

## İnsanda Uzun Kodlanmayan RNA'ların Doku İfade Örüntüleri ve Varsayımsal İşlevlerinin Benzeşme Yayılması Algoritması ile İncelenmesi

Gökhan KARAKÜLAH<sup>1</sup>

**ÖZET:** Kümeleme yöntemleri, yüksek hacimli gen ifadesi örüntülerinden biyolojik olarak anlamlı bilginin elde edilmesinde yaygın olarak kullanılmaktadır. Benzeşme yayılması algoritması, veri noktaları arasından örnekler adı verilen küme merkezlerinin belirlendiği ve bunların etrafında kümelerin oluşturulduğu yeni bir yaklaşımdır. Bu çalışmada, hastalık, gelişim ve farklılaşma gibi farklı hücrel olaylarda düzenleyici transkriptler olan uzun kodlanmayan RNA'ların, benzeşme yayılması algoritması ile 16 farklı sağlıklı insan dokusundaki ifade örüntüleri incelendi. Bununla beraber uzun kodlanmayan RNAların varsayımsal işlevleri, kümeleme yaklaşımı ile bilgisayar tabanlı olarak tahminlendi ve kapsamlı bir ifade örüntüsü – işlev kataloğu hazırlandı.

**Anahtar Kelimeler:** Benzeşme yayılması algoritması, biyoenformatik, kümeleme, uzun kodlanmayan RNA

## Investigation of Tissue Expression Patterns and Putative Functions of Human Long Non-Coding RNAs with Affinity Propagation Algorithm

**ABSTRACT:** Clustering methods have been widely used for extracting biologically relevant information from high-throughput gene expression patterns. Affinity propagation algorithm is a novel clustering approach based on the determination of cluster centers, called exemplars, among data points and the generation of clusters around these exemplars. In this study, expression patterns of human long non-coding RNAs, new class of regulatory transcripts in diverse cellular events, including disease, development and differentiation, were investigated using affinity propagation algorithm in 16 different healthy human tissues. Besides, their putative functions were computationally predicted with clustering approach and a comprehensive long non-coding RNA expression pattern – function catalogue was created.

**Keywords:** Affinity propagation algorithm, bioinformatics, clustering, long non-coding RNA

<sup>1</sup> İzmir Uluslararası Biyotıp ve Genom Enstitüsü, Dokuz Eylül Üniversitesi, İnciraltı, 35340, İzmir, Türkiye  
Sorumlu yazar/Corresponding Author: Gökhan KARAKÜLAH, gokhan.karakulah@deu.edu.tr

## GİRİŞ

Biyoteknoloji alanında 2000’li yılların başından itibaren yaşanan gelişmeler ile birlikte günümüzde, binlerce genin zaman boyunca ya da farklı durumlar altında ifade düzeylerindeki değişimler eş zamanlı olarak ölçülebilmektedir (Schena et al., 1995; Brenner et al., 2000; Chaitankar et al., 2016). Genom düzeyindeki gen ifade profilleri içerisinde, biyolojik ya da klinik olarak anlamlı örüntülerin belirlenebilmesi ancak özel istatistiksel yöntemlerin kullanılması ile mümkün olmaktadır. Bu yöntemler arasında kümeleme analizi, bir biri ile benzer özelliklere sahip nesnelere benzerliklerine göre aynı grup ya da kümeler ayrılması için sıkça tercih edilen bir çözümlenme yöntemidir (Kerr et al., 2008). Benzer ifade örüntülerine sahip gen kümelerinin belirlenmesi, özellikle rolleri tam olarak anlaşılammış genlerin biyolojik işlevlerinin aydınlatılmasına yardımcı olabilmektedir. Bir biri ile ilişkili biyolojik görevleri üstlenen genler, çoğunlukla ifade profilleri açısından aynı gen kümelerinde bulunmaya daha yatkındırlar ve bu yolla işlevleri bilinmeyen bazı genlerin biyolojik rolleri tahminlenebilmektedir (D’Haeseleer, 2005). Hiyerarşik, k-ortalama lar ya da özdüzenleyici harita kümeleme yöntemleri, gen ifade profillerinin anlamlandırılmasında yaygın bir şekilde kullanılan başlıca yaklaşımlardır (D’Haeseleer, 2005).

Son yıllarda geliştirilen kümeleme algoritmalarından, benzeşme yayılması (=affinity propagation) algoritmasının, pek çok kümeleme metodundan farklı olarak küme sayısının önceden belirlenmesini gerektirmemesi, her veri noktası için tercihlerin değerlerinin düzeltilmesi ile küme sayısını kontrol edilebilmesi, girdilerin metrik olmayan benzerlikleri de içerebilmesi, kolaylıkla uygulanabilmesi, tüm veri noktalarını olası örnekler (=exemplars) olarak içermesi gibi avantajları bulunmaktadır (Frey and Dueck, 2007; Li and Wang, 2015).

Benzeşme yayılması algoritması, şu ana kadar beyin fonksiyonlarının tanımlanmasında fonksiyonel manyetik rezonans görüntüleme verilerinin analizi (Li et al., 2010; Zhang et al., 2011; Li and Wang, 2015), DNA dizileme ve genetik çözümlenmeler (Sun et al., 2015; Cao and Amendt, 2016; Jiang et al., 2016), top-

luluk (=ensemble) özellik seçimi (Meng et al., 2016), hedef takibi (Zhang and Wu, 2015) ve kapalı mekanlarda konumlandırma sistemleri için parmak izi oluşturma tasarımları (Caso et al., 2015) gibi sağlık, fen ve mühendislik alanındaki pek çok çalışmada kullanılmıştır. Homojen dağılmayan veri setlerinde, verilerin ideal kümelerle gruplanması durumu, problemi algoritmanın genişletilen versiyonunda çalışılmıştır (Zhao and Xu, 2015). Yöntemin uygulanmasında, her nokta büyük bir ağ yapısındaki birer nokta olarak alınmakta, noktaların birbirlerine gönderdikleri farklı tipteki iletiler aracılığı ile her noktanın merkezi ağ içerisinde belirlenmektedir (Baş ve İkizler-Cinbis, 2013).

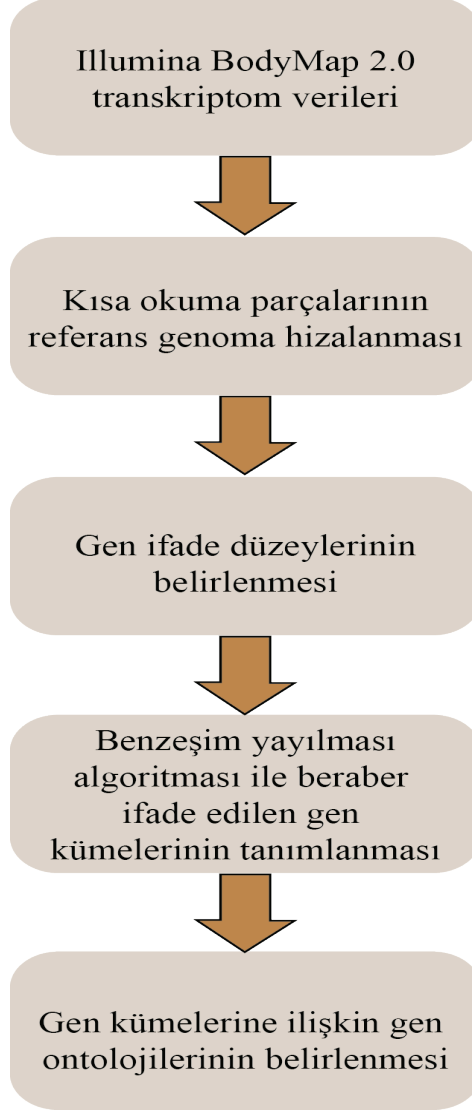
Bu çalışmada, gelişim ve farklılaşma gibi biyolojik süreçlerde gen ifadesinin düzenlenmesinde rol oynayan uzun kodlanmayan RNAların (=long non-coding RNA, ukRNA), (Fatica and Bozzoni, 2014) farklı insan dokularındaki ifade örüntülerinin aydınlatılması ve ilişkili oldukları olası protein kodlayan RNA (pkRNA) transkriptlerinin belirlenmesi amaçlanmıştır. Bu nedenle, *Illumina BodyMap 2.0* RNA dizileme verileri üzerinde benzeşme yayılması kümeleme yöntemi uygulanarak, birlikte ifade edilen ukRNA ve pkRNA gen kümeleri belirlenmiş, ukRNAların olası rolleri bu yolla tahminlenmeye çalışılarak, kapsamlı bir ukRNA işlev kataloğu hazırlanmıştır.

## MATERYAL VE YÖNTEM

*Illumina BodyMap 2.0* projesi kapsamında üretilen 16 farklı doku tipine (adipoz, adrenal bez, akciğer, beyin, böbrek, iskelet kası, kalp, karaciğer, kolon, lenf nodu, lökosit, meme, ovaryum, prostat, testis, tiroid) ait eşli dizileme transkriptom kütüphanesi, *Sequence Read Archive* veri tabanından (Leinonen et al., 2011) (veri tabanı erişim tanımlayıcısı: ERP000546) indirildi (Şekil 1). Transkriptom kütüphanelerine ait kısa okumaların kalite kontrolleri FastQC (Andrews, 2010) yazılımı kullanılarak yapıldı.

Trimmomatic (sürüm 0.36) (Bolger et al., 2014) aracılığı ile *Illumina* adaptör dizileri kısa okuma parçalarından kesilerek uzaklaştırıldı. Her bir transkriptom kütüphanelerine ait kısa okumalar daha sonra Ensembl veri tabanından (Hubbard et al., 2002) indirilen insan referans genomuna (GRCh38.

p7 asemblesi, sürüm 85) STAR (sürüm 2.5) çalıştırılırken, ön tanımlı parametrelerde herhangi (Dobin et al., 2013) yazılımı aracılığı ile hizalandı. bir değişiklik yapılmadı. STAR hizalama algoritması iki-geçiş modunda



Şekil 1. Çalışmaya ait iş akışı şeması

STAR programından elde edilen gen düzeyindeki kısa okuma sayım değerleri, R (sürüm 3.3.1) (Team, 2014) istatistiksel hesaplama ortamında edgeR (Robinson et al., 2010) paketi kullanılarak kütüphane büyüklüğüne göre normalize edildi ve her bir gen için o genin ifade seviyesini belirtmekte kullanılan kilobaz milyondaki toplam okuma sayısı (RPKM değeri) hesaplandı. RPKM değerlerinin hesaplanmasında kullanılan gen dizisi uzunlukları, Ensembl veri

tabanından BioMart ara yüzü kullanılarak indirildi. Çalışmaya 16 farklı dokunun en az sekizinde ifade edilen genler dahil edilirken, bu koşulu sağlamayan genler filtrelendi ve çalışmanın bundan sonraki basamaklarından çıkarıldı.

Benzeşme yayılması algoritmasının gen ifadesi örüntülerinin üzerine uygulanması ile dokuya özgü ukRNA kümeleri ve ayrıca bu ukRNAlar ile birlikte ifade edilen pkRNAlar belirlendi. Benzeşme yayılması

algoritmasının çalıştırılmasında, algoritmanın R ortamına uyarlanmış hali olan APCluster (Bodenhofer et al., 2011) paketinden faydalanıldı. corSimMat() fonksiyonu ile gen ifade profilleri arası korelasyon benzerlik matrisi hesaplandı ve her bir kümeye ilişkin örnek genler belirlendi.

Benzeşme yayılması algoritması ile tanımlanan gen kümelerinin biyolojik rollerinin tahmini için gen ontolojisi (GO) (Ashburner et al., 2000) çözümlenmesi kullanılarak, ukRNAların biyolojik süreçlerdeki olası işlevleri bilgisayar tabanlı olarak tahminlendi.

GO çözümlenmesi yine R ortamında gProfileR paketi (Reimand et al., 2016) aracılığı ile gerçekleştirilirken, 'Benjamini Hochberg' düzeltmesi uygulanan olasılık değerlerinden, 0.05 eşik değerinden daha küçük hatalı keşif oranına (=False Discovery Rate) sahip GO terimleri anlamlı olarak kabul edildi.

## BULGULAR VE TARTIŞMA

FastQC yazılımı ile gerçekleştirilen transkriptom kütüphanelerinin kalite kontrolünde, herhangi bir RNA dizileme artefaktına rastlanmadı ve çalışmaya tüm kütüphaneler dahil edildi.

Illumina adaptör dizileri belirlenip, uzaklaştırılan kısa okuma parçalarının STAR hizalama algoritması ile insan referans genomu üzerinde eşleştirilmesi sonrasında ortalama %87.74 eşsiz hizalama oranı elde edilirken, ortalama okuma uzunluğu 99 nükleotit ve kütüphanelere ait toplam ortalama okuma sayısı 69 milyon civarında belirlendi. Çizelge 1'de referans genoma hizalama sonrası elde edilen tanımlayıcı istatistikler yer almaktadır. Gerek toplam okuma sayıları gerekse hizalama yüzdeleri açısından değerlendirildiğinde, transkriptom kütüphaneleri biyolojik çıkarımlar yapmak için yeterli okuma sayısı ve kaliteye sahip olarak belirlendi.

**Çizelge 1.** *Sequence Read Archive* veri tabanından elde edilen transkriptom kütüphanelerine ait toplam okuma sayıları ve insan referans genomuna hizalanma oranları.

Doku tipi	SRA veri tabanı erişim kimliği	Toplam okuma sayısı	Ortalama eşsiz hizalama oranı (%)
Adipoz	ERX011220	70844000	83.07
Adrenal bez	ERX011217	68143585	85.07
Akciğer	ERX011227	72141860	89.69
Beyin	ERX011200	42454601	89.49
Böbrek	ERX011182	72844954	83.60
İskelet kası	ERX011212	73466073	91.56
Kalp	ERX011197	74643822	87.43
Karaciğer	ERX011229	71109359	88.92
Kolon	ERX011204	75602271	86.97
Lenf nodu	ERX011193	73613850	86.37
Lökosit	ERX011195	68119332	91.10
Meme	ERX011225	58098948	85.14
Ovaryum	ERX011184	69769385	87.36
Prostat	ERX011205	74635206	90.68
Testis	ERX011203	73065173	89.24
Tiroid	ERX011194	72947915	88.20
Ortalama değer		69468770.88	87.74

Ensembl veri tabanından elde edilen gen anotasyonu verilerinde toplam 22298 pkRNA, 7842 ukRNA yer aldığı belirlendi. Bu genlerin edgeR paketi ile hesaplanan RPKM değerlerine göre yapılan filtreleme işlemi sonrası, 4061 ukRNA çalışmaya dahil edilirken, pkRNAların ise 11401 tanesi filtreleme koşulunu sağlamadığı için veri setinden çıkarıldı. Düşük oranlarda ifade edilen genlerin filtrelenmesi, gen ifadesi örüntülerinin istatistiksel olarak çözümlenmesinde sıklıkla kullanılan bir yöntem olmakla beraber, bu genlerin veri setinden uzaklaştırılmasının duyarlılığı anlamlı bir şekilde artırdığı daha önceki çalışmalarda gösterilmiştir (Sha et al., 2015). Bu bağlamda, çalışmanın bundan sonraki basamaklarında Ensembl veri tabanında tanımlanan tüm genler yerine belirlenen eşik değerinin üzerinde ifade edilen genler üzerine yoğunlaşmıştır.

Günümüzde pek çok ukRNA'nın dokuya özgü ifade biçimine sahip olduğu ve ifade edildikleri dokularda gen ifadesinin düzenlenmesinde rol oynadığı bilinmektedir (Gloss and Dinger, 2016). Ağustos 2016 tarihi itibarı ile Ensembl veri tabanında ukRNA olarak tanımlanmış genlerin, dokuya özgü ifade örüntülerinin ortaya konması için uygulanan benzeşme yayılması algoritması ile sadece bir doku tipinde ifade edilen ukRNA kümeleri belirlendi. Şekil 2'de 16 farklı insan dokusunun yalnızca birinde yüksek miktarlarda ifade edilen ukRNA kümeleri ısı haritası üzerinde görülmektedir.

Normalize edilmiş RPKM değerleri kullanılarak oluşturulan ısı haritasında, her bir sütun doku tipini ifade ederken her bir satır bir gene karşılık gelmektedir. Bunun yanında Şekil 2'de sağ alt köşede yer alan renk anahtarında da gösterildiği gibi koyu kırmızı yüksek gen ifade düzeyini simgelerken, mavi renkler düşük seviyeler de ya da sıfır seviyesinde gen ifadesini işaret etmektedir.

Benzeşme yayılması algoritması ile gerçekleştirilen kümeleme analizinde, tiroid ve testis dokularında az sayıda ukRNA'nın dokuya özgü ifade edildiği belirlenirken, diğer dokularda görece daha çok elemandan oluşan ukRNA kümelerinin var olduğu belirlendi. Şekil 2'nin sağ sütununda yer alan ısı haritasında karaciğerde yüksek düzeylerde ifade edildiği belirlenen ukRNA genleri gösterilmektedir.

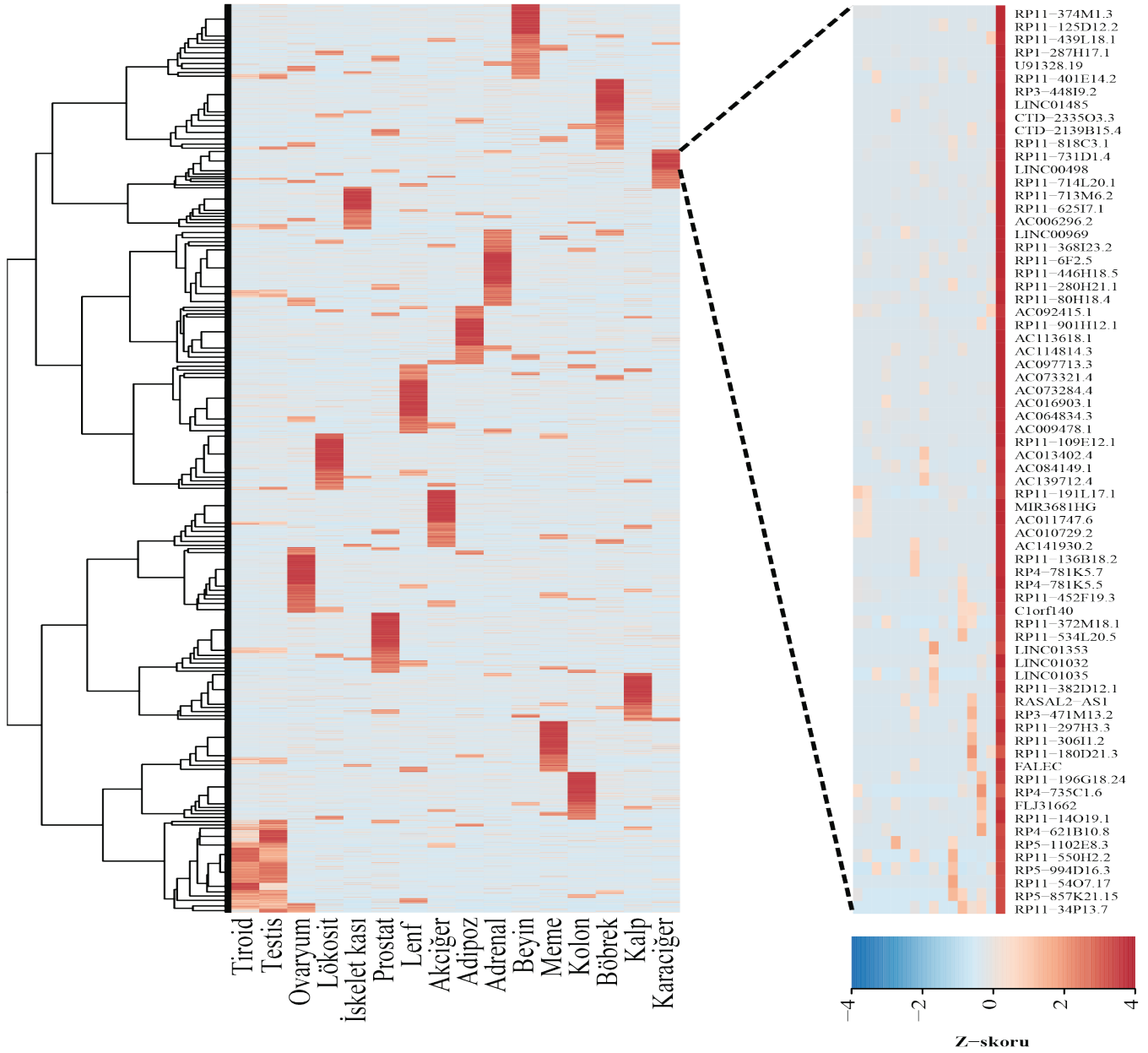
Mikrodizin ve yeni nesil dizileme yöntemleri gibi yüksek hacimli profillemeye verileri kullanılarak oluşturulan eş ifade ağları (=co-expression networks), son yıllarda özellikle işlevi iyi karakterize edilememiş genlerin biyolojik rollerinin belirlenmesinde sıklıkla kullanılmaktadır (Childs et al., 2011; Wong et al., 2014). Buradaki yaklaşım, benzer ifade profillerine sahip olan genlerin birlikte kümelenildiği ve aynı biyolojik yollarda yer alacağı üzerine kuruludur (Wong et al., 2014).

Aynı yaklaşım ile ukRNA-pkRNA birlikte ifade ağlarının oluşturulması ve ukRNAların hücre içerisindeki işlevlerinin tanımlanmasına yönelik çeşitli çalışmalar bulunmaktadır. Bu şekilde, şizofreni ve kanser gibi karmaşık hastalıklarda ukRNAların rolleri tahminlenmeye çalışılmış ve hastalıklarla olan olası ilişkileri ortaya konulmuştur (Ren et al., 2015). Bu çalışmada, benzer bir yaklaşımı *Illumina BodyMap 2.0* gen ifadesi verilerine uygulayarak, dokuya özgü ifade edilen ukRNAların hangi biyolojik olaylar ile ilişkili olduğu, benzeşme yayılması algoritması ile tahminlenmiş ve kapsamlı bir insan ukRNA anotasyon kataloğu oluşturulmuştur.

Yürütülen analizlerde 255 farklı ukRNA-pkRNA birlikte ifade kümesi belirlenirken, bu kümelerin her birinin ayrı ayrı GO terimi analizi sonrası toplam 2748 GO biyolojik işlev teriminin bu kümeler ile anlamlı bir şekilde ilişkili olduğu belirlenmiştir (FDR<0.05). Şekil 3'de ukRNA-pkRNA birlikte ifade kümelerinin ısı haritası gösterimi ile istatistiksel anlamlılık düzeyi en yüksek ilk 10 GO biyolojik süreç terimi gösterilmektedir. Gerçekleştirilen GO analizinde, ukRNAların gen ifadesinin düzenlenmesi (Geisler and Collier, 2013) ve gelişim ile ilişkili genlerin (Fatica and Bozzoni, 2014) yanı sıra öğrenme ve davranış gibi karmaşık biyolojik süreçlerde de rol alan genler ile de ilişkili olabileceği biyoformatik olarak ortaya konmuştur.

Ancak bulguların gerçekleşmesi için fonksiyonel genetik çalışmalar mutlaka gereklidir. Bu bağlamda çalışmada ortaya konan olası ukRNA işlevleri ile ukRNAlar ve hedeflendikleri protein kodlayan gen kümeleri test edilebilir hipotezlerin oluşturulmasında ukRNA araştırmacılarına yardımcı olabilecek potansiyeldedir.



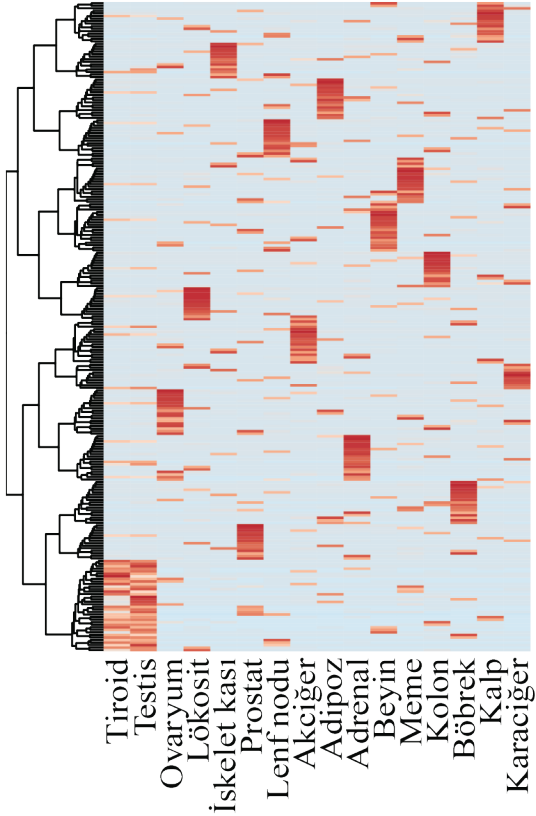


Şekil 2. İnsan dokularında ukrNA ifade profillerine ait ısı haritası ve karaciğere özgü ukrNAlar

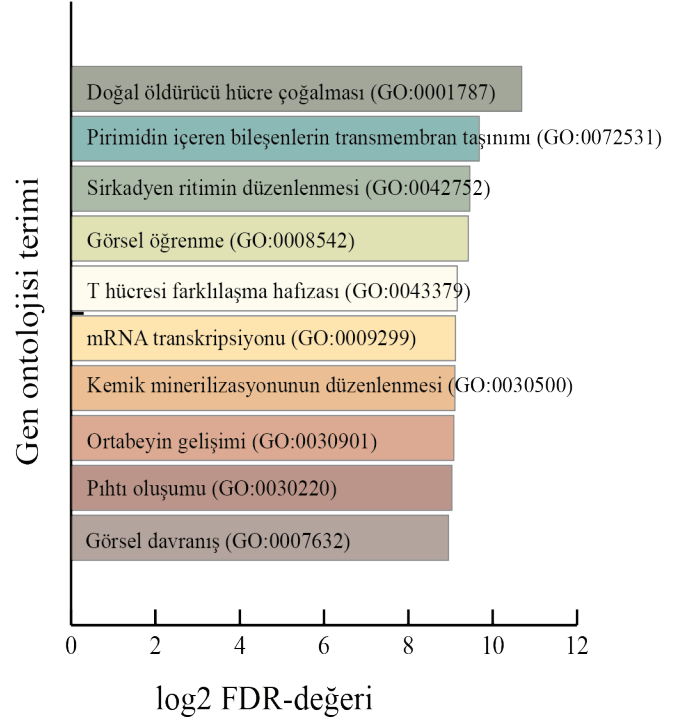
Şekil 4’de APCluster R paketi ile yapılan çözümlene sonrası 227 numaralı kümede yer alan ukrNAlar ve protein kodlayan genler arasındaki eş ifade ağı temsil edilmektedir. Açık kırmızı renk ile ifade edilen düğümler ukrNAları temsil ederken sarı renkli düğümler pkRNA genlerini simgelemektedir. Kırmızı ve sarı renkli ağ düğümleri arasındaki kenarlar ukrNA – pkRNA çifti arasında benzeşme yayılması

algoritması ile ortaya konan olası ilişkiyi ifade etmekte olup, ukrNAnın ilişkili olduğu protein kodlayan genini hedefleyebileceği ve ifadesini düzenleyebileceğini simgelemektedir. Bu çalışmada elde edilen verilere, mikroRNAlar gibi diğer düzenleyici genetik elementlerinde eklenmesi ile biyolojik süreçlere ait tamamlanmış ve bütüne yakın düzenleyici ağların oluşturulması mümkün olabilir.

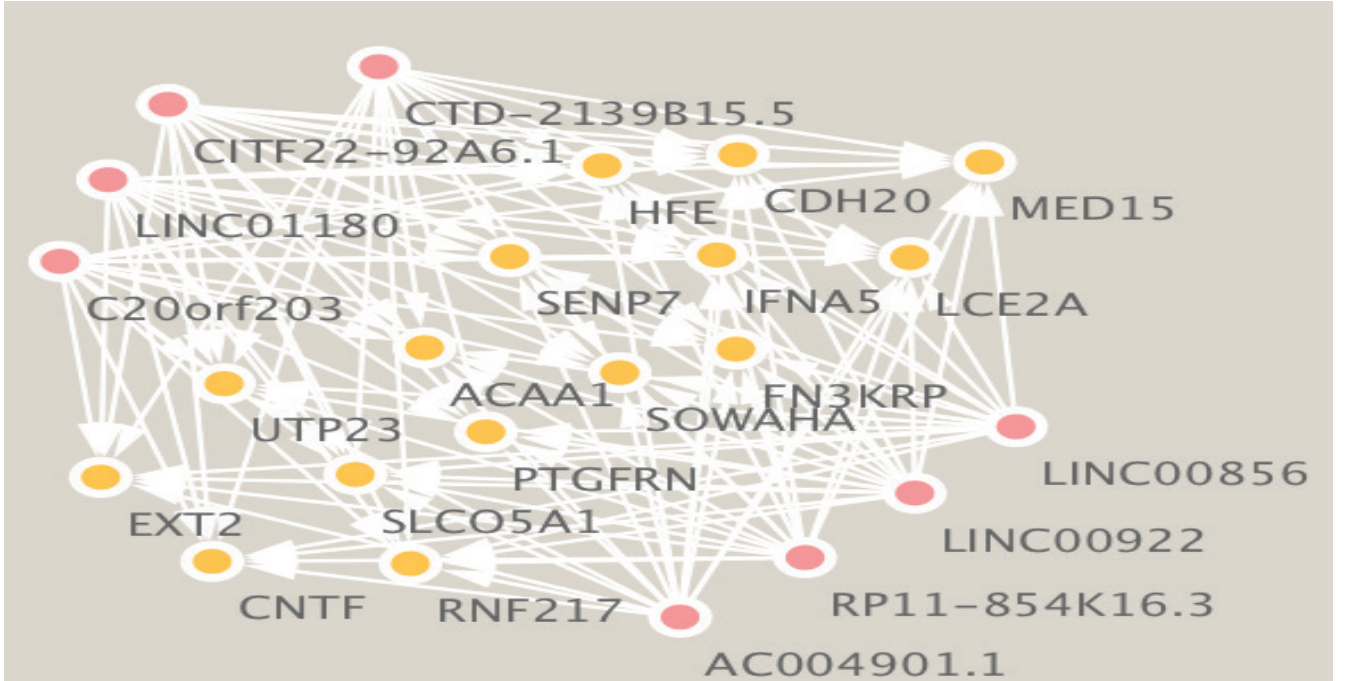
A



B



Şekil 3. ukRNA ve pkRNA ifade profillerine ait ısı haritası ve en yüksek istatistiksel anlamlılık düzeyine sahip ilk 10 GO terimi



Şekil 4. Benzeşme yayılması ile belirlenmiş örnek bir ukRNA-pkRNA eş ifade ağı

## SONUÇ

Bu çalışmada farklı insan dokularından RNA dizileme yöntemi ile elde edilen gen ifadesi örüntüleri, benzeşme yayılması algoritması ile irdelenmiş, birlikte ifade edilen ukRNA ve protein kodlayan gen kümeleri tanımlanmıştır. Elde edilen gen kümeleri doğrultusunda biyolojik işlevleri karakterize edilememiş pek çok ukRNA için biyoenformatik yöntemler aracılığı ile olası biyolojik roller tanımlanmıştır.

Çalışmada, yüksek işlem hacimli moleküler biyolojik yöntemlerin ve bilgisayar tabanlı işlemsel yöntemlerin doğası gereği çok sayıda çıktı elde

edilmiştir. Bu bağlamda, (i) dokuya özgü ifade edildiği tahminlenen ukRNA kümeleri ve ilgili ukRNA Ensembl veri tabanı tanımlayıcıları, (ii) ukRNA ve pkRNA birlikte ifade kümeleri ve (iii) bu kümelerin ilişkili oldukları biyolojik süreçler ve bunların istatistiksel anlamlılık düzeyleri <https://goo.gl/hIjF3s> adresinden herhangi bir kısıt olmadan indirilebilir. İnternet üzerinden erişilebilir bu kapsamlı kataloğun, sağlık ve temel bilim araştırmalarında hipotez oluşturmada ve bulguların desteklenmesinde kullanılabileceği düşünülmektedir. Ayrıca bu çalışmada kullanılan yaklaşımın fare, zebra balığı gibi diğer model organizmalardaki ukRNAların annotasyonu içinde kullanılması olasıdır.

## KAYNAKLAR

- Andrews S, 2010. FastQC: a quality control tool for high throughput sequence data. from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. (Erişim tarihi: 1 Ağustos, 2016)
- Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S S, Eppig J T, Harris M A, Hill D P, Issel-Tarver L, Kasarskis A, Lewis S, Matese J C, Richardson J E, Ringwald M, Rubin G M, Sherlock G, 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1): 25-29.
- Baş Ç, İkizler-Cinbis N, 2013. Comparison of clustering methods for pose based video summarization. In *Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4. IEEE.
- Bodenhofer U, Kothmeier A, Hochreiter S, 2011. APCluster: an R package for affinity propagation clustering. *Bioinformatics*, 27(17): 2463-2464.
- Bolger A M, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120.
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd D H, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams S R, Moon K, Burcham T, Pallas M, DuBridgde R B, Kirchner J, Fearon K, Mao J, Corcoran K, 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, 18(6): 630-634.
- Cao H, Amendt B A, 2016. pySAPC, a python package for sparse affinity propagation clustering: Application to odontogenesis whole genome time series gene-expression data. *Biochim Biophys Acta*.
- Caso G, de Nardis L, di Benedetto M G, 2015. A Mixed Approach to Similarity Metric Selection in Affinity Propagation-Based WiFi Fingerprinting Indoor Positioning. *Sensors*, 15(11): 27692-27720.
- Chaitankar V, Karakülah G, Ratnapriya R, Giuste F O, Brooks M J, Swaroop A, 2016. Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog Retin Eye Res*.
- Childs K L, Davidson R M, Buell C R, 2011. Gene Coexpression Network Analysis as a Source of Functional Annotation for Rice Genes. *PLoS One*, 6(7).
- D'Haeseleer P, 2005. How does gene expression clustering work? *Nat Biotechnol*, 23(12): 1499-1501.
- Dobin A, Davis C A, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras T R, 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1): 15-21.
- Fatica A, Bozzoni I, 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*, 15(1): 7-21.
- Frey B J, Dueck D, 2007. Clustering by passing messages between data points. *Science*, 315(5814): 972-976.
- Geisler S, Collier J, 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol*, 14(11): 699-712.
- Gloss B S, Dinger M E, 2016. The specificity of long noncoding RNA expression. *Biochim Biophys Acta*, 1859(1): 16-22.
- Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyraas E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M, 2002. The Ensembl genome database project. *Nucleic Acids Res*, 30(1): 38-41.
- Jiang J, Huang J, Wang X R, Quan Y H, 2016. Investigating key genes associated with ovarian cancer by integrating affinity propagation clustering and mutual information network analysis. *Eur Rev Med Pharmacol Sci*, 20(12): 2532-2540.
- Kerr G, Ruskin H J, Crane M, Doolan P, 2008. Techniques for clustering gene expression data. *Comput Biol Med*, 38(3): 283-293.
- Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C, 2011. The sequence read archive. *Nucleic Acids Res*, 39(Database issue): D19-21.



- Li K M, Guo L, Li G, Nie J X, Faraco C, Zhao Q, Miller L S Liu T M, 2010. Cortical Surface Based Identification of Brain Networks Using High Spatial Resolution Resting State Fmri Data. 2010 7th Ieee International Symposium on Biomedical Imaging: From Nano to Macro: 656-659.
- Li X, Wang, H X, 2015. Identification of functional networks in resting state fMRI data using adaptive sparse representation and affinity propagation clustering. *Frontiers in Neuroscience*, 9.
- Meng J, Hao H, Luan Y, 2016. Classifier ensemble selection based on affinity propagation clustering. *J Biomed Inform*, 60: 234-242.
- Reimand J, Kolde R, Arak M T R, Curl I, 2016. Package 'gProfileR'.
- Ren Y, Cui Y, Li X, Wang B, Na L, Shi J, Wang L, Qiu L, Zhang K, Liu G Xu Y, 2015. A co-expression network analysis reveals lncRNA abnormalities in peripheral blood in early-onset schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry*, 63: 1-5.
- Robinson M D, McCarthy D J, Smyth G K, 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1): 139-140.
- Schena M, Shalon D, Davis R W, Brown P O, 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235): 467-470.
- Sha Y, Phan J H, Wang M D, 2015. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. *Conf Proc IEEE Eng Med Biol Soc*, 2015: 6461-6464.
- Sun C X, Huo H W, Yu Q, Guo H T, Sun Z G, 2015. An Affinity Propagation-Based DNA Motif Discovery Algorithm. *Biomed Research International*.
- Team R C (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013, ISBN 3-900051-07-0.
- Wong D C J, Sweetman C, Ford C M, 2014. Annotation of gene function in citrus using gene expression information and co-expression networks. *Bmc Plant Biology*, 14.
- Zhang J, Tuo X G, Yuan Z, Liao W, Chen H F, 2011. Analysis of fMRI Data Using an Integrated Principal Component Analysis and Supervised Affinity Propagation Clustering Approach. *Ieee Transactions on Biomedical Engineering*, 58(11): 3184-3196.
- Zhang T, Wu R B, 2015. Affinity Propagation Clustering of Measurements for Multiple Extended Target Tracking. *Sensors*, 15(9): 22646-22659.
- Zhao X L, Xu W X, 2015. An Extended Affinity Propagation Clustering Method Based on Different Data Density Types. *Computational Intelligence and Neuroscience*.