

# Multipl Miyelom Kanseri Veri Setlerinde Genetik Uzaklık Ölçütlerinin Karşılaştırılması

Anıl AKTAS SAMUR<sup>a</sup>, Mehmet Kemal SAMUR<sup>a</sup>, Özgür TOSUN<sup>a</sup>,  
Osman SAKA<sup>a</sup>

<sup>a</sup> Biyoistatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi, Antalya

## Comparison of Genetic Distance Measures using Multiple Myeloma Datasets

**Abstract:** Quantification of the genetic distance between populations is instrumental in many genetic research initiatives, and a large number of formulas for this purpose have been proposed. A fundamental step in the analysis of gene expression and other high-dimensional genomic data is the calculation of the similarity or distance between pairs of individual samples in a study. Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product.

In this study, we compared the results obtained with five widely used genetic distance measures: Euclidean, Maximum, Manhattan, Canberra and Minkowski distance calculation methods. These five measures were applied to three gene expression datasets of Multiple Myeloma cancer those are publicly available for scientific use. We aimed to understand which of these measures is better to distinguish the patients and control cases. In order to evaluate the differences among these five measures, we used Procrustes analyses. Procrustes analysis can assess the similarity of results produced with different genetic distance measures. Distance values computed with different genetic distance measures using the Generalised Procrustes Analysis function within the FactoMineR library of R. Our findings showed that Maximum distance measure differed the most from remaining measures. Maximum distance measure may be the most effective distance measure for better clustering of Multiple Myeloma cancer patients and controls.

**Key Words:** Gene expression, Procrustes analysis, genetic distance measures

**Özet:** Çoğu genetik araştırmada kitleler arası genetik uzaklığın belirlenmesi gereklidir ve bu nedenle birçok formül önerilmiştir. Gen ekspresyonu ve yüksek boyutlu gen verilerinin analizinde en önemli adım, örneklerdeki birey çiftleri arasındaki benzerlikler ya da uzaklıkların hesaplanmasıdır. Gen ekspresyonu, fonksiyonel gen ürünlerinin sentezlenmesi sırasında kullanılan genden elde edilen bilgilerin kullanıldığı süreçtir.

Bu çalışmada, en sık kullanılan Öklid, Maksimum, Manhattan, Canberra ve Minkowski genetik uzaklık ölçütlerinin sonuçlarını karşılaştırdık. Bu beş ölçüt, bilimsel kullanım için herkese açık olan Multipl Miyelom kanseri gen ekspresyon veri setlerine uygulandı. Çalışmada amacımız, bu uzaklık ölçütlerinden hangisinin hasta ve kontrol olgularını daha iyi ayırdığını anlamaya çalışmaktır. Bu beş ölçüt arasında farklılığı değerlendirmek için Procrustes analizi kullanıldı. Bu analiz farklı genetik uzaklık ölçütleri ile üretilen sonuçların benzerliklerini değerlendirmek için kullanılır. Uzaklık değerleri, R programlama dili FactoMineR kütüphanesi içindeki Genelleştirilmiş Procrustes Analizi fonksiyonu kullanılarak hesaplanmıştır. Sonuç olarak, Maksimum uzaklık ölçütü diğer ölçütlere göre daha farklı bulunmuştur. Maksimum uzaklık ölçütü, Multipl Miyelom kanser hastalarını ve kontrol gruplarını iyi sınıflandırmada diğer ölçütlere göre daha etkili ölçüt olabilir.

**Anahtar Kelimeler:** Gen ekspresyonu, Procrustes analizi, genetik uzaklık ölçütleri

## Sorumlu Yazarın Adresi

Anıl AKTAS SAMUR, Akdeniz Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı,  
[anilaktas@akdeniz.edu.tr](mailto:anilaktas@akdeniz.edu.tr)