

## ON THE EQUIVALENCE OF METHODS FOR POPULATION STRATIFICATION AND THEIR APPLICATION IN GENETIC ASSOCIATION STUDIES

Nubia Esteban DUARTE<sup>1</sup>  
Suely Ruiz GIOLO<sup>2</sup>  
Mariza de ANDRADE<sup>3</sup>  
Júlia Pavan SOLER<sup>4</sup>

- **ABSTRACT:** *Population-based association studies with unrelated individuals have been used in the mapping of genes involved in the regulation of complex diseases. However, when subjects are from different ethnic ancestries, these studies may yield spurious associations due to population stratification, with an excess of false positive or negative results. Principal components analysis based either on genotype values from known genetic markers (columns of the matrix) or on individuals (rows of the matrix) are the most common approaches used for correction of the confounding effect due the population stratification in genetic association studies. In this paper, results from the singular value decomposition theory of matrices are used to show the analytical equivalence between these approaches, focusing mainly in their relevant role in population stratification analysis. It is also shown the importance of using the biplot as a visualization tool not only to explain the joint information of samples and genetic markers but also to detect informative markers. Although both procedures can be used to correct for population stratification, principal components analysis based on samples is more computationally feasible due to the large number of genetic markers ( $n \ll p$  problem). As an application, it is used genotype data from four HapMap populations..*
- **KEYWORDS:** *Principal components; singular value decomposition; duality in genetic matrix..*

---

<sup>1</sup> Universidade de São Paulo - USP, Faculdade de Medicina, Laboratório de Genética e Cardiologia Molecular, CEP: 05403-900, São Paulo, SP, Brasil. E-mail: [nued03@gmail.com](mailto:nued03@gmail.com)

<sup>2</sup> Universidade Federal de Paraná - UFPR, Departamento de Estatística, CEP: 81531-990, Curitiba, PR, Brasil. E-mail: [giolo@ufpr.br](mailto:giolo@ufpr.br)

<sup>3</sup> Mayo Clinic, Division of Biomedical Statistics and Informatics, Rochester, Minnesota, USA. E-mail: [mandrade@mayo.edu](mailto:mandrade@mayo.edu)

<sup>4</sup> Universidade de São Paulo - USP, Instituto de Matemática e Estatística, CEP: 05508-090, São Paulo, SP, Brasil. E-mail: [pavan@ime.usp.br](mailto:pavan@ime.usp.br)