

Hastalıklarla İlişkili Genetik Yolakların Bulunmasında İki Farklı Genomik Veri Tipinin Entegrasyonunu Sağlayan Bir Yöntem

Mehmet Kemal SAMUR^a, Özgür TOSUN^a, Uğur BİLGE^a

^a Biyoistatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi, Antalya

Method For Integrating Two Different Genomic Data Types To Identify Disease Related Genetic Pathways

Abstract: Depending on the technological advances and reduced costs, the amount of data produced with genetic studies increased dramatically. Parallel to this increase, the need for new methodologies to process and analyze these data rapidly accelerated, as well. One of the main goals of Bioinformatics is to develop appropriate methods and algorithms and apply them on genetic data to obtain biologically and clinically significant information. SNP and gene expression data are extensively used in genetic research. However, the analysis methods specifically used for these data have their own limitations. Such limitations may be related with either the nature of the data or the methods applied. It is believed that developing new approaches to integrate the analysis methods used for both data types would be crucial to overcome the limitations. The purpose of this study is to develop appropriate statistical procedures and tools those will use SNP and gene expression based analysis methods in an integrative manner. In order to evaluate the accuracy of the methodology, it will be applied on publicly available SNP and gene expression datasets. In the second step, statistically significant genes those contribute explaining the complex phenotypes in both data types will be determined. At the last step, the findings obtained with the analysis of two data types will be combined and gene set enrichment analyses will be applied to the statistically significant genes. The results will be re-evaluated based on current literature and expert assessment.

Key Words: Gene Expression, SNP, Integrative Analysis

Özet: Teknolojik gelişmeler ve maliyetlerdeki azalmalar sayesinde genetik alanında üretilen verinin miktarı hızla artmaktadır. Bu artışa paralel olarak elde edilen verilerin işlenmesi ve çözümlenmesi için yeni analiz yöntemlerine gereksinim duyulmaktadır. Biyoinformatik alanının esas amaçlarından birisi de uygun yöntem ve algoritmalar geliştirilerek genetik verilerden biyolojik ve klinik olarak anlamlı bilgiler elde edebilmektir. SNP ve gen ekspresyonu temelli veriler genetik araştırmalarda yaygınlıkla kullanılmaktadır. Ancak bu verilere özelleşmiş analiz yöntemlerinin çeşitli kısıtlılıkları bulunmaktadır. Bu kısıtlılıklar hem verinin özelliklerinden hem de uygulanmakta olan yöntemlerden kaynaklanabilmektedir. Bu iki veri tipi için kullanılan analiz yöntemlerinin geliştirilecek olan yeni bir yaklaşımla entegre edilmesinin benzer kısıtlılıkların üstesinden gelebilmek için gerekli olduğu düşünülmektedir. Bu çalışmanın amacı SNP ve gen ekspresyonu tabanlı genetik araştırma yöntemleri entegre şekilde kullanılarak uygun istatistiksel yöntem ve araçlar geliştirmektir. Geliştirilen metodolojinin doğruluğunu değerlendirmek için, metot açık olarak erişilebilen SNP ve gen

ekspresyonu verisetlerine uygulanacaktır. Bir sonraki adımda iki veri tipi için de kompleks fenotiplerin açıklanabilmesini sağlayacak istatistiksel önemlilik gösteren genler belirlenecektir. Son aşamada ise iki veri tipinden elde edilen bulgular istatistiksel yöntemlerle birleştirilip anlamlılık gösteren genlere gen seti zenginleştirme analizi uygulanacaktır. Elde edilen tüm bulguların literatürdeki bilgilerle ve uzman değerlendirmesi ile uyumluluğu araştırılacaktır.

Anahtar Kelimeler: Gen ekspresyonu, SNP, Entegre analiz

Sorumlu Yazarın Adresi

Mehmet Kemal Samur. Akdeniz Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı.
Antalya. samur@akdeniz.edu.tr