

Yeni Nesil Dizileme Teknolojisi ile Transkriptom Analizi

Transcriptome Analysis Using Next Generation Sequencing

Melda Sarıman¹, Sema Sırma Ekmekci¹, Neslihan Abacı¹, Aris Çakiris¹, Ferda Paçal¹, Zeliha Emrence¹, Duran Üstek¹, Şükrü Öztürk*

¹*Istanbul Üniversitesi, Deneysel Tıp Araştırma Enstitüsü, Genetik AD*

**Sorumlu yazar. İstanbul Üniversitesi, İstanbul Tıp Fakültesi, İç Hastalıkları AD, Tıbbi Genetik BD*

ÖZET

Yeni Nesil Dizileme platformları sayesinde, RNA materyali paralel ve eş zamanlı olarak transkriptom dizileme yöntemiyle transkripte olan tüm genler ve izoformları, genlerin 5' ve 3' uçlarının belirlenmesi, tek nükleotid değişimleri, insersiyon, delesyon, füzyon genler, ve genlerin ifadesi incelenebilmektedir. Bu makalede transkriptom dizileme uygulamaları ile dizileme sonuçlarının analizinde kullanılan biyoenformatik araçlar hakkında genel bir bakış sunulmuştur.

Anahtar Kelimeler: Transkriptom, Yeni Nesil Dizileme, RNA Dizileme, RNA Dizileme Data Analizi, Semikonduktor Dizileme Sistemi

SUMMARY

By means of the next generation sequencing platforms, RNA material can be analyzed all genes and isoforms, annotation of 5' and 3' ends of genes, single nucleotide polymorphisms, insertion, deletion, fusion genes, and gene expression by using massively parallel and simultaneously transcriptome sequencing method. This article provides a general overview to the transcriptome sequencing, applications and the bioinformatics tools to the analysis of the sequencing data.

Keywords: Transcriptome, Next Generation Sequencing, RNA Sequencing, RNA Sequencing Data Analysis, Semiconductor Sequencing System

Yeni Nesil Dizileme Teknolojisi ile Transkriptom Analizi

İnsan genom projesinden sonra ilk çalışmalar maya, fare ve Arabidopsis 'te yapılan RNA dizileme çalışmalarıdır (1,2,3). İnsan Genom Projesinin tamamlanmasından sonra gen sayısının beklenenden çok daha az bulunması ve DNA'daki protein kodlayan bölge oranının düşük olması, bilim insanlarını genomdaki kodlanmayan bölgelerin araştırılmasına ve RNA dizilemelerine yönlendirmiştir.

Son yıllarda tüm genom çaplı ENCODE projesi ile çok sayıda hasta ve kontrol gruplarının kısa sürede ve yüksek çözünürlükte analiz edilmesi sonucunda RNA'nın organizmaların gelişimde ve hücre içerisinde kilit rol oynadığı gösterilmiştir (4,5).

On yıl öncesine kadar genlerin anlatım düzeyleri ile ilişkili aday genlerin analizi real-time PCR ve mikroarray gibi yöntemler ile kısıtlı kalmaktaydı. Tüm genlerin eş zamanlı analizi hem karmaşık bir olaydır ve hem de geniş kapsamlı teknolojiler gerektirmektedir (6). Günümüzde yeni nesil dizileme (YND) teknolojileri; DNA, RNA, miRNA, *Piwi* RNA gibi hücre içerisindeki biyo moleküller hakkında düşük maliyetle, hızlı ve paralel olarak çok büyük miktarda veri elde edilmesini sağlamıştır. YND yöntemleri ve biyoenformatik araçların çok kısa sürede gelişmesi sonucu kapsamlı teknolojilerin de sağlanması ile tüm genom çaplı çalışmalarda RNA dizileme ön plana çıkmaya başlamıştır (6).

Transkriptom, bir hücrede bulunan bütün transkriptlerin çeşidinin ve miktarının eş zamanlı olarak incelenmesidir. Bir örnekte bulunan RNA miktarına bağlı olarak, genlerin seçilmiş bir alt grubunun veya tamamının ekspresyon düzeyini, tek nükleotid değişimlerini (SNP), insersiyonları, delesyonları, translokasyonları ve kodlanmayan RNA'ları da incelemek mümkündür (7,8).

Böylece transkriptom dizileme ile hücre içerisindeki dinamik işleyişin eş zamanlı paralel olarak incelenmesi mümkündür. Biyoenformatik analizler sayesinde elde edilen veriler değerlendirilerek, RNA dünyası aydınlatılabilmektedir.

DNA dizileme ile benzer prensiplere sahip olmasına rağmen, örnek hazırlama aşamasında ve biyoenformatik analizlerde farklılıklar içermektedir. Ökaryotik hücrel RNA'nın %80'i ribozomal RNA (rRNA)' dan oluşmaktadır. Ekspresyon analizinde rol alan RNA'ları belirlemek için total RNA'dan rRNA'yı uzaklaştırmak gerekmektedir. Uzaklaştırma işlemi uygun kitler kullanarak ya da enzimatik degradasyon ile yapılabilir. RNA dizilemede RNA, önce daha stabil olan komplementer DNA (cDNA)'ya çevrilir, rastgele fragmente edildikten sonra DNA yeni nesil dizileme teknolojisi kullanılarak dizilenir. Dizileme sonucu elde edilen bu fragmentler referans genom ile karşılaştırılarak biyoenformatik yöntemlerle analiz edilir (Şema 1).

Yeni Nesil Dizileme teknolojilerinde günümüzde pek çok platform bulunmaktadır. Roche 454 (artık kullanılmamaktadır) pirodizileme yöntemini, Illumina Miseq/Hiseq/GA baz senteziyle dizileme yöntemini ve Solid Sistemi (platformu) ligasyon yoluyla dizileme yöntemini kullanmaktadır. Diğer gelişmekte olan platformlar semikondöktör çip bazlı dizileme ile Ion-Torrent, katı faza dayanan genetik dizileme platformu Helicos' ve en son olarak tek molekülü eş zamanlı olarak dizileme sistemine sahip Pasific Biosciences veya Oxford Nanopore teknolojileridir (9,10,11,12,13).

YÖNTEM

Yöntemler kullanılacak dizileme platformuna göre farklılıklar göstermektedir. Yöntem bölümünde laboratuvarımızda bulunan Ion Torrent cihazında kullanılan transkriptom dizileme yönteminden bahsedilecektir.

ÖRNEKLERİN HAZIRLANMASI

Periferik kan, kemik iliği veya doku gibi biyolojik materyallerden hücre izolasyonunun gerçekleştirilmesi gerekmektedir. Elde edilen hücre popülasyonundan veya alt hücre gruplarına ayrıştırılan spesifik hücrelerden transkriptom analizi yapmak için RNA izole edilir.

Agilent Biyoanalizör 2100 cihazında RNA 6000 Pico Kit (Agilent Technologies, ABD) kullanılarak RNA konsantrasyonu ve kalite kontrolü yapılır (11).

Total RNA'dan mRNA Eldesi

Total RNA örneklerinde ribozomal RNA'ları uzaklaştırmak için mRNA+poly A zenginleştirilmesi yapılır. Elde edilen

mRNA'ların kalitesi ve miktarını belirlemek için Agilent Biyoanalizör 2100 cihazında Agilent Bioanalyzer 6000 Nano Kit (Agilent Technologies, ABD) kullanılır. RNA miktarı ve kalitesine göre düşük ya da standart protokol uygulanarak kütüphane hazırlanır.

Agilent Biyoanalizör de ölçülen herbir numunenin mRNA örnekleri Ion Total RNA dizileme kiti (Life Technologies, ABD) ile az miktarda RNA örnekleri için uygun olan Low-input protokol yada RNA miktarı yeterli ise Standart-input protokol kullanılarak kütüphane hazırlanmalıdır (11).

Kütüphane Hazırlanması ve emPZR

Elde edilen mRNA dizileme için çok uzun ya da çok kısadır. Yeni nesil dizileme yapabilmek için mRNA örneklerini uygun büyüklükte parçalara ayırmak gerekir. RNA'yı parçalara ayırmak için RNAaz III enzimi kullanılır. RNAaz III enzimi ile muamele edildikten sonra dizileme için uygun büyüklükte olmayan parçalar purifikasyon yapılarak uzaklaştırılır. Daha sonra RNA parçalarına Ion Adaptor Mix V2 bağlanır ve ters transkripsiyon ile cDNA sentezi gerçekleştirilir.

cDNA kütüphanesi oluşturulurken barkodsuz ya da barkodlu kütüphane prosedürü seçilerek devam edilir. Her aşamada olduğu gibi cDNA kütüphanesi için de kalite ve konsantrasyon kontrolü yapılır. Biyoanalizör 2100 Cihazında Agilent® High Sensitivity DNA (Agilent Technologies, ABD) kiti kullanılarak cDNA'ların kalitesi ve konsantrasyonları ölçülür. Dizileme için uygun uzunluk 100-300 baz çiftidir. Örnekler bu uzunluk aralığında ise Ion PGM™ 200 Xpress™ Şablon (Life Technologies, ABD) kiti ile cDNA kütüphanesi dilüe edilir (Şekil 1).

Hazırlanan kütüphaneden elde edilen fragmanlar, 1 yakalama Ion sphere partiküllerine 1 fragman gelecek şekilde hesaplanıp, boncuk ve fragmanların birbirine bağlanması sağlanır. Daha sonra bu partiküller yağ içeren su karışımında emülsiyonu gerçekleştirilir. Böylece her bir partikül Emülsiyon Polimeraz Zincir Reaksiyonu (emPZR) amplifikasyonunun gerçekleşeceği bir mikroreaktör içine hapsedilmiş olur. Bu mikroreaktörler kırılarak temizlenir. Bu amplifiye olan partiküller dizileme aşamasında kullanılır (11).

TRANSKRİPTOM DİZİLEME

Yeni nesil dizileme teknolojilerinden, Ion Torrent Personal Genome Machine® (PGM™) dizileme cihazında transkriptom dizileme gerçekleştirilir.

Ion PGM™ 200 semikondöktör dizileme sisteminin temeli, DNA polimerizasyonu sırasında Hidrojen iyonu salınımı sonucu

meydan gelen pH değişiminin yarı iletken sensörler tarafından dijital sisteme aktarılmasına dayanır. Tek kalıp diziyi içeren mikroçipte tek tip nükleotid akışı uygulanarak, tamamlayıcı nükleotid yeni oluşan ipliğin yapısına katıldığında Hidrojen iyonu salınımı olur. Hidrojen iyonları hipersensitif iyon sensörleri tarafından algılanır. Herhangi bir kamera, ışık olmaksızın kimyasal bilgi dijital bilgiye aktarılarak dizileme gerçekleştirilir (11) (Şekil 2).

Biyoformatik Analizler

Yeni nesil dizileme sonuçlarında elde edilen ham veriler oldukça fazladır. Başlangıçta ham verilerin içerisinde nonspesifik diziler, duplikasyon bölgeleri ve homopolimer diziler bulunmaktadır. Doğru analiz için bu istenmeyen dizileri ham veriden uzaklaştırmak gerekir. Bu işlem sırasıyla; dizilenen bölgelerden adaptörler atılır, *quality score*'ları 30 dan düşük olanlar *trimlenir* ve yalancı okumalar dışlanır.

Dizileme sonrasında elde edilen dotalar *CLC Genomics Workbench - CLC bio v6.01* (www.clcbio.com), *NextGene v2.3.2* (www.softgenetics.com) gibi yazılımlar ile analiz edilir. Elde edilen RNA transkriptom dizileri tüm insan analiz genomu (Feb.2009 (GRCh37/hg19) (www.genome.ucsc.edu) ile karşılaştırılarak elde edilen dizi sayısı, SNP, insersiyon-delesyon, frameshift (çerçeve kayması) mutasyonlar, translokasyonlar ile transkripte olan tüm genlerin ifadeleri belirlenir.

Diğer tüm genom dizilemelerden farklı olarak transkriptom analizinde RPKM (Read of per kilobase exon per million reads) değerleri analiz edilir (11).

Biyoformatik Analizlerin Basamakları

Transkriptom dizileme sadece bir run'da yüzlerce gigabayt veri elde edilmektedir. Bunların analizi sonucunda transkriptom verilerinde anlamlı bilgiler elde edilerek füzyon genler, tek nükleotid değişimleri insersiyon delesyon ve genlerin ifadesi eş zamanlı olarak değerlendirilebilir.

RNA dizileme de tüm bu verileri analiz etmek için donanımlı bir bilgisayar sistemine, data depolama kaynaklarına, basit biyoformatik analizlerin yapılabileceği ayrı yüksek kapasitede bir alana ihtiyaç vardır (6).

Biyoformatik analiz için kullanılan programlar

Galaxy (14) kullanımı kolay ve internet üzerinden herkesin erişimine açık bir platformdur. *UNIX shell* programı ve *Perl/Python* ise veri analizi için modifikasyon yapmaya elverişli olduğundan daha avantajlıdır (6).

Buna ilave olarak bazı yardımcı programlarda yararlı olacaktır

(Tablo 1). *Cufflinks* (15), (Alternatif kırılma software), *Galaxy* (14) (RNA-seq data analizi), *Gene Ontology* (16) (Genlerin işlevleri), *Blast2GO*, *SATSUMA*, *SPINES* (17) (Gen ismi belirleme), *KeGG pathway* (18), *STRING database* (19) (genlerin birbiriyle etkileşimi), *DESeq* (20), *baySeq* (21) (ekspresyon software), *fastQCtoolkit* (23), *NGSQCtoolkit* (24) (işlem öncesi ve kalite kontrol araçları), *MapView* (25), *IGV* (26) (data görüntüleme araçları) gibi programlar kullanılabilir (6).

Dosya Formatları

Dosya formatı olarak çapraz-platform (cross-platform) içeren *fasta*, *fastaq*, *sam*, *bam*, *vcf*, *gtf* yada *gff* dosyaları ile elde edilen veriler tavsiye edilmektedir (6).

Kalite Kontrol

Dizileme sonucunda elde edilen bütün sonuçlar ham (*Raw*) veri olarak isimlendirilmektedir ve bu veriler işe yaramayan verileri de içermektedir. Bu sebeple bu verileri ayıklayıp gereksiz olanları uzaklaştırmak gerekmektedir. Bu aşamadan sonra eşleştirme ve haritalama işlemi yapılabilir. Öncelikle adaptör diziler uzaklaştırılır ve duplike dizilerin miktarı belirlenir ve *fastQC* (23) ve benzeri araçlar ile *Quality Score* hesaplanır. Verilerin doğruluğu ve güvenilirliği için bir çok farklı yönden analiz yapılması gerekmektedir. Bunun için *MapView* (25) ve *IGV* (26) gibi verileri görüntüleme araçları kullanılır (6).

Varyant Çağırma

Biyoformatik işleyiş basamaklarından biride, referans genoma eşleşen kısa okuma verilerindeki tek nükleotid değişimi (SNP), insersiyon veya delesyonların bulunması için kullanılan varyant çağırma. *The Genome Analysis Toolkit (GATK)* (27) *software*'ı yaygın olarak varyant çağırma kullanılmaktadır (32). Bunun gibi farklı istatistiksel yaklaşımlara sahip (baz eşleşme kalitesi, yerdeğiştirme oranı, populasyon allel frekansı vb) pek çok varyant çağırma software bulunmaktadır. Genel bir strateji uygulanabildiği gibi, farklı programların birbiriyle kesişiminin kullanılmasında önerilmektedir (6).

Haritalama Stratejisi

Dizileme sonrasında ham verilerin işlenmesi sonucu elde edilen okumalar referans genoma eşleştirilmelidir ya da ulaşılacak bir genom yoksa eğer, de *novo* olarak veriler bir araya toplanmalıdır. İlk olarak yeni nesil dizileme teknolojileri alanında geliştirilen ve ulaşılabilir software programlarından uygun bir haritalama aracı seçilmelidir (33). Hizalanan okumalar çok kısadır (36-125 baz). Okumaların kısa olması sonucunda özellikle ortaya çıkan biyoformatik hata oranlarının çoğunu ekzon-ekzon bağlantısı (kesişmesi) kapsar (34). Burada dikkat edilmesi gereken, hizalanan kısa okumalardaki ekzon-ekzon bağlantıları ile birlikte okuma eşleştirmelerinin belirsizliğinin

(uyumsuzluğunun) nasıl üstesinden gelineceğidir (6).

Referans genoma transkriptom dizilemenin haritalanmasında iki ana algoritma yaklaşımı bulunmaktadır. Birincisi **kırpılma (Splice) olmayan okuma hizalanması** olup referans genoma, hizalanan tüm okumalar arada hiç boşluk bırakılmadan bir araya getirilir. Bu yaklaşım ekzon ve bağlantıların tanımlanmasında sınırlıdır ve yeni ekzon kırılma durumlarının tanımlanmasına izin vermez (34).

İkincisi **kırpılma hizalanması** olup, geliştirilen biyoenformatik metodlar ve *TopHat* (35), *MapSplice* (36), *GSNAP* (37) benzeri araçlar kullanılarak önce hizalanan okumalar referansa aralıksız hizalanır, sonrasında referansa eşleşmeyen okumalar kısa segmentlere ve bağımsız olarak hizalanmaya ayrılırlar.

Referansa haritalanan transkriptom verilerinin tüm transkriptlerinin ve izoformlarının dikkatlice tanımlanması gerekmektedir. Bu işleyişe transkriptom datasının yeniden yapılandırılması denilmektedir. Biyoenformatik olarak genomdan bağımsız ve genom yönlendirmeli olmak üzere iki ana yaklaşımla incelenmektedir. *Cufflinks*, *Scripture*, *Velvet*, *TransABYSS* gibi pek çok biyoenformatik program kullanılacak yaklaşıma göre bilinen bir referans genomda yeni transkriptler tanımlanmasına ya da bilinen referans genom olmaksızın yeni genlerin ve transkript izoformlarının tanımlanmasına olanak sağlayacaktır (34).

Gen İsmi Tayini

Gen ismi tayini RNA dizileme deneylerinden elde edilen verilerin biyolojik anlamının belirlenmesi gereken en önemli basamaktır. Haritalama yaklaşımında açıklamalı referans genoma haritalanarak gen isimleri belirlenir. Bununla birlikte *de novo transkriptom* dizileme olduğunda, *contig*'ler gen dizisi hakkında hiçbir bilgiyi sağlayamayacaktır ve ilişkili türlerde ortolog genlerin belirlenmesi de her zaman işe yarayan bir yol değildir. *NucMer* ve *ProMer* gibi *suffix-tree* –tabanlı metodlar yakın ilişkili türlerde gen ismi tayini için uygundur. BLAST tabanlı ortolog yöntemi ise uzak ilişkili türler için daha iyi bir alternatiftir (6).

Normalizasyon

RNA dizilemenin can alıcı unsurlarından biridir. Farklı uzunluktaki transkriptleri karşılaştırırken, uzunlukların kontrolü için normalizasyon gereklidir. Çünkü uzun transkriptler, eşit anlatım yapan kısa okumalardan pek çoğu ile birlikte örtüşebilir. Transkript uzunluğu aynı transkriptlerin farklı ekspresyonlarına göre daha az problemdir ve böylece farklı gen ekspresyonları için kullanılan istatistiksel metodları sıklıkla içine almamaktadır.

İki örnek arasındaki gen anlatım profili karşılaştırıldığında, normalizasyonun diğer önemli bir durumuda iki örnek arasındaki kalite ve dizileme gücünün farklılıkları için kontrol olmasıdır (6).

RPKM değeri genlerin ekspresyon düzeylerinin normalize edilmesi için hesaplanır. RPKM bir gene haritalanmış olan okuma sayısının diziden elde edilen toplam okuma sayısı ile bir gen için baz çifti değerinden ekzon uzunluğuna oranıdır (11) (Tablo 2).

Şüphesiz ki, derin bir kütüphaneden gelen bir gen için RPKM değeri, çok sık bir kütüphaneden gelen denk değerlerden istatistiksel olarak daha anlamlı olabilecektir (33).

Gen-Gen Etkileşimi ve Protein Özellikleri

Başarılı bir RNA dizi analizi sonucunda elde edilen veriler aday gen seti listesi oluşturacaktır. Bu listedeki genlerin birbiriyle olan ilişkileri, genin hastalıkla ilişkisini, metabolik yollardaki etkileşimlerini gösterebilecek çeşitli bilgisayar programları mevcuttur. Bu programlara *Gen Set Enrichment Analysis* (GSEA) Molecular signature database programı, Protein Etkileşimleri ise *String versiyon 9.05* programı örnek olarak verilebilir (11).

BİYOENFORMATİK ZORLUKLAR

Transkriptom dizilemede ortaya çıkan en büyük biyoenformatik sorun, elde edilen okuma dizilerinin referans genoma doğru bir şekilde hizalanıp hizalanmamasıdır. Ayrıca gen ifadesinin belirlenmesi ve ifade farklılıklarının karşılaştırılabilmesi için yeni algoritmaların geliştirilmesine ihtiyaç duyulmaktadır (40).

RNA dizileme sonrası elde edilen ham veriler tek bir dizileme sonrası gigabaytlara kadar ulaşmaktadır. Yapılan çalışmalar artıkça biyoenformatik analizlerin yapılabilmesi ve verilerin saklanması için bilgisayar teknolojilerinde hızlı depolama sistemlerine ihtiyaç duyulmaktadır (6).

Transkriptom sonucunda çok geniş kapsamlı veriler elde edilir. Bu veriler sonucunda duplikasyon, delesyon, insersiyon, translokasyonlar, frameshift mutasyonlar, SNP'ler füzyonlar, genlerle ilgili yeni ekzonlar veya alternatif splicing olabilecek aday bölgeler ve RPKM değeriyle bağlantılı olarak genlerin ekspresyon düzeyleri hakkında eş zamanlı olarak geniş bilgiler elde edilir. Transkriptom gibi YND teknolojileri son yıllarda kullanılan yeni yöntemler olup, yapılan çalışmaların sonuçlarının kantitatif PZR ile doğrulanması gerekmektedir (11).

KAYNAKLAR

- 1- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker, JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 2008; 133 (3): 523-536.
- 2- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008; 320: 1344-1349.
- 3- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 2008; (7): 621-628.
- 4- The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; 489: 57-74.
- 5- Akkaya ZY, Diñer P. Derleme:Tedavi yaklaşımlarında yeni bir dönem:Kodlamayan RNA'lar ve hastalıklar. *Marmara Medical Journal* 2013; 26 :5-10
- 6- Wolf JB. Principles of transcriptome analysis and gene expression quantification :an RNA-seq tutorial. *Mol Ecol Resour.* 2013 Jul;13(4):559-72
7. Gündođdu AK, Karahan AG. Nutrigenomik Teknolojileri. *SDÜ Mühendislik Mimarlık Fakültesi Dergisi.* 2008; 33 (4): 183-191.
8. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol.* 2010 ; Chapter 4:Unit 4.11.1-13.
- 9- Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* 2012 ;22(4): 271-4.
- 10- Ustek D, Abacı N, Sırma S, Cakiris A. Derleme :Yeni Nesil DNA Dizileme .İ.Ü. DETAE Dergisi 2011 ;(1):11-18.
- 11-Sarıman Melda (2013). Multiple Myeloma Hastalarının Kemik İliđi Myeloma Hücrelerinde Gen İfadesinin Transkriptom Dizileme Yöntemi İle Araştırılması . Yüksek Lisans .İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü .İstanbul
- 12- Mutz KO, Heilkenbrinker A, Lönne M, Walter JG, Stahl F. Transcriptome analysis using next-generation sequencing . *Curr Opin Biotechnol.* 2013 Feb;24(1):22-30.
- 13- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacif Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012 Jul 24;13:341.
- 14- Goecks J, Nekrutenko A, Taylor J, The Galaxy Team Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology,* 2010 ;11, R86.
- 15- Trapnell C, Roberts A, Goff L et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols,* 2012; 7, 562-578.
- 16-Gene Ontology Consortium The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research,* 2004; 32, 258D-261.
- 17-Vijay N, Poelstra JW, K?unstner A, Wolf JBW Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology,* 2013; 22, 620-634.
- 18- KeGGpathway (www.genome.jp/kegg/pathway.html)
- 19- STRING database (<http://string-db.org/>)
- 20- Anders S, Huber W Differential expression analysis for sequence count data. *Genome Biology,* 2010; 11, R106.
- 21- Hardcastle TJ, Kelly KA baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics,* 2010; 11, 422.
- 22- Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A Differential expression in RNA-seq: a matter of depth. *Genome Research,* 2011; 21,2213-2223.
- 23- fastQCtoolkit <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 24- Patel RK, Jain M NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE,* 2012; 7, e30619.
- 25-Bao H, Guo H, Wang J et al. MapView: visualization of short reads alignment on a desktop computer. *Bioinformatics,* 2009; 25, 1554-1555.
- 26-Thorvaldsd_ottir H, Robinson JT, Mesirov JP .Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 2013; 14, 178-192.
- 27-DePristo MA, Banks E, Poplin R et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics,* 2011; 43, 491-498.

28- Freebayes (<http://bioinformatics.bc.edu/marthlab/FreeBayes>)

29- Anders S, Reyes A, Huber W Detecting differential usage of exons from RNA-seq data. *Genome Research*, 2012; 22, 2008–2017.

30-Leng N, Dawson JA, Thomson JA et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, 2013; 29, 1035–1043.

31- Katz Y, Wang ET, Airoidi EM, Burge CB Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, 2010; 7, 1009–1015.

32-Warden CD, Adamson AW, Neuhausen SL, Wu X .Detailed comparison of two popular variant calling packages for exome and targeted exon studies. *PeerJ*. 2014 Sep 30;2:e600.

33- Manuel Ares Jr. Methods for Processing High-Throughput RNA Sequencing Data . *Cold Spring Harbor Protocols* 2014 (11): 1139-48.

34-Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat methods* .2011 Jun ;8(6):469-77.

35-De Bona, F., Ossowski, S., Schneeberger, K. & Ratsch, G. Optimal spliced alignments of short sequence reads. *Bioinformatics*, 2008; 24, i174–i180 .

36- Mikkelsen, T.S. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* , 2007; 447, 167–177.

37-Robertson, G. et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods* , 2010; 7, 909–912 .

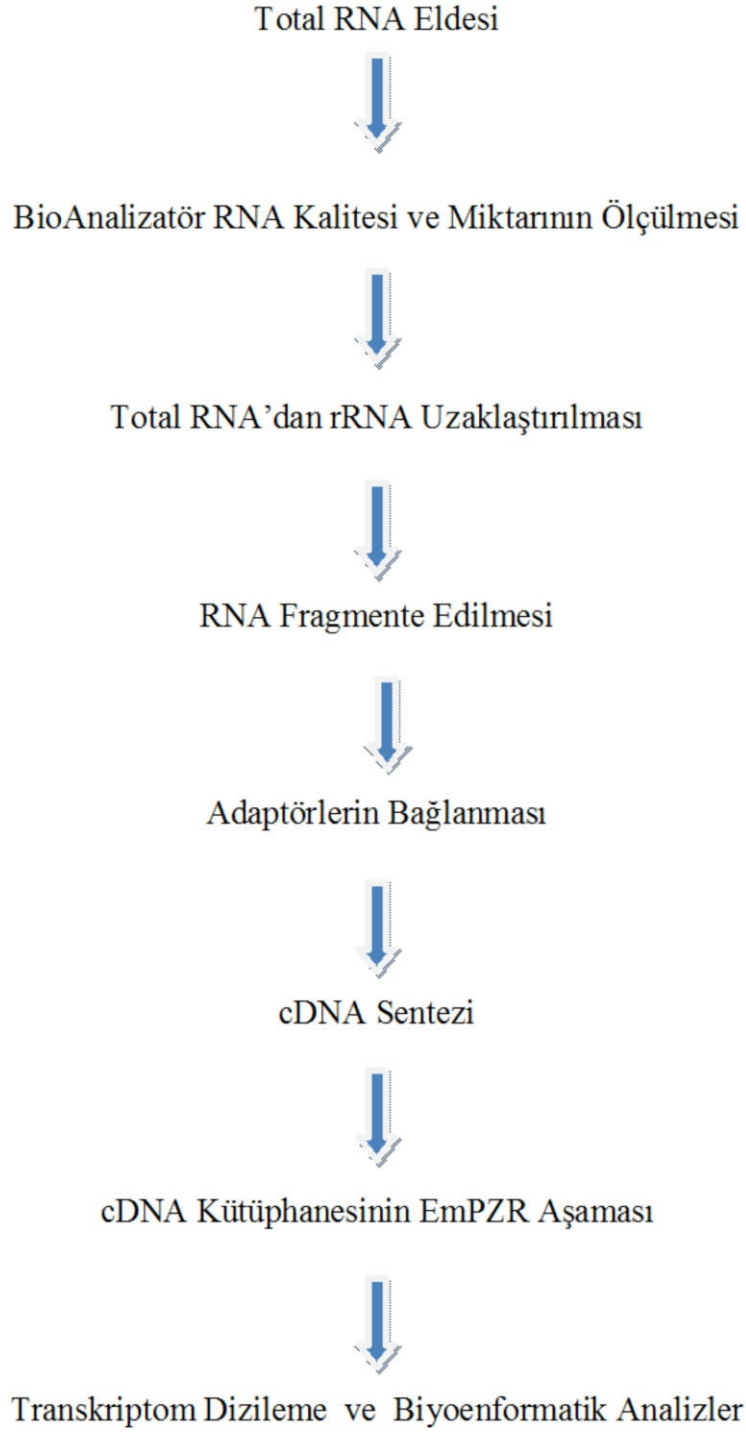
38- Ferragina, P. & Manzini, G. An experimental study of a compressed index. *Inf. Sci.* 2001; 135, 13–28.

39-Griffith, M. et al. Alternative expression analysis by RNA sequencing. *Nat. Methods* ,2010; 7, 843–847.

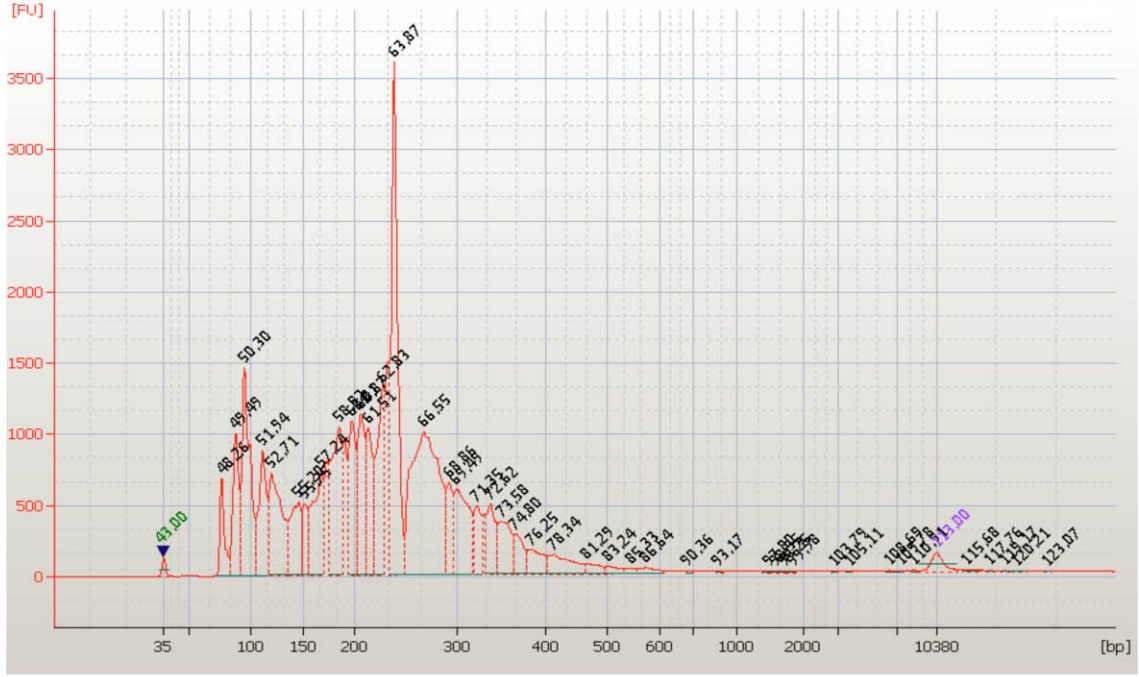
40-McGettigan PA . Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol*. 2013 Feb;17(1):4-11

TEŞEKKÜR

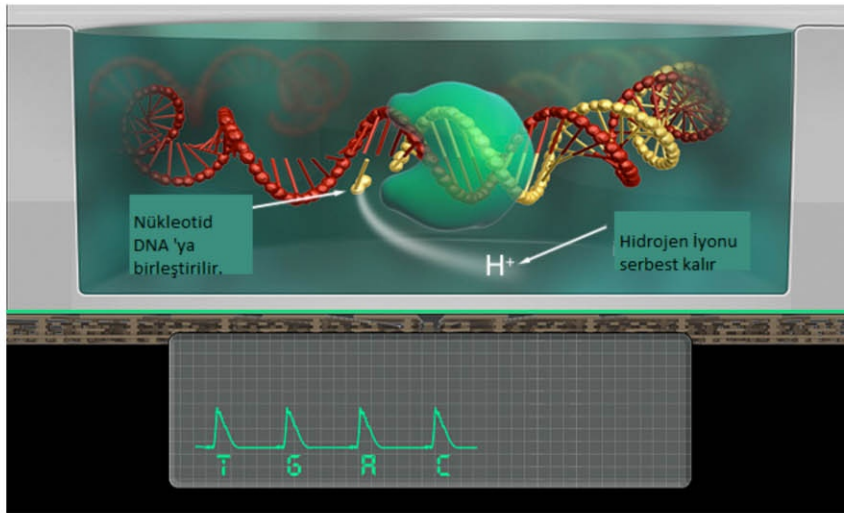
Bu makale İstanbul Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenen 7348 numaralı projeden elde edilen deneyim ve tecrübelerle yazılmıştır.



Şema1: Yöntem Akış Şeması



Şekil 1: cDNA Fragmentlerinin Büyüklüklerinin Biyoanalizator Cihazında Ölçümü.



Şekil 2: Hidrojen İyonu Salınımı ile İon Semikondüktör Dizileme (11).

Data Görüntüleme Araçları	MapView (25), IGV(26)
RNA - Dizileme Analiz için Online Platformlar	Galaxy (14), UnixSHELL (6), Perl/Phyton (6)
Kalite Kontrol ve İşlem Öncesi Araçlar	NGSQCtoolkit (24), fastQCtoolkit (23)
Varyant Çağırma ve Genotipleme	GATK (27), freebayes(28)
Alternatif Kırılma Programları	Cufflinks (15), DEXSeq (29), EBSeq (30), MISO (31)
Genlerin Özellikleri	Gene Ontology(16)
Gen İsmi Belirleme	BLAST2GO(17), SATSUMA(17), SPINES (17)
Genlerin Birbirleriyle Etkileşim Yolakları	KeGGpathway (18),STRING database(19)
Farklı Ekspresyon Programları	DEseq (20), baySeq (21), NOIseq(22)
Haritalama Programları	Kırılma Olan Okuma Hizalaması Tophat (35), MapSplice (36), GSNA (37) Kırılma Olmayan Okuma Hizalaması Bowtie (38), BWA (39)

Tablo 1: Biyoenformatik Analizlerde Kullanılan Araçlar ve Programlar (6)

<p>C = Bir gene haritalanmış okuma sayısı</p> <p>N = Dizilemeden elde edilen toplam okuma sayısı (genoma haritalanmış)</p> <p>L = Bir gen için baz çifti değerinden ekzon uzunluğu</p> <p style="text-align: center;">$RPKM = (10^9 \times C) / (N \times L)$</p>

Tablo 2: RPKM Değerinin Hesaplanma Formülü (11)