



MikroRNA Ekspresyon Profillemesinde Yaygın Kullanılan Normalizasyon Yaklaşımları

Ali Osman TURGUT^{1,a}, Özgecan KORKMAZ AĞAOĞLU^{2,b}

¹Burdur Mehmet Akif Ersoy Üniversitesi, Veteriner Fakültesi, Zootekni Anabilim Dalı, Burdur-TÜRKİYE

²Burdur Mehmet Akif Ersoy Üniversitesi, Veteriner Fakültesi, Genetik Anabilim Dalı, Burdur-TÜRKİYE

ORCID : ^a0000-0001-6863-0939; ^b0000-0002-7414-1725

Corresponding author: Ali Osman TURGUT; E-posta: aoturgut@mehmetakif.edu.tr

How to cite: Turgut AO, Korkmaz Ağaoğlu Ö. MikroRNA ekspresyon profillemesinde yaygın kullanılan normalizasyon yaklaşımları. Erciyes Univ Vet Fak Derg 2022; 19(2): 152-159

Öz: MikroRNA (miRNA) ekspresyonlarının belirlenmesinde RT-qPCR, mikroarray platformları ve miRNA dizileme en yaygın kullanılan tekniklerdir. Tüm bu tekniklerin kullanıldığı çalışmalarda en önemli hususlardan biri verilerin uygun normalizasyon yöntemi ile normalize edilmesidir. Normalizasyon ile biyolojik ve teknik varyasyonların sonuçlar üzerine olan etkisinin elimine edilmesi amaçlanmaktadır. MiRNA ekspresyonu çalışmalarında, farklı tekniklerden elde edilen verilerin normalizasyonunda kullanılan çok sayıda normalizasyon yaklaşımı kullanılmaktadır. Bu derlemede, miRNA ekspresyonu çalışmalarında en yaygın kullanılan normalizasyon yaklaşımları hakkında bilgiler özetlenmiştir.

Anahtar kelimeler: Mikroarray, mikroRNA, miRNA dizileme, normalizasyon, RT-qPCR

Commonly Used Normalization Approaches in MicroRNA Expression Profiling

Abstract: RT-qPCR, microarray platforms and miRNA sequencing are the most common techniques used to determine microRNA (miRNA) expressions. One of the most important issues in studies these techniques are used is the normalization of the data by using appropriate normalization method. The purpose of normalization is to eliminate the effects of biological and technical variations on study results. Numerous normalization approaches are used for normalization of data obtained from different techniques in miRNA expression studies. In this review, information about the most commonly used normalization approaches in miRNA expression studies is summarized.

Keywords: Microarray, microRNA, miRNA sequencing, normalization, RT-qPCR

Giriş

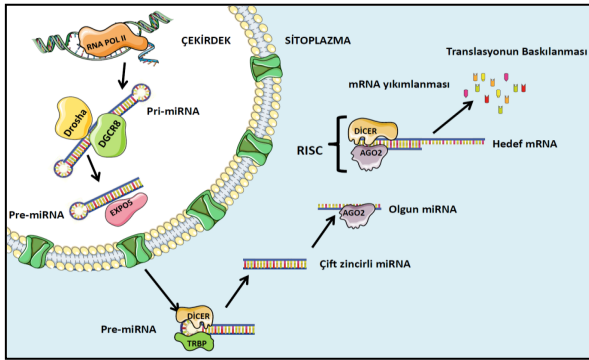
miRNA'lar; yaklaşık 18-22 nükleotid uzunluğunda, gen ekspresyonlarının transkripsiyon sonrası baskılanmasında rol oynayan ve kodlama yapmayan RNA sınıfına dahil olan RNA'lardır (Faraldi ve ark., 2018). MiRNA'lar hedef genlerin ekspresyonunu düzenleyerek önemli biyolojik yollarda etkili olmaktadır (Qureshi ve Sacan, 2013; Sidekli ve Korkmaz Ağaoğlu, 2019). Ayrıca, miRNA'ların değişen ekspresyon düzeyleri kanser ve bazı önemli patolojik olaylarla da ilişkili olabilmektedir (Heneghan ve ark., 2010; Lange ve ark., 2017). MiRNA gen ekspresyonlarının belirlenmesinde gerçek zamanlı kantitatif polimeraz zincir reaksiyonu (real time quantitative polymerase chain reaction; RT-qPCR), miRNA mikroarray platformları ve miRNA dizileme tekniklerinden yararlanılmaktadır (Saliminejad ve ark., 2019). Bu tekniklerden elde edilen sonuçların güvenilir olabilmesi için uygun normalizasyon yaklaşımı ile verilerin normalize edilmesi gerekmektedir. Normalizasyon, biyolojik varyasyonların ve deneysel uygulamalar sırasında oluşan farklılıkların sonuçlara olan etkisini en az düzeye indirge-

mek için gereklidir (Bustin ve ark., 2009). Bu nedenle; gen ekspresyonu çalışmalarında normalizasyon amacıyla çok sayıda yaklaşım geliştirilmiştir (Livak ve Schmittgen, 2001; Qureshi ve Sacan; 2013; Tam ve ark., 2015). Bu yaklaşımlardan bazıları kullanılan moleküler tekniğe özeldir (Livak ve Schmittgen, 2001; López-Romero ve ark., 2010; Robinson ve ark., 2010). Bazı normalizasyon yaklaşımları ise, birden fazla teknikte bazı modifikasyonlarla kullanılabilir (Bolstad ve ark., 2003; Ritchie ve ark., 2015; Tam ve ark., 2015). Bu derlemede, miRNA gen ekspresyonu çalışmalarında kullanılan moleküler teknikler ve normalizasyon yaklaşımları hakkında bilgiler verilmiştir.

miRNA'ların biyogenezi

miRNA'lar kodlama yapmayan RNA'lardır. Bu sınıftaki RNA'lar gen ekspresyonlarının transkripsiyon sonrası düzenlenmesinde rol oynamaktadır (Faraldi ve ark., 2018). Biyogenezde ilk olarak; miRNA gen bölgesi RNA polimeraz II enzimi tarafından transkripsiyona uğramakta ve stem-loop (gövde ilmek) yapıda primer transkript (pri-miRNA) sentezlenmektedir. Pri-miRNA'lar; çekirdekte bulunan Drosha enzimi ve onun kofaktörü (DiGeorge kritik sendrom bölgesi 8; DiGeorge syndrome critical region 8; DGCR8) tara-

findan kesilerek, yaklaşık 60 baz uzunluğunda prekürsör-miRNA'ya (precursor-miRNA; pre-miRNA) dönüştürülmektedir (Lee ve ark., 2003; Faraldi ve ark., 2018). Oluşan miRNA'lar çekirdek porlarından sitoplazmaya, Ekspörtin 5-Ran-GTP kompleksi aracılığıyla taşınmaktadır (Lund ve ark., 2004). Stoplazmaya taşınan pre-miRNA'lar, Dicer enzimi tarafından kesilerek yaklaşık 22 nükleotit uzunluğundaki çift zincirli (double-stranded RNA; dsRNA) miRNA'lara dönüştürülmektedir (Faraldi ve ark., 2018). Oluşan dsRNA'lar; helikaz enzimi tarafından tek zincirli hale dönüştürülmektedir. Olgun miRNA stem-loop yapısının 5' kolundan sentezlenirse "5p", 3' kolundan sentezlenirse "3p" ifadesi miRNA isminin sonuna eklenmektedir. Tek zincir yapıdaki olgun miRNA'lar; RNA ile uyarılan susturma kompleksine (RNA induced silencing complex; RISC) bağlanmaktadır. RISC; Dicer proteini ve mRNA'yı hedef alan argonat-2 (AGO2) proteininden oluşmaktadır (Faraldi ve ark., 2018). RISC; tam anlamıyla bir eşleşme olması durumunda mRNA'yı parçalamaktadır. Kısmi bir eşleşme olması durumunda ise RISC; hedef mRNA'nın translasyonunu baskılamaktadır (Qureshi ve Sacan, 2013). MiRNA'ların hedef mRNA'larla eşleşmesi çoğunlukla; 5' uçta yer alan 2-8 baz uzunluğundaki çekirdek (seed) dizilerin eşleşmesiyle olmaktadır. Bu çekirdek dizilerin birden fazla genin mRNA'sı ile eşleşebilmesi nedeniyle tek bir miRNA çok sayıda genin ekspresyonunu etkileyebilmektedir (Hammond, 2015).



Şekil 1. miRNA'ların biyogenezini. Şekil, Servier Medical Art görselleri kullanılarak hazırlanmıştır (<https://smart.servier.com>).

miRNA ekspresyonu analiz yöntemleri

1- RT-qPCR

RT-qPCR miRNA gen ekspresyon düzeylerinin belirlenmesinde yaygın olarak kullanılan bir tekniktir (Mestdagh ve ark., 2009; Sidekli ve Korkmaz Ağaoğlu, 2021). MiRNA'ların az sayıda nükleotit içermeleri; miRNA'ların izolasyonu, kalite ve kantite kontrolleri, komplementer DNA (complementary DNA; cDNA) sentezi ve PCR reaksiyonu gibi aşamalarında farklı yaklaşımların ele alınmasına neden olmuştur

(Redshaw ve ark., 2013; Schwarzenbach ve ark., 2015; Sidekli ve Ağaoğlu, 2021). RT-qPCR tekniğinde, özgün olmayan floresan boyalar veya özgün prob dizaynı ile miRNA'ların kantifikasyonu gerçekleştirilmektedir. RT-qPCR reaksiyonu sonrası elde edilen veriler, hedef bölgenin yükseltgenmesiyle açığa çıkan floresan ışımının tespit edilebildiği siklus eşik değerleridir (cycle threshold; Ct). MiRNA çalışmalarında 35'ten küçük Ct değerleri dikkate alınmaktadır. Bunun sebebi tek bir miRNA kopyası olması durumunda dahi floresan sinyalinin en fazla 35. PCR siklusunda tespit edilebilir düzeye ulaşmasıdır (Mestdagh ve ark., 2009). Elde edilen ham Ct değerleri referans (housekeeping) genlerin kullanımı (Izumi ve ark., 2012) veya diğer normalizasyon yaklaşımları ile normalize edilerek verilerin istatistiksel analizi gerçekleştirilmektedir (Mestdagh ve ark., 2009; Qureshi ve Sacan, 2013).

2- Mikroarray platformları

Gen ekspresyonu çalışmalarında, cDNA mikroarray platformları ve oligonükleotid mikroarray platformları olmak üzere iki farklı mikroarray platformu kullanılmaktadır (Mandruzzato, 2007). Gen ekspresyonu çalışmalarında kullanılan mikroarray platformları DNA array, DNA çip, biyo çip ve gen çip olarak da isimlendirilmektedir (Gershon, 2002). Bu mikroarray platformları belirli bir hücre veya doku örneğinde ekspresyon olan çok sayıda genin ekspresyon düzeyinin belirlenmesine olanak tanımaktadır. Mikroarray platformlarında çapı 200 µm'den küçük olan binlerce gözenek bulunmaktadır. Bu gözenekler cDNA veya oligonükleotid problemlerini içermektedir ve özel lazer tarayıcılarla okunmaktadır (Gershon, 2002). Mikroarray platformlarında tarayıcı tarafından okunan floresan ışımaya değerleri farklı normalizasyon yaklaşımları ile normalize edildikten sonra gen ekspresyon değişimleri analiz edilmektedir. MiRNA mikroarray platformlarının en büyük dezavantajı inceleme de yüksek miktarda RNA örneğine ihtiyaç duyulmasıdır. Prob sinyallerinin güvenilir bir şekilde değerlendirilebilmesi için yaklaşık 10 µg total RNA örneğine ihtiyaç duyulduğu belirtilmektedir (Mandruzzato, 2007).

3- miRNA dizileme tekniği

MiRNA dizileme (miRNA sequencing; miRNA-seq), yeni nesil sekanslama temelli bir yaklaşımdır. Bu yaklaşımda analiz, diğer yöntemlerden farklı olarak genellikle zenginleştirilmiş miRNA'lar ile gerçekleştirilmektedir (Saliminejad ve ark., 2019). Bu yöntem pahalı bir yöntem olmasının yanı sıra veri analizi ve yorumlamasındaki zorluklar nedeniyle bazı dezavantajları bulunmaktadır (Pritchard ve ark., 2012). Ancak yeni nesil sekanslama platformlarının sayısının artması bu yöntemin maliyetinin düşmesini ve miRNA ekspresyon profillerinin değerlendirilmesinde miRNA-seq tekniğinin daha yaygın kullanımını mümkün hale getirmektedir (Tam ve ark., 2015). MiRNA dizileme

platformlarında genel analiz şeması; kısa okumaların belirlenmesi, düşük kaliteli okumaların ve adaptör dizilerinin filtrelenmesi, filtrelenen dizilerin bir referans dizi ile tespit edilmesi (bu aşamada miRBase veya genom veri tabanlarından yararlanılmaktadır), miRNA ekspresyon düzeyinin belirlenmesi, verilerin normalize edilmesi ve miRNA biyolojik fonksiyonlarına ilişkin analizlerin gerçekleştirilmesi aşamalarından oluşmaktadır (Tam ve ark., 2015). MiRNA dizileme yeni miRNA'ların keşfinin yanısıra (Metpally ve ark., 2013), miRNA mutasyonlarının tespitine de olanak tanımaktadır (Pritchard ve ark., 2012).

RT-qPCR verilerinin normalizasyonu

RT-qPCR çalışmalarında; örnek toplama ve hazırlanması, RNA izolasyonu, PCR etkinliği ve ters transkripsiyon reaksiyonu sırasındaki farklılıklar ve hatalar sonuçların güvenilirliği açısından risk teşkil etmektedir. Bu hataları en az düzeye indirmek için uygun bir normalizasyon metodu ile elde edilen verilerin normalize edilmesi gerekmektedir (Bustin ve ark., 2009). Veri normalizasyonu sadece mRNA'lar değil, miRNA gen ekspresyonu çalışmalarında da gereklidir (Meyer ve ark., 2010). MiRNA RT-qPCR verilerinin normalizasyonunda farklı yaklaşımlar kullanılmaktadır (Livak ve Schmittgen, 2001; Mestdagh ve ark., 2009; Qureshi ve Sacan, 2013).

Referans genler ile normalizasyon

RT-qPCR çalışmalarında normalizasyon amacıyla referans genlerin kullanımı altın standart olarak kabul edilmektedir (Faraldi ve ark., 2018). İdeal bir referans genin farklı dokularda, patolojik koşullarda, hücre döngüsü aşamalarında ve fizyolojik süreçlerde ekspresyonunun stabil olması sahip olması istenmektedir (Dheda ve ark., 2005). RT-qPCR çalışmalarında kullanılan referans genler genel olarak temel hücre fonksiyonlarının sürdürülebilmesi için gerekli olan, fizyolojik ve patolojik koşullarda ekspresyonu stabil olan genlerdir (Zhu ve ark., 2008). MiRNA RT-qPCR verilerinin normalizasyonunda da referans gen yaklaşımı yaygın olarak kullanılmaktadır (Schwarzenbach ve ark., 2015). RT-qPCR çalışmalarında normalizasyon işleminden önce çalışılan dokuda referans gen adaylarının stabilitesi farklı bilgisayar programları ile değerlendirilmektedir (Korkmaz Ağaoğlu ve Sidekli, 2020; Mestdagh ve ark., 2009; Qureshi ve Sacan, 2013). Bu amaçla en yaygın olarak kullanılan programlar; geNorm, NormFinder ve BestKeeper programlarıdır (Faraldi ve ark., 2018). MiRNA çalışmalarında çalışılan dokuda belirlenen en stabil olan referans genler ile verilerin normalizasyonu yoluna gidilmektedir (Faraldi ve ark., 2018; Inyawilert ve ark., 2015; Zhang ve ark., 2017).

MiRNA RT-qPCR verilerinin normalizasyonunda endojen referans genler veya sentetik spike-in miRNA'lar kullanılmaktadır. Normalizasyon amacıyla en yaygın kullanılan endojen referans genler; miR-93,

miR-425 (Zalewski ve ark., 2017), miR-191 (Zheng ve ark., 2013), miR-320, miR-16 (Schwarzenbach ve ark., 2015), miR-221, miR-26a (Zhu ve ark., 2012), küçük kodlama yapmayan RNA genleri (RNU6, RNU43, RNU44, RNU48) (Inyawilert ve ark., 2015; Donati ve ark., 2019) ve ribozomal RNA genleridir (5S rRNA ve 18S rRNA) (Montenegro ve ark., 2009; Zhang ve ark., 2017). Yaygın olarak kullanılan spike-in miRNA'lar ise; *Coehabtidis Elegans* (cel) miR-39, cel-miR-54, cel-miR-238'dir (Izumi ve ark., 2012).

RT-qPCR çalışmalarında genlerin mutlak (absolute) ve relatif (relative) ekspresyon düzeyleri belirlenmektedir. Ancak, RT-qPCR çalışmalarının çoğunda genlerin relatif ekspresyon düzeyi saptanmaktadır (Schwarzenbach ve ark., 2015). Relatif ekspresyon düzeyi deneysel bir gruptaki hedef bölgenin PCR sinyalinin, kontrol grubundaki PCR sinyaline kıyaslanması ile belirlenebilmektedir (Livak ve Schmittgen, 2001). Livak ve Schmittgen (2001) tarafından geliştirilen delta delta Ct ($\Delta\Delta Ct$) matematiksel modeli genlerin relatif ekspresyon düzeylerinin belirlenmesinde en yaygın kullanılan modeldir. Bu modelde relatif ekspresyon; $\Delta Ct = Ct$ (hedef gen) - Ct (referans gen), $\Delta\Delta Ct = \Delta Ct$ (çalışma grubu) - ΔCt (kontrol grubu) formülü ile belirlenmektedir. Çalışılan bir genin kontrol grubuna kıyasla ekspresyon düzeyinin değişimi ise kat değişimi (fold change; FC) olarak isimlendirilmektedir. Bir genin ekspresyon düzeyi değişimi; $FC = 2^{-\Delta\Delta Ct}$ formülasyonu ile hesaplanmaktadır (Livak ve Schmittgen, 2001). MiRNA gen ekspresyonlarının araştırıldığı çalışmalarda bu matematiksel model sıklıkla kullanılmaktadır (Inyawilert ve ark., 2015; Zhang ve ark., 2017).

Global ortalama normalizasyonu

MiRNA RT-qPCR verilerinin normalizasyonunda kullanılan diğer bir yöntem; global ortalama normalizasyon (global mean normalization) yöntemidir. Bu yöntemde, bir örnek içerisindeki tüm genlerin Ct değerleri ortalaması endojen kontrol olarak kullanılmaktadır. ΔCt değeri, genlerin Ct değerlerinin bu ortalama Ct değerinden çıkarılmasıyla elde edilmektedir (Mestdagh ve ark., 2009). Global ortalama yaklaşımında yüksek Ct değerlerine sahip genler olması durumunda varyasyonun artacağı ve örnek içerisinde çok düşük miktarlarda bulunan miRNA'ların kantifikasyonunda güvenilir olmadığı belirtilmektedir (Qureshi ve Sacan, 2013).

Ağırlıklı ortalama normalizasyonu

Ağırlıklı ortalama (weighted mean) normalizasyonu Ct varyasyonlarını da normalizasyona dahil eden bir yaklaşımdır. Bu yaklaşımda, tüm örneklerde bir genin Ct değerinin standart sapmasının tersi alınarak ağırlıklı ortalama gücü (weighted mean power; wmp) hesaplanmaktadır. Bu değer tüm genler için standart sapma değerlerinin tersinin toplamına bölünmektedir. Bu değer ile bir örnekteki tüm ham Ct değerlerinin

toplamı çarpılarak Ct_0 değeri hesaplanmaktadır. Ct_0 değeri;

$$Ct_0 = \sum_j^n x \left(\frac{1}{STD(Ct_j)} \right) / \sum_{i=1}^n 1/STD(Ct_i)$$

formülasyonuna göre belirlenmektedir (Qureshi ve Sacan, 2013). ΔCt değeri ise, her bir genin Ct değerinden Ct_0 değerinin çıkarılmasıyla hesaplanmaktadır. Bu yöntemin, standart sapma değeri düşük olan genlerin normalizasyondaki ağırlığını artırdığı ifade edilmektedir. Ayrıca düşük ve yüksek ekspresyon düzeyine sahip miRNA'ların kantifikasyonunda normalizasyon amacıyla kullanılabileceği bildirilmiştir (Qureshi ve Sacan, 2013).

Mikroarray verilerinin normalizasyonu

Mikroarray platformlarında; RNA örneklerinin hazırlanması, floresan işaretleme, hibridizasyon, yıkama aşamasındaki uygulamalar ve bazı teknik problemler sonuçlarda farklılıklara neden olabilmektedir (Meyer ve ark., 2010). Bu nedenle, miRNA mikroarray verilerindeki bu farklılıkları en az düzeye indirmek için veriler normalize edilerek sonuçlar değerlendirilmektedir. Bu amaçla yaygın kullanılan normalizasyon yaklaşımları; toplam gen sinyali, üst çeyrek normalizasyonu, kuantil normalizasyon, medyan normalizasyonu, RMA (robust microarray algorithm) ve siklik loess (cyclic locally estimated scatterplot smoothing veya cyclic local regression) yaklaşımıdır.

Toplam gen sinyali

Oligonükleotid problemlerinde genelde her bir miRNA'nın birkaç farklı kopyası bulunmaktadır. Agilent mikroarray platformlarında bu birkaç kopyanın floresan sinyallerinin ortalaması alınarak toplam gen sinyali değeri (total gene signal; TGS) hesaplanmaktadır (López-Romero ve ark., 2010). Oligonükleotid mikroarray platformlarındaki prob hücreleri, iki farklı oligonükleotid probu (perfect match; PM; mis-match; MM) içermektedir. Aynı hücredeki bu iki farklı prob sadece bir baz farklılığa (13. baz) sahiptir (Mandrzzato, 2007). TGS, PM floresan sinyali değerinin MM floresan sinyali değerinden çıkarılması ile hesaplanmaktadır. Hesaplanan bu TGS direkt olarak istatistiksel analizde kullanılabilir (López-Romero ve ark., 2010). Ancak bu değer bazı miRNA problemlerinde MM floresan sinyalinin PM floresan sinyalinden yüksek olması nedeniyle negatif değer alabilmektedir. Böyle durumda, minimum TGS değerinin eklenmesiyle negatif değerler pozitif hale gelmektedir. Bu aşamadan sonra elde edilen veriler log₂ tabanında değerlere dönüştürülerek değerlendirilmektedir (López-Romero ve ark., 2010).

Üst çeyrek normalizasyonu

Bu yöntemde TGS değerlerindeki 75. yüzdellik değer

(75th percentile) baz alınarak normalizasyon gerçekleştirilmektedir. MiRNA mikroarray platformlarının kullanıldığı çalışmalarda 75. yüzdellik değer üst çeyrek (upper quartile) veya üçüncü çeyrek (third quartile; Q3) olarak da isimlendirilmektedir (Pradervand ve ark., 2009). Her bir miRNA için elde edilen TGS sinyallerinin bu değere bölünmesiyle veriler normalize edilmektedir. Normalizasyon sonrası veriler log₂ tabanına dönüştürülerek sonuçlar değerlendirilmektedir. Bu yaklaşımın normalize edilmemiş TGS yaklaşımına kıyasla, verilerin standart sapma değerini azalttığı ve sonuçlarının güvenilirliğini artırdığı ifade edilmektedir (Pradervand ve ark., 2009; López-Romero ve ark., 2010).

Kuantil normalizasyon

Kuantil normalizasyonda tüm mikroarray platformunda prob floresan sinyal yoğunluğunun değişmeyeceği varsayılmaktadır. Bu nedenle elde edilen floresan sinyal yoğunlukları kuantil normalizasyon yöntemiyle eşitlenerek elde edilen yeni veriler istatistiksel analizde kullanılmaktadır (Bolstad ve ark., 2003; Meyer ve ark., 2010). Normalizasyon yapılırken; p sayıda miRNA ve n sayıda örneğin floresan sinyali verilerini içeren X matrisi ($p \times n$) oluşturulur (Rao ve ark., 2008). Bu matriste her örnek için sıralama en küçükten en büyüğe doğru gerçekleştirilmektedir. Birinci sırada yer alan floresan sinyali en küçük floresan sinyalidir (Ballman ve ark., 2004). Daha sonra X matrisinin her satırında yer alan değerlerin ortalaması alınarak yeni değerler oluşturulur ki bu değerlerin tüm örneklerde aynı olduğu kabul edilir (Rao ve ark., 2008). Kuantil normalizasyon işlemi, Bioconductor gibi R programlama ile aynı dili kullanan bilgisayar programları ile gerçekleştirilebilmektedir (Liu ve ark., 2008). Diğer normalizasyon yöntemleri ile kıyaslandığında bu yöntemde normalizasyonun daha hızlı yapılabildiği ifade edilmektedir (Ballman ve ark., 2004).

Medyan normalizasyonu

Medyan normalizasyon yönteminde; p sayıda miRNA ve n sayıda örneğin floresan sinyali verilerini $p \times n$ matrisi oluşturulmaktadır. Bu matriste yer alan her kolondaki medyan değerleri alınarak n-boyutlu medyan vektörü M oluşturulmaktadır. Daha sonra bu M vektöründe de medyan değeri (overall median) belirlenmektedir. M vektörü medyan değeri ile her örneğin medyan değeri arasındaki fark belirlenmekte ve bu fark miRNA'nın ekspresyon değerinden çıkartılarak veriler normalize edilmektedir (Rao ve ark., 2008). Ancak, medyan normalizasyon yaklaşımının miRNA ekspresyonlarındaki global değişimlerin belirlenmesinde yetersiz kaldığı ifade edilmektedir (Wu ve ark., 2013).

RMA

Oligonükleotid temelli mikroarray platformlarından elde edilen verilerin normalizasyonu amacıyla kullanı-

lan RMA algoritması, genlerin ekspresyon seviyelerinin daha kesin bir şekilde tespit edilmesine olanak tanımaktadır (Wu ve ark., 2013). Genellikle Affymetrix GeneChip mikroarray platformlarında kullanılan standart bu algorithmada, elde edilen anormal değerler çıkarılmakta ve sadece pozitif sinyal değerleri dikkate alınmaktadır (Irizarry ve ark., 2003). RMA algoritmasındaki son adımda veriler kuantil normalizasyon yöntemi ile normalize edilmektedir (Bolstad ve ark., 2003; Wu ve ark., 2013). Bu algoritmanın aynı zamanda Agilent mikroarray platformlarından elde edilen TGS verilerinin arka plan sinyallerinin düzeltilmesi amacıyla kullanılabilceği bildirilmektedir (López-Romero ve ark., 2010).

Siklik loess

MiRNA mikroarray verilerinin normalizasyonunda kullanılan diğer yöntem ise, siklik loess yöntemidir (Rao ve ark., 2008). Bu yöntemde ilk olarak iki farklı örnek seçilerek MA plotu (M versus A plot) verisi oluşturulmaktadır. Bu plotta M; bir miRNA probu için iki örneğin ekspresyon değerleri arasındaki farktır. A ise iki ekspresyon değerinin ortalamasıdır. Her prob için oluşturulan M ve A değerleri regresyonu ile bir eğrisi oluşturulur (Rao ve ark., 2008). Bu eğriden elde edilen düzeltme faktörü ile verilerin normalizasyonu gerçekleştirilmektedir (Rao ve ark., 2008; Ballman ve ark., 2004). Tüm problemler için bu ikili işlem yapılarak analiz tamamlanmaktadır. Ancak siklik loess yönteminde aynı anda iki örnek için normalizasyon gerçekleştirilmektedir. Bu nedenle normalizasyon işleminin nispeten daha uzun süre aldığı ifade edilmektedir (Ballman ve ark., 2004).

miRNA dizileme verilerinin normalizasyonu

MiRNA dizileme tekniğinde gerekli filtrelemeler yapıldıktan sonra iyi kalitedeki okumalarla miRNA dizileri ilgili veri tabanları ile tespit edilmektedir. Okuma sonrası elde edilen veriler, miRNA'ların toplam okuma sayıdır. Bu nedenle veri normalizasyonu temelinde miRNA okuma sayıları ile gerçekleştirilmektedir (Tam ve ark., 2015). MiRNA dizileme verilerinin normalizasyonunda farklı yöntemler kullanılmaktadır. Bu amaçla yaygın olarak kullanılan yöntemler; milyon başına düşen okuma sayısı (counts per million; CPM veya read per million; RPM), toplam sayı normalizasyonu, medyan normalizasyonu, üst çeyrek normalizasyonu, M'in budanmış ortalaması (trimmed mean of M; TMM) ve kuantil normalizasyon ve DEseq2 yöntemleridir.

CPM

CPM, miRNA dizileme verilerinin normalizasyonunda kullanılan en basit yöntemdir. Bu yöntemde normalizasyon, karşılaştırılan örnekler arasında; bir miRNA'nın toplam okuma sayısının toplam okuma sayısına bölünmesi ve elde edilen değer 1×10^6 ile çarpılması ile hesaplanmaktadır. Bu yaklaşımda, toplam

okuma sayısının 10^6 'ya göre orantılanmasıyla, miRNA okuma sayıları tekrar hesaplanmaktadır (Tam ve ark., 2010). Bu sayede, okuma sayısı bir milyondan fazla veya az olsa dahi, milyon başına düşen miRNA okuma sayısı dikkate alınmaktadır. Daha sonra veriler log2 tabanındaki değere dönüştürülerek analiz gerçekleştirilmektedir (Robinson ve ark., 2010; Tam ve ark., 2015).

Toplam sayı normalizasyonu

Toplam sayı normalizasyonunda (total count; TC) her örnekte bir miRNA'nın toplam okuma sayısı örnekteki tüm okumaların sayısına bölünerek hesaplanmaktadır. Tüm örnekler için bu hesaplama yapılmaktadır. Daha sonra elde edilen değer log2 tabanındaki değere dönüştürülmektedir (Dillies ve ark., 2013). Ancak miRNA dizileme verilerinin normalizasyonunda kullanılan TC ve CPM yaklaşımlarında varyasyonun yüksek olduğu bu nedenle bu iki yaklaşımın güvenilirliğinin diğer yöntemlere kıyasla düşük olduğu bildirilmiştir (Tam ve ark., 2015).

Medyan normalizasyonu

Medyan normalizasyonunda, bir miRNA'nın okuma sayısı, tüm örneklerdeki okumaların median değerine bölünmesi ve 1×10^6 ile çarpılmasıyla veriler normalize edilmektedir. Normalizasyon sonrası elde edilen değer log2 tabanına dönüştürülerek analiz gerçekleştirilmektedir (Dillies ve ark., 2013).

Üst çeyrek normalizasyonu

Üst çeyrek normalizasyonunda; bir genin toplam okuma sayısı, her örnekteki 75. yüzdelikteki okuma sayısına bölünerek verilerin normalizasyonu gerçekleştirilmektedir. Bu normalizasyon işlemi Bioconductor programında EdgeR package ile gerçekleştirilebilmektedir (Robinson ve ark., 2010; Tam ve ark., 2015).

Kuantil normalizasyon

Kuantil normalizasyonda çoğu miRNA'nın ekspresyon seviyesinin değişmediği varsayılmaktadır. Mikroarray verilerinin kuantil normalizasyonu ile benzer olan bu yaklaşımda, normalizasyon işleminde miRNA okuma sayıları kullanılmakta ve tüm örneklerde genlerin çoğunun ekspresyon düzeyi dağılımının benzer olduğu kabul edilmektedir (Bolstad ve ark., 2003; Tam ve ark., 2015). Bu analiz Bioconductor Limma paketi ile gerçekleştirilebilmektedir (Ritchie ve ark., 2015).

TMM

MiRNA dizileme tekniğinde genellikle örneklerdeki RNA miktarı göz ardı edilmektedir. RNA miktarının örnekler arasında farklı olması, okuma sayılarının farklı olmasına neden olmaktadır. Öyle ki; analiz so-

nucunda bir miRNA'nın ekspresyon düzeyinin değişmediği tespit edilse bile, bu durumun örneklerdeki RNA miktarı farklılığından kaynaklanabileceği belirtilmektedir (Tam ve ark., 2015). TMM normalizasyon yöntemi ile RNA miktarından kaynaklanan varyasyonun önüne geçilmesi amaçlanmaktadır. TMM yönteminde normalizasyon işlemi; genlerin okuma sayıları ve toplam okuma sayıları ile hesaplanan M ve A değerlerinin belirli bir oranda budanmasıyla gerçekleştirilmektedir (Robinson ve Oshlack, 2010). Budama oranı yapılan çalışmaya göre %5-%30 arasında değişiklik gösterebilmektedir (Robinson ve ark., 2010; Tam ve ark., 2015). Bu normalizasyon Bioconductor EdgeR paketi ile gerçekleştirilebilmektedir (Robinson ve ark., 2010).

DEseq2

DEseq2 yaklaşımında; bir örnekteki bir miRNA'nın okuma sayısı, tüm örneklerdeki tüm genlerin okuma sayılarının geometrik ortalamalarına bölünerek veriler normalize edilmektedir. Bu normalizasyon işlemi Bioconductor DEseq2 paketi ile gerçekleştirilmektedir (Love ve ark., 2014).

Sonuç

MiRNA ekspresyon profillerinin belirlenmesine yönelik yapılan çalışmalardan yüksek miktarda veri elde edilmektedir. Elde edilen bu verilerin normalizasyonunda çok sayıda yöntem kullanılmaktadır. Referans gen yaklaşımı RT-qPCR verilerinin normalizasyonunda en yaygın kullanılan yaklaşımdır. MiRNA mikroarray platformları ve miRNA dizileme tekniğinde ise kullanılan normalizasyon yaklaşımlarının sayısı nispeten daha fazladır ve bu yaklaşımların güvenilirliği yapılan çalışmaya göre değişebilmektedir. Bu nedenle, özellikle miRNA mikroarray platformları ve miRNA dizileme çalışmalarında veri normalizasyonu hususunda fikir birliğinin bulunmadığı söylenebilir. Bu bağlamda, miRNA ekspresyonu profillerinin değerlendirilmesi amacıyla gerçekleştirilen çalışmalarda uygun normalizasyon yaklaşımlarının belirlenmesi sonuçların güvenilirliği açısından önemlidir. Bu sayede, miRNA'ların fizyolojik olaylar, kanser ve diğer patolojik olaylardaki rollerinin daha güvenilir bir şekilde belirlenmesi mümkün olabilecektir.

Kaynaklar

Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* 2004; 20(16): 2778-86.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; 19: 185-93.

Bustin SA, Benes V, Garson J, Hellems J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Ship-

ley GL, Vandesompele J, Wittwer CT. The MIQE guide-lines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009; 55: 611-22.

Dheda K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, Rook GA, Zumla A. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal Biochem* 2005; 344 (1): 141-3.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* 2013; 14: 671-83.

Donati S, Ciuffi S, Brandi ML. Human circulating miRNAs real-time qRT-PCR-based analysis: an overview of endogenous reference genes used for data normalization. *Int J Mol Sci* 2019; 20: 1-19.

Faraldi M, Gomasca M, Banfi G, Lombardi G. Free circulating miRNAs measurement in clinical settings: the still unsolved issue of the normalization. *Adv Clin Chem* 2018; 87: 113-39.

Gershon D. Microarray technology: An array of opportunities. *Nature* 2002; 416: 885-91.

Hammond SM. An overview of microRNAs. *Adv Drug Deliv Rev* 2015; 87: 3-14.

Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin M. Circulating micromas as novel minimally invasive biomarkers for breast cancer. *Ann Surg* 2010; 251: 499-505.

Inyawilert W, Fu TY, Lin CT, Tang PC. Let-7-mediated suppression of mucin 1 expression in the mouse uterus during embryo implantation. *J Reprod Dev* 2015; 61: 138-44.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4(2): 249-64.

Izumi H, Kosaka N, Shimizu T, Sekine K, Ochiya T, Takase M. Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions. *J Dairy Sci* 2012; 95(9): 4831-41.

Korkmaz Ağaoğlu Ö, Sidekli Ö. Kantitatif RT-PCR (RT-qPCR) Çalışmalarında uygun housekeeping genlerin (HKGs) seçimi ve validasyonu. *Erciyes Üniv Vet Fak Derg* 2020; 17(1): 76-83.

Lange T, Stracke S, Rettig R, Lendeckel U, Kuhn J, Schlüter R, Rippe V, Endlich K, Endlich N. Identifi-

- cation of miR-16 as an endogenous reference gene for the normalization of urinary exosomal miRNA expression data from CKD patients. *PLoS One* 2017; 12(8): 1-13.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 2003; 425: 415-9.
- Liu CG, Calin GA, Volinia S, Croce CM. MicroRNA expression profiling using microarrays. *Nat Protoc* 2008; 3(4): 563-78.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 2001; 25(4): 402-8.
- López-Romero P, González MA, Callejas S, Dopazo A, Irizarry RA. Processing of agilent microRNA array data. *BMC Res Notes* 2010; 3(1): 1-6.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15(550): 1-21.
- Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. *Science* 2004; 303(5654): 95-8.
- Mandrizzato S. Technological platforms for microarray gene expression profiling. Simone M. eds. In: *Microarray Technology and Cancer Gene Profiling*, New York: Springer Science+Business, 2007; pp.12-8.
- Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009; 10(6): 1-10.
- Metpally RP, Nasser S, Malenica I, Courtright A, Carlson E, Ghaffari L, Villa S, Tembe W, Keuren-Jensen V. Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. *Front Genet* 2013; 4(20): 1-13.
- Meyer SU, Pfaffl MW, Ulbrich SE. Normalization strategies for microRNA profiling experiments: A 'normal' way to a hidden layer of complexity? *Bio-technol Lett* 2010; 32(12): 1777-88.
- Montenegro D, Romero R, Kim SS, Tarca AL, Draghici S, Kusanovic JP, Kim JS, Lee DC, Erez O, Gotsch F, Hassan SS. Expression patterns of microRNAs in the chorioamniotic membranes: a role for microRNAs in human pregnancy and parturition. *J Pathol* 2019; 217(1): 113-21.
- Pradervand S, Weber J, Thomas J, Bueno M, Wirapati P, Lefort K, Dotto GP, Harshman K. Impact of normalization on miRNA microarray expression profiling. *RNA* 2009; 15(3): 493-501.
- Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: Approaches and considerations. *Nat Rev Genet* 2012; 13(5): 358-69.
- Qureshi R, Sacan A. A novel method for the normalization of microRNA RT-PCR data. *BMC Med Genomics* 2013; 6(1): 1-13.
- Rao Y, Lee Y, Jarjoura D, Ruppert AS, Liu CG, Hsu JC, Hagan JP. A comparison of normalization techniques for microRNA microarray data. *Stat Appl Genet Mol Biol* 2008; 7(1): 1-18.
- Redshaw N, Wilkes T, Whale A, Cowen S, Huggett J, Foy CA. A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability. *Biotechniques* 2013; 54(3): 155-64.
- Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43(7): 1-13.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26: 139-40.
- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010; 11(3): 1-9.
- Saliminejad K, Khorram Khorshid HR, Soleymani Fard S, Ghaffari SH. An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *J Cell Physiol* 2019; 234(5): 5451-65.
- Schwarzenbach H, Da Silva AM, Calin G, Pantel K. Data normalization strategies for microRNA quantification. *Clin Chem* 2015; 61(11): 1333-42.
- Sidekli Ö, Korkmaz Ağaoğlu Ö. Kantitatif RT-PCR (RT-qPCR) ile mikroRNA (miRNA) ekspresyon profillemesi. *Erciyes Üniv Vet Fak Derg* 2021; 18(1): 48-56.
- Sidekli Ö, Korkmaz Ağaoğlu Ö. Gebelik sürecinde rol oynayan mikroRNA (miRNA)'lar. *Lalahan Hay Araşt Enst Derg* 2019; 59(1): 36-48.
- Tam S, Tsao MS, McPherson JD. Optimization of miRNA-seq data preprocessing. *Brief Bioinform* 2015; 16(6): 950-63.
- Wu D, Hu Y, Tong S, Williams BR, Smyth GK, Gantier MP. The use of miRNA microarrays for the analysis of cancer samples with global miRNA dec-

rease. RNA 2013; 19(7): 876-88.

Zalewski K, Misiek M, Kowalik A, Bakula-Zalewska E, Kopczyński J, Zielińska A, Bidziński M, Radziszewski J, Gózdź S, Kowalewska M. Normalizers for microRNA quantification in plasma of patients with vulvar intraepithelial neoplasia lesions and vulvar carcinoma. *Tumour Biol* 2017; 39: 1-6.

Zhang L, Liu XR, Liu JZ, Song YX, Zhou ZQ, Cao BY. miR-182 selectively targets HOXA10 in goat endometrial epithelium cells in vitro. *Reprod Domest Anim* 2017; 52(6): 1081-92.

Zheng G, Wang H, Zhang X, Yang Y, Wang L, Du L, Li W, Li J, Qu A, Liu Y, Wang C. Identification and validation of reference genes for qPCR detection of serum microRNAs in colorectal adenocarcinoma patients. *PLoS ONE* 2013; 8: 1-10.

Zhu HT, Dong QZ, Wang G, Zhou HJ, Ren N, Jia HL, Ye QH, Qin LX. Identification of suitable reference genes for qRT-PCR analysis of circulating microRNAs in hepatitis B virus-infected patients. *Mol Biotechnol* 2012; 50: 49-56.

Zhu J, He F, Hu S, Yu J. On the nature of human housekeeping genes. *Trends Genet* 2008; 24(10): 481-4.