Statistical prediction of pathogenic variant sites in human mitochondrial genomes.

Marcella Attimonelli [(1)], Matteo Accetturo [(1)], and Daniela Lascaro [(1)]

[(1)]Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Bari, Via Orabona, 4 – 70126 Bari, Italy.

m.attimonelli@biologia.uniba.it

Keywords. Mitochondrial disorders, site variability, pathological potential.

Introduction

Mitochondrial DNA disorders – disorders associated with dysfunctions of the oxidative phosphorylation system (OXPHOS) – are caused by inborn metabolism errors and have an estimated frequency of 1 out of 10000 live births.

Due to the relevant role played by the OXPHOS system in ATP production, causes and effects of mitochondrial disorders are highly heterogeneous and complex [1]. Major origin of mitochondrial disorders is in both nuclear and mitochondrial DNA mutations. Although prenatal diagnosis is routine for nuclear DNA mutations, the cases of prenatal diagnosis of mtDNA mutations are rare, even though urgent, as no real therapies exist [2]. However thanks to bioinformatics support, the gap may be reduced in a short time. Indeed, up to now, the pathogenicity of mtDNA mutations has been, in most cases, prevalently validated by their segregation with the disease and their consequent loss of function when the mutation involves a structural gene, but no systematic statistical analysis of the mtDNA SNPs has been performed. Moreover the criteria commonly followed to associate a mutation to a given pathology are:

- aminoacidic change in a strictly conserved site; - presence in patients only; - heteroplasmy condition; - presence in phenotipically similar, but ethnically different families.

However a strict correlation mutation-phenotype in patients is not always verified.

Here we propose a statistical approach aimed to contribute in the estimation of the pathogenic variation sites.

The analysis is based on the estimation of site-specific relative variability in a sets of homologous sequences, through the application of SiteVarProt [3] and SiteVariability [4] softwares, in order to infer a correlation between site variability and pathogenicity of a given mutation.

Methods

Site-specific variability indexes have been calculated starting from nucleotidic and amminoacidic multialignments, through the application of SiteVariability and SiteVarProt softwares respectively. Site-specific relative variability values for each i_{th} site (v_i) of a dataset of N sequences, have been estimated according to the following formula:

$$v_i = \sum_{j=1}^{N(N-1)/2} \delta ij / Kj$$

where, as far as nucleotidic sequences are concerned, δ is a parameter assuming value 1 when the variation is present and 0 elsewhere in the position i of the j pair sequences and K_j the mean genetic distance calculated for the j pair on the entire alignment with the GTR model [5] [6], while as far as aminoacidic sequences are concerned, δ is a Blosum-like index (giving the level of similarity between two amminoacids) for the position i of the j pair sequences, and K_j the mean genetic distance calculated with the Kimura model. In both cases values of v_i are normalized respect to the maximum value of variability (v_{max}) calculated for that particular dataset of sequences, obtaining a new value γ_i, in order to make site variability indexes comparable between two or more dataset of sequences.

$$\gamma_i = v_i / v_{max}$$

Sample

Two datasets have been used. The nucleotidic site variability estimate has been performed on mtDNA sequences of the 13 mitochondrial genes coding for OXPHOS proteins, belonging to 687 human subjects from different geographic origin. Most of the sequences have been retrieved from literature except for 7 belonging to West New Guinea individuals, which have been sequenced in our laboratory as part of a broader complete mitochondrial genome sequencing project. Whereas the aminoacidic site variability estimate has been calculated on the same set of sequences as far as human genomes are concerned, plus mtDNA sequences of the 13 mitochondrial genes coding for OXPHOS proteins belonging to 60 mammalian different species retrieved from AMmtDB database [7].

Results

The comparative analysis of site variability patterns together with the association of MITOMAP [8] data relative to polymorphic and pathological mutations, has allowed to infer a relationship between the variability of a given position and the pathogenic potential of the corresponding mutation.

Table 1. Example of mutation classification on the basis of site variability data and their association with MITOMAP data. FP=frequent polymorphism; PPP=Potential Pathological Polymorphism; RP = Rare Polymorphism; PM=pathological mutation; MERRF = Myoclonous Epilepsy with Ragged Red Fibers.

<div align="center">ATP8</div>

aminoacidic site	nucleotidic site	nucleotidic change	aminoacidic change	blosum index	human nucleotidic var.	human aminoacidic var.	mammalian aminoacidic var.	annotations
	8414	C-T	L-F	0	0,61			FP
	8415	T-C	L-D	-4	0			
17	8416				0	1,000	0,601	
	8426	T-C	F-L	0	0			RP
	8427	T-C	F-S	-2	0			PPP
21	8428	T-C			0,095	0,000	0,191	RP
	8429				0,043			
	8430				0			
22	8431	C-T	S-L	-2	0	0,067	0,480	PM (MERRF)

This kind of approach has allowed us to define 4 different classes of mutations as shown in table 1, defined by their own site variability range.
Moreover nucleotidic and aminoacidic site variability patterns of the 13 mitochondrial coding for proteins genes have given the possibility to distinguish between high variable genes and low variable genes, in order to assess their proneness to accumulate pathogenic mutations, also in relationships with the previous defined classes of mutations.
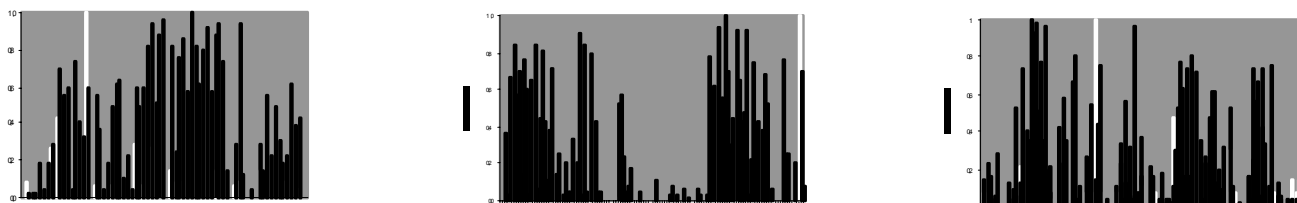

Fig.1. Aminoacidic site variability pattern of ATP8, ND3, COX3; in white human aminoacidic variability, in black mammalian aminoacidic variability.

We can conclude that this new approach could contribute to complete the knowledge about mitochondrial disorders, giving the possibility to shed light on the pathogenic potential of new mitochondrial mutations.

References

[1] E.A. Schon, E. Bonilla, S. DiMauro, Mitochondrial DNA mutations and pathogenesis. *J Bioenerg Biomembr,*29:131-149, 1997

[2] J. Smeitink, L. van den Heuvel, S. DiMauro, The Genetics and Pathology of Oxidative Phosphorylation. *Nature Reviews Genetics,* 2: 342-352, 2001

[3] D. S. Horner, G. Pesole, The estimation of relative site variability among aligned homologous protein sequences. *Bioinformatics* 19: 600–606, 2003

[4] G. Pesole, C. Saccone, A novel method for estimating substitution rate variation among sites in a large dataset of homologous DNA sequences. *Genetic*s, 157: 859–865, 2001

[5] C. Lanave, G. Preparata, C. Saccone, G. Serio, A novel method for calculating evolutionary substitution rates. *J Mol Evol,* 20:86-93, 1984

[6] C. Saccone, C. Lanave, G. Pesole, G. Preparata, Influence of base composition on quantitative estimates of gene evolution. *Methods Enzymol,* 183:570-583, 1990

[7] C. Lanave, S. Liuni, F.Licciulli, M. Attimonelli, Update of AMmtDB: a database of multi-aligned metazoa mitochondrial DNA sequences. *Nucleic Acids Research,* 28:153-154, 2000

[8] A.M. Kogelnik, M. T. Lott, M. D. Brown, S. B. Navathe, D. C. Wallace, MITOMAP: a human mitochondrial genome database—1998 update. *Nucleic Acids Res,* 26:112-115, 1998