported by Fundação para a Ciência e a Tecnologia (FCT).

Author contributions
M.P.A. and C.S. contributed to the study design and supervised the work. M.U. performed experiments and gathered the data. M.U., A.R., M.P.A. and C.S. analyzed the data and interpreted the results. M.U. wrote the main manuscript. All authors revised and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59631-x.

Correspondence and requests for materials should be addressed to M.P.A. or C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020