InGene: a novel approach for gene analysis and cluster definition in patients with hyperckemia

G. Bertocci[1], I. Ceppa, A. Rubegni[1], C. Dosi[1], D. Tolomeo[1], J. Baldacci[1], G. Astrea[1], E. Dati, S. Frosini, D. Cassandrini[1], R. Conte[3], M. Calderisi[2], F.M. Santorelli[1].

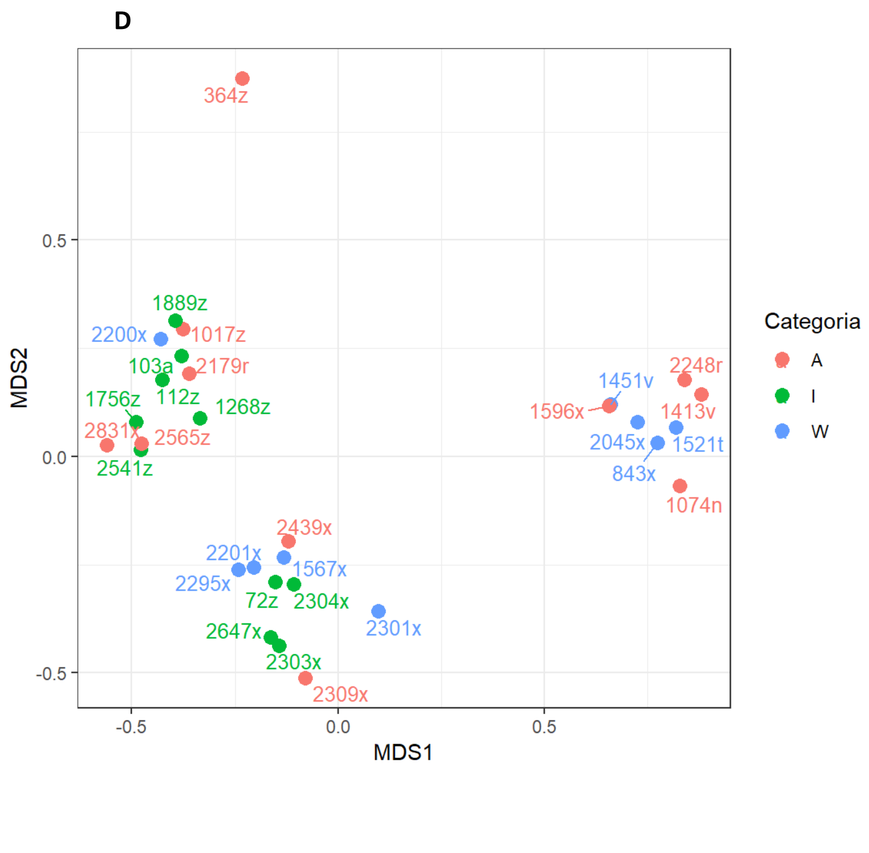
[1]IRCCS Fondazione Stella Maris, Pisa, Italy, [2]Kode Solutions, Pisa, Italy, [3]IFC-CNR, Pisa, Italy

Introduction

Elevated serum creatine kinase (CK) can be found in several genetically well-defined myopathies, usually accompanied by muscle weakness. Next-generation sequencing (NGS) has recently been proposed as a cost-effective strategy for the molecular diagnosis of genetically heterogeneous conditions such as inherited neuromuscular disorders. We evaluated 66 patients presenting high CK levels in blood and performed targeted NGS studies using a neuromuscular multigene panel. Within the Regione Toscana Health project **Ingene**, genetic data, clinical information and MRC values in positive patients were analyzed using non-metric multidimensional scaling (nMDS) and clustering techniques to disclose unanticipated correlations.

Methods

We performed targeted NGS studies using a neuromuscular multigene panel containing a total of 78 genes known to be associated with inherited muscular disorders. Aligning, call, and interpretation for the analysis of the data were performed using SureCall (Algilent) for the assembly and alignment phase, Ingenuity Variant analysis (QIAGEN, Hilden, Germany) and wANNOVAR for the variant calls phase and interpretation. To understand if the plethora of common variants could address specific phenotypes and assist in clustering specific gene/clinical phenotype correlations, we analyzed the variants dataset using a well-established multivariate algorithm nMDS dataset. This is a way of visualizing the level of similarity of individual cases in a dataset. To further explore the results of the nMDS analysis, we then applied a clustering algorithm on the data projection to define clusters between different clinical phenotypes and list of gene variants (rare and common).



Results & Comments

Mutations in *RYR1* were found in 11 cases, whereas 4 subjects had mutations in *ANO5* and 4 in genes encoding sarcoglycans. Three cases harbored mutations in *CAPN3*. Within **Ingene**, genetic data, clinical information in the HPO language (Fig. A and B) and MRC values in 33 positive patients were analyzed using multidimentional nMDS. The algorithm used the information about the specific gene variants found in each patient identifying if specific genes or mutations could cluster (as application for the exploration of the Ingenuity, see Fig. C) and present association with clinical manifestations. With this method, we identified three definite clusters characterized by presence/absence of weakness and dystrophic features in muscle but not matching our earlier categorization based on clinical data only (Fig. D: A=asymptomatic, I=high CK levels, W= weakness). We anticipate that a computer assisted integration of data will help identify more precisely genotype/phenotype correlations and also pinpoint possible genetic modifiers among hyperckemias.