

# Gen ve miRNA ifade profillerinin ağ tabanlı entegrasyonu için web sunucusu

Mehmet Kemal SAMUR<sup>ab</sup>, Zhenyu YAN<sup>b</sup>, Parantu SHAH<sup>b</sup>, Cheng LI<sup>b</sup>

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

<sup>b</sup> Biyoistatistik ve Hesaplamalı Biyoloji AD, Dana Farber Kanseri Enstitüsü-Harvard School of Public Health, Boston, USA

## A web server for network based integration of gene and miRNA expression profiles

**Abstract:** A large amount of genome-wide gene expression and microRNA expression profiles for the same set of patients covering multiple subtypes are increasingly becoming available from focused efforts of multiple laboratories and large projects such as TCGA and ICGC. The proposed method identifies candidate transcription factor and microRNA regulators by statistically ranking computationally predicted feed-forward loops consisting of these regulators and their common target genes. Therefore, the method integrates not only gene and miRNA expression profiles but also computationally derived information about transcription factor-target and microRNA-target interactions and generates biologically testable hypotheses. The web server is available at <http://www.canevolve.org/dChip-GemiNi>.

**Key Words:** mRNA, miRNA, Transcription Factors, Feed Forward Loops, Integrative Analysis

**Özet:** TCGA ve ICGC gibi büyük çaplı projelerin ve bireysel laboratuvarların bu ananda odaklanması ile büyük miktarda genom çapında gen ifadesi verisi ve microRNA ifadesi verisi erişilebilir duruma gelmiştir. Bu çalışmada önerilen metod aday transkripsiyon faktörlerini ve microRNA düzenleyicilerini ortaya koymaktadır. Bu noktada metod sadece gen ve miRNA ifadelerini entegre etmeyip aynı zamanda transkripsiyon faktörü ve miRNA hedeflerinin etkileşimlerini entegre etmekte ve biyolojik olarak test edilebilir hipotezler ortaya koymaktadır. Bu metodu temel alan web sunucusuna <http://www.canevolve.org/dChip-GemiNi> adresinden erişilebilir.

**Anahtar Kelimeler:** mRNA, miRNA, Transkripsiyon Faktörü, İleri Beslemeli Ağ, Entegre Analiz

## 1. Giriş

Transkripsiyon faktörü ve miRNA'daki düzensizliklerin hücre içindeki düzenleyici ağları etkiledikleri ve kanser gibi kompleks hastalıklara neden oldukları bilinmektedir [1-2]. TCGA ve ICGC gibi büyük organizasyonların ve bireysel laboratuvarların

girişimiyle çeşitli hastalılara ait aynı hasta grubundan elde edilmiş genom çapında gen ve miRNA ifade verileri hızla artan şekilde erişilebilir hale gelmektedir. Bu artışın sonucu olarak da gen ve miRNA ifadesi verilerini entegre eden pek çok metot ortaya konulmuştur. Ancak regülasyonları göz önüne alarak gen ifadesi ve miRNA ifadesini entegre eden pek az araç geliştirilmiştir [3-6].

Örneğin MAGIA ve MMIA yöntemleri korelasyon ve karşılıklı bilgi gibi ilişki metriklerinden faydalanmaktadır. MAGIA farklı ilişki ölçümlerini kullanarak eş veya eş olmayan ifade profillerini için hedef tahminlerini ve gen ifadesi entegre etmektedir [3]. MMIA ise miRNA ifadeleri ve mRNA hedef genleri arasındaki negatif korelasyonu temel alarak oluşturulmuştur [7-8]. miRGator[5,7], DIANA-microT[6,9] gibi araçlar ise miRNA ve mRNA profillerini entegre etmek için geliştirilmiş diğer araçlardır. mirConnX ise daha TF-hedef motif taraması, miRNA hedef tahminleri ve literatür taraması yapılarak önceden derlenmiş ağları üst üste koyarak sonuca gitmeye çalışmıştır [4].

Bu çalışmadan önce mevcut bilgiler ışığında diğer yöntemlerden farklı olarak ileri beslemeli ağları dikkate alan dChip-GemiNi [10] yöntemi mRNA ve miRNA verilerini entegre olarak analiz edebilmek için ekibimiz tarafından geliştirilmiştir. Bu yöntem TF-motif hedef taraması ve miRNA hedef tahminlerini kullanarak hücre içindeki etkilenen ağları tespit etmeyi hedeflemektedir. dChip-GemiNi TF'leri, hedefleri olan miRNA'ları ve ikisinin ortak hedefi olan genler aralarındaki ileri beslemeli ağı temel alarak hücre içindeki etkilerini ortaya koymayı hedefleyen bir yöntemdir. Bu yöntem ileri beslemeli ağlar yerine motifler üzerine de uygulanabilecekken, şu anda ileri beslemeli ağların önemi ve bunların biyolojik ağlar üzerindeki etkileri nedeniyle ileri beslemeli ağlara odaklanmıştır. Bu metotta göre iki farklı durumdan (örn: Normal birey ve hasta birey) elde edilen gen ve miRNA ifade profilleri ile ileri beslemeli ağlar oluşturulmakta, ağ motif skorlama prosedürü ile oluşturulan ağlardan iki durum arasında düzensiz regülasyona sebep olan ileri beslemeli ağlar tespit edilmektedir. Bu çalışmada geliştirilen yeni yöntemin biyolog ve araştırmacılar tarafından kolay kullanılmasını sağlamak amacıyla bir internet tabanlı bir araç sunulması hedeflenmiştir.

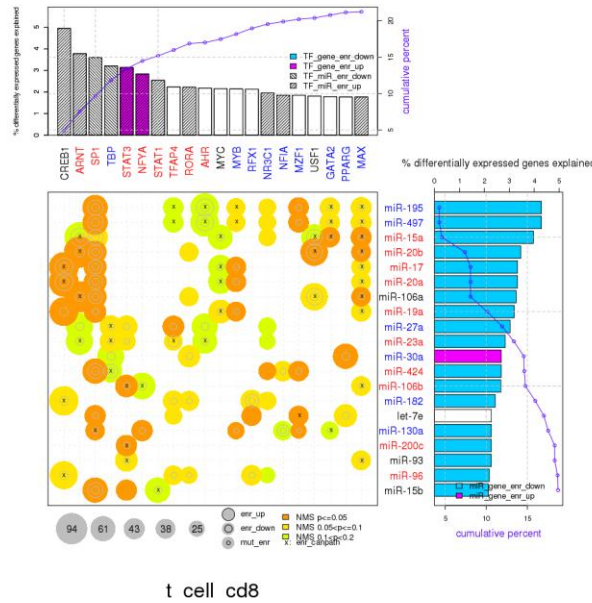
## **2. Analiz Sunucularının Mimarisi**

Geliştirilen analiz sunucusu çok katmanlı mimari ile geliştirilmiştir. En üst katmanda kullanıcıların analizde kullanacakları dosyaları sunuculara gönderebilmesini sağlayan ve gerekli analiz parametrelerini kendi analiz modellerine göre ayarlayabilmelerine olanak sağlayan kullanıcı ara yüzü katmanı yer almaktadır. Bu katman PHP ve AJAX araçları kullanılarak geliştirilmiştir. İkinci katman kullanıcılar tarafından sunucuya gönderilen işleri takip etmede ve yönlendirmede kullanılan mantık katmanıdır. Bu katman canEvolve projesi içindeki framework yardımı ile yönetilmekte ve sunucuya gönderilen taleplerin gerekli kontrollerden sonra R istatistiksel programlama dilini tetikleyerek analiz edilmesine olanak vermektedir. Analiz katmanı ise mantık katmanı tarafından çalıştırılan analizlerin sürdürülmesi ve analiz sonuçlarının veritabanı ve dosya sunucularında depolanmasından sorumludur. Geliştirilen metodoloji permutasyon analizi nedeni ile uzun zamanlar alabileceğinden analiz sunucuları tarafından analizler tamamlandıktan sonra kullanıcılara geri bildirim elektronik iletileri gönderilmektedir.

### 3. Kullanım

Geliştirilen web ara yüzü (<http://www.canevolve.org/dChip-GemiNi>) kullanıcılara analizlerinin parametrelerini daha kolay anlayabilmeleri için çeşitli yardım dokümanları ve veri setlerinin olması gereken formatı daha iyi anlayabilmeleri için örnek veri setleri sağlamaktadır. Genel olarak kullanıcılar analiz sunucularına mRNA ve miRNA ifade verilerini içeren veri matrislerini, analiz parametrelerini ve analizler tamamlandığında sunucular tarafından geri dönüş alabilmek için geçerli bir elektronik posta adresini sunuculara sağlayarak bu analiz hizmetinden faydalanabilmektedirler. Ayrıca eğer kullanıcıların analiz için sunucuya yüklediği veri matrisleri HUGO gen sembollerini ve miRbase miRNA isimlerini içermiyorsa gerekli eşleştirme dosyalarını da sunucuya yüklemelidirler.

Gerekli analiz sunucular tarafından tamamlandıktan sonra kullanıcılara analiz sonuçlarına ulaşabilecekleri bir linkle beraber bilgilendirme postası gönderilmektedir. Kullanıcılar ayrıca analiz sonuçlarının detaylı olarak nasıl yorumlanabileceğine aynı web sitesinde yer alan dokümanlardan ulaşabilirler. Şekil 1 analiz sonuçlarının genel olarak yorumlanabilmesini sağlayan bir analiz çıktısı göstermektedir.



Şekil 1-Örnek analiz baloncuk grafik çıktısı. Şekilde her bir baloncuk TF ile miRNA arasında oluşan istatistiksel olarak anlamlı ileri beslemeli ağı temsil etmektedir. Her bir baloncunun büyüklüğü ise iki durum arasında anlamlı farklılık gösteren ve bu ileri beslemeli ağdan etkilenen gen sayısını, renk ise p değerine göre anlamlılık düzeyini göstermektedir. Seklin üstünde ve yanında yer alan kırmızıdan maviye doğru farklı renk düzeylerinde çizilmiş çubuk grafikler TF ve miRNA'ların sırasıyla yüksek ve düşük regülasyona sahip olduklarını ifade etmektedir.

#### 4. Teşekkür

Yazarlar Cheng Li, Nikhil Munshi ve Dana Gabuzda laboratuvarı üyelerine metodolojinin ve web tabanlı aracın geliştirilmesi sırasında verdikleri desteklerden dolayı teşekkür ederler.

#### 5. Kaynakça

- [1] Calin, G. A., and Croce, C. M. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6,857-866.
- [2] Nebert, D. W. Transcription factors and cancer: an overview. *Toxicology*. 2002; 181-182, 131-141.
- [3] Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. *Nucleic Acids Res*. 2010; 38, W352-359.
- [4] Huang, G. T., Athanassiou, C., and Benos, P. V. mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res*. 2011; 39, W416-423.
- [5] Cho, S., Jun, Y., Lee, S., Choi, H. S., Jung, S., Jang, Y., Park, C., Kim, S., and Kim, W. miRGator v2.0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Res*. 2011, 39, D158-162.
- [6] Maragkakis, M., Reczko, M., Simossis, V. A., Alexiou, P., Papadopoulos, G. L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., et al. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res*. 2009; 37, W273-276.
- [7] Nam, S., Kim, B., Shin, S., and Lee, S. miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res*. 2008; 36, D159-164.
- [8] Xin, F., Li, M., Balch, C., Thomson, M., Fan, M., Liu, Y., Hammond, S. M., Kim, S., and Nephew, K. P. Computational analysis of microRNA profiles and their target genes suggests significant involvement in breast cancer antiestrogen resistance. *Bioinformatics*. 2009; 25, 430-434.
- [9] Maragkakis, M., Alexiou, P., Papadopoulos, G. L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V. A., et al. Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*. 2009; 10, 295.
- [10] Yan Z, Shah PK, Amin SB, Samur MK, Huang N, Wang X, Misra V, Ji H, Gabuzda D, Li C. Integrative analysis of gene and miRNA expression profiles with transcription factor-miRNA feed-forward loops identifies regulators in human cancers. *Nucleic Acids Res*. 2012 Sep 1;40(17):e135. Epub 2012 May 29.

#### 6. Sorumlu Yazarın Adresi

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