

Kanserde Somatik ve Germ-Line Tüm Genom Dizileme ve Transkriptom Profillemeye Yönelik Biyoinformatik Analiz Algoritmalarının Geliştirilmesi

Development of Bioinformatics Analysis Algorithms for Somatic and Germ-line Genome Sequencing and Transcriptome Profiling in Cancer

İbrahim BOGA^{1*}, Atıl BİŞGİN¹

¹ Çukurova Üniversitesi Tıp Fakültesi, Tıbbi Genetik Anabilim Dalı & Çukurova Üniversitesi AGENTEM (Adana Genetik Hastalıklar Tanı ve Tedavi Merkezi), Adana / TÜRKİYE

ÖZET

Amaç: Yeni Nesil Dizileme yöntemi (YND), hassas ve güvenilir bir çalışma yöntemi olması sebebiyle rutin kullanımda yerini almış olup kanserden nadir hastalıklara kadar bütün genetik temelli hastalıklarda değerini ortaya koymaktadır. Ancak hem elde edilen verinin büyüklüğü hem de maliyetleri düşürmek amacıyla yapılan çalışmalar özellikle kansere yönelik somatik çalışmalarda belirli gen bölgeleri ile sınırlıdır. Bu durum kanser gibi karmaşık hastalıkların patogenezinin aydınlatılmasını ve/veya yeni biyobelirteçlerin tespit edilmesini güçleştirmektedir. Tüm genom dizileme gibi kapsamlı bir çalışmadan dahi elde edilen veriler mevcut literatür bilgisinin yetersiz olmasından dolayı yeteri kadar iyi analiz edilememekte ve transkriptom dizileme gibi fonksiyonel testler ile destekleyici uygulamalara ihtiyaç duyulmaktadır.

Materyal ve Metot: Bu çalışmada, Çukurova Üniversitesi Adana Genetik Hastalıklar Tanı ve Tedavi Merkezi (ÇÜ AGENTEM) laboratuvar alt yapısı ile seçilen nadir kanser hastasına ait periferik kan örneğinden germ-line ve FFPE doku ile likit biyopsi materyallerinden tüm genom dizileme ve transkriptom dizileme yapılmıştır.

Bulgular: Elde edilen sekans verileri kalite kontrol aşamalarından sonra biyoinformatik analiz algoritmasının oluşturulması sağlanarak varyantlar tespit edildi. Sonuç olarak her materyal türünden yaklaşık 4'er milyon varyant tespit edildi.

Sonuç: Farklı somatik materyallerden tüm genom dizileme ve transkriptom dizileme yapılarak bu verilerin analizlerinin gerçekleştirilmesine yönelik biyoinformatik analiz algoritması geliştirilmiştir. Geliştirilen algoritma sayesinde karşılaştırmalı analizler yapılabilmeye ve yeni biyobelirteç tespit edilebilecek bir çalışma algoritması oluşturulmuştur.

Anahtar Kelimeler: Biyoinformatik, tüm genom dizileme, transkriptom dizileme, somatik varyant analizi

ABSTRACT

Aim: Next Generation Sequencing (NGS) technologies have constituted a turning point toward clinical routine genetic testing of both cancer and rare disease to deliver reliable and high sensitivity results. However, due to the challenges in management of extensive ease of analyse and cost most of NGS testing, somatic sequencing for cancers mostly limited to specific gene(s) or gene regions. Even though, this targeted approach is useful for routine testing, discovering new biomarkers to understand pathogenesis of cancer. The analysis of the entire genomic DNA sequence called as WGS provides the most comprehensive characterization; WGS still cannot be interpreted in clinical level of precision oncology without the transcriptome profiling yet due to the lack of literature and knowledge.

Materials and Methods: In this study, we performed both germ-line and somatic WGS and transcriptome sequencing within the Cukurova University AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Center) infrastructure of samples from peripheral blood, FFPE tissue and liquid biopsy in a rare cancer patient, then the comparative bioinformatics analysis. Through this study, we aimed to define an algorithm and implement the bioinformatics analysis using the germ-line and somatic data including the liquid biopsy, FFPE tissue and peripheral blood samples to establish the practices in WGS and transcriptome sequencing data analysis.

Results: As the result of sequencing quality controls, the bioinformatics algorithm was created and resulted with 4 million variations for each study conducted.

Conclusion: In conclusion, an optimal bioinformatics algorithm was developed for analysing whole genome and transcriptome sequencing of different somatic materials. Due to this algorithm we paved the way for somatic/germ-line comparative analysis for novel biomarker discovery.

Keywords: Bioinformatic, whole genome sequencing, transcriptome sequencing, somatic variant analysis.

GİRİŞ

Günümüzde kanser ilişkili çok sayıda yapılmış çalışma olmakla beraber, kanser türlerinde hastalık patogenezi ve tedavisine yönelik biyobelirteçlerin tespiti için tüm genom dizileme ve transkriptom çalışmalarının birlikte gerçekleştirildiği yeterli sayıda araştırma ve bu araştırmaların pratikte kullanımlarına yönelik bilgi birikimi bulunmamaktadır. Tüm genom ve transkriptom çalışmaları taşıdığı veri boyutu, kapsamı ve literatürde oldukça az bilgi olmasından dolayı büyük öneme sahiptir. Özellikle oldukça zor bu iki metodun birlikte yapıldığı çalışmalar ise sınırlı olup ülkemizde henüz bu yönde gerçekleştirilmiş bir çalışma bulunmamaktadır. Kanser çalışmalarında büyük önem taşıyan parafinize tümör dokusu ve periferik kandan alınan likit biyopsi örneğinden yapılan somatik varyant analizleri bakımından genom ve transkriptom dizileme düzeyinde kombine olarak yapılmış ne ulusal ne de uluslararası literatürde bir çalışma bulunmamaktadır. Bu doğrultuda, bu çalışmada kanser patogenezinin aydınlatılması ve etkili tedavi yöntemlerinin geliştirilebilmesi için biyobelirteç(ler) bulunmasına yönelik hem somatik hem de germ-line örneklerden tüm genom dizileme ve transkriptom profileme analizleri uygulanarak, biyoinformatik algoritmaların oluşturulması amaçlanmıştır.

Parafine gömülü formalin ile fikse edilmiş [Formalin Fixed Paraffin Embedded (FFPE)] dokular ile yapılan kansere yönelik yeni nesil dizileme temelli genetik çalışmalar rutin olarak uygulanan ve optimize edilmiş bir çalışma yöntemidir. FFPE doku örnekleri ile yapılan çalışmalarda, FFPE hazırlık aşamalarından kaynaklanan oluşması muhtemel DNA hasarları neticesi materyal kalitesizliğinin olabilmesi oldukça önemli bir sorundur (1). Son dönemlerde rutin testlerde çalışılan gen panelleri genişletilmiş olsa da bu testler hala tedaviye yönelik olup hastalık patogenezi için yeterli bilgi sağlanamamaktadır. Hastalık hakkında yeni bilgiler elde edilebilmesi için tüm ekzom sekanslama (TES) ve tüm genom sekanslama (TGS) gibi oldukça kapsamlı testlerin çalışılma ihtiyacı vardır. Ancak FFPE doku örneklerinden yapılan çalışmalarda yaşanan biyoinformatik analiz sorunları böylesi kapsamlı çalışmalarda çok daha fazla hissedilmekte ve bu çalışmaların tercih edilememesine sebebiyet vermektedir (2).

Periferik kandan elde edilen likit biyopsi örneklerinden izole edilen dolaşımdaki serbest DNA (ccfDNA) örnekleriyle yapılan çalışmalar son dönemlerde rutin klinik genetik testlere entegre olmuş durumdadır (3). Likit biyopsi örnekleri ile yapılan çalışmalar FFPE doku örneklerinden yapılan çalışmalardaki bazı sorunları bertaraf etse de rutin uygulamalar açısından benzer sorunlar yaşanabilmektedir. Likit biyopsi materyallerinden yapılan rutin çalışmaların sadece çoklu gen panelleri ile sınırlı şekilde gerçekleştirilmektedir ve bu çalışmalar sonucunda, hastalık patogenezinin aydınlatılması ve yeni biyobelirteçlerin tespit edilmesi için yeterli genetik veri sağlanamamaktadır (4).

Bütün bu bahsedilen metodolojilerin uygulandığı ekstramedüller miyeloid tümörler (granülositik sarkom, kloroma, miyeloblastom ve miyelosarkom olarak da adlandırılırlar) olgunlaşmamış miyeloid hücrelerinin kemik iliği dışındaki bölgelerde tümörleşmesiyle oluşmaktadır. Tek başına oluşabileceği gibi genelde başka miyeloid neoplazmalar (AML, MDS, KML vs.) hastalığa eşlik edebilmektedir (5). Hastalık oldukça nadir görülmekle (2/1.000.000) birlikte cinsiyetler arası dağılım farkı bulunmamaktadır. Ancak vakaların %60'ı 15 yaş ve altı grupta olduğundan gençlerde daha yaygın görülmektedir (6). Hastalık tanısında patoloji testleri majör rol oynamakla beraber yapılan son

çalışmalarda, yanlış tanı oranının %27 ile lenfoma %45 arasında olduğu bildirilmiştir. Bu hastalara yaygın olarak non-Hodgkin lenfoma, histiyositik, timoma, myeloma, eozinofilik sarkom, ekstra-medüller hematopoezis, mukoza ilişkili lenfoid doku lenfoması, Ewing sarkom ve karsinom tanıları konmakta ve akut lösemi şüphesiyle kemik iliği aspirasyonu, biyopsisi ya da periferik kan yayması yapılabileceği kadar tanının yanlışlığı fark edilememektedir. Hastalığın tedavisi için en yaygın kullanılan yöntem sistemik kemoterapiler olmakla beraber literatürde sınırlı sayıda da olsa hedefe yönelik tedavilerin uygulanabilirliği gösterilmiştir (7).

Bu kapsamda kanser patogenezinin aydınlatılması ve kansere yönelik yeni biyobelirteçlerin tespit edilmesi hedeflendiğinden öncelikli olarak elde edilecek verilerin işlenebilmesi adına etkin bir biyoinformatik algoritmanın geliştirilmesi hedeflenmiştir. Bu doğrultuda örneklerden tüm genom dizileme yapılmıştır. Ek olarak, somatik materyallerden yapılacak tüm genom dizi analizlerinde somatik varyant analizlerinin yapılması ve karşılaştırmalı analizlerle çalışmanın sürdürülebilmesi amacı ile periferik kan örneklerinden de elde edilen germ-line DNA ile tüm genom dizileme yapılmıştır. Tüm genom boyutundaki genetik verinin işlenmesi, biyoinformatik analizlerin doğru biçimde yapılması ve elde edilen verilerin anlamlandırılması ancak ve ancak aynı materyallerden transkriptom dizilemesi yapıp karşılaştırmalı analizlerin yapılması ile mümkün olabileceğinden bu çalışma kapsamında multi-omik yaklaşım sergilenerek transkriptom dizileme de çalışmaya dahil edilmiştir.

MATERYAL ve METOT

Çalışmada, granülositik sarkoma tanı hastadan FFPE tümör dokusu, likit biyopsi ve periferik kan örnekleri temin edilmiştir. Tüm çalışmalar Çukurova Üniversitesi Tıp Fakültesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu Tarafından onaylanmıştır (Karar no: 05042019-87).

Çalışma kapsamında FFPE dokudan DNA (Qiagen GeneReader FFPE kiti, Qiagen, Almanya), periferik kandan elde edilen likit biyopsiden ccfDNA Circulating Cell-Free DNA izolasyon kiti (Qiagen QIAamp Circulating Nucleic Acid Kit, Qiagen, Almanya) ve periferik kan lökositlerinden ise genomik DNA (QIAamp DNA Blood Mini Kit, Qiagen, Almanya) izole edildi. Ek olarak, transkriptom çalışmaları kapsamında FFPE dokudan (RNeasy FFPE Kit, Qiagen, Almanya), periferik kandan elde edilen likit biyopsiden (Qiagen QIAamp Circulating Nucleic Acid Kit, Qiagen, Almanya) ve periferik kandaki lökositlerden RNA izolasyonu (QIAamp RNA Blood Mini kit, Qiagen, Almanya) gerçekleştirildi. Genetik materyallerin (DNA ve RNA) kalite kontrolleri ve konsantrasyon ölçümleri florometrik olarak gerçekleştirildi (Qubit 4, Thermo Scientific, ABD) (1).

Tüm genom dizileme çalışmalarında NEBNext® DNA Library Prep Kit (New England Biolabs, Fransa) kiti ile üretici firmanın belirlediği protokole uygun olarak yapıldı. Transkriptom çalışmaları ise NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (New England Biolabs, ABD) kiti aracılığıyla üretici firmanın protokolü takip edilerek gerçekleştirildi. DNA ve RNA temelli yeni nesil sekanslamalar, Illumina NovaSeq (Illumina, ABD) platformunda gerçekleştirildi.

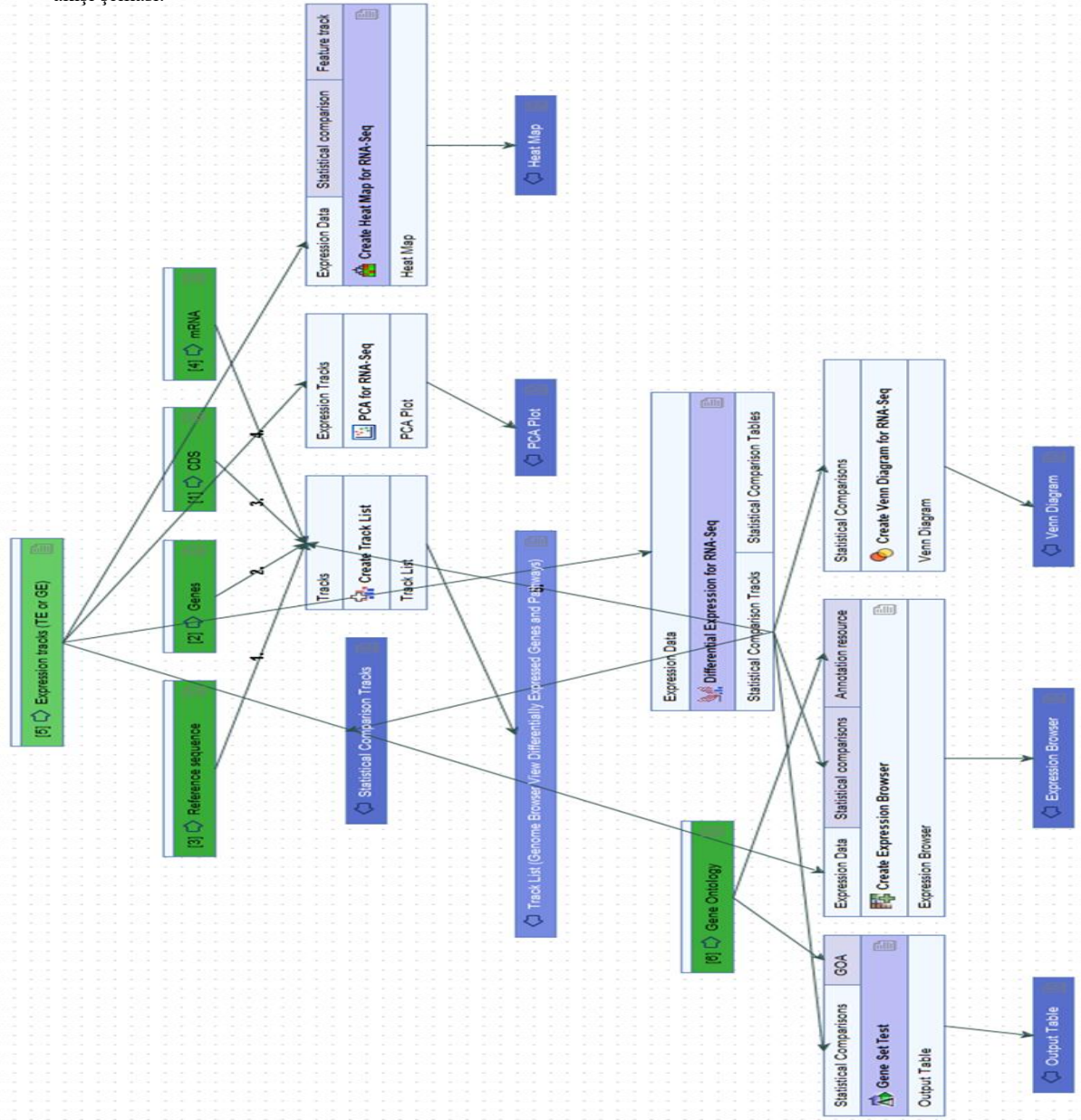
Tüm genom dizileme sonu yeterli kalite ve miktardaki verilerden hizalama (haritalama) işlemi GRCh38 – hg38 referans genomu

doğrultusunda GATK (Genome Analysis Tool Kit, The Broad Institute, Massachusetts Teknoloji Enstitüsü, ABD) biyoinformatik aracı kullanılarak yapıldı ve hizalama sonucunda BAM formatında hizalanmış veriler elde edildi. Varyant anotasyonları için ANNOVAR (Annotate Variation) biyoinformatik aracı kullanıldı. Tespit edilen varyantlar ACMG (American College of Medical Genetics and Genomics) kriterleri doğrultusunda sınıflandırıldı. Varyant anotasyonunda varyantların allel fraksiyonu ve klinik bilgileri için avSNP147, ClinVar, GWASCatalog, 1000g2015aug_eas, 1000g2015aug_sas, 1000g2015aug_eur, 1000g2015aug_afr, 1000g2015aug_amr, 1000g2015aug_all, esp6500siv2_all, ExAC_ALL, ExAC_AFR, ExAC_AMR, ExAC_EAS, ExAC_FIN, ExAC_NFE, ExAC_OTH, ExAC_SAS veri tabanlarından; varyantların fonksiyonel özelliklerini öngörmek için SIFT, Polyphen2_HVAR, Polyphen2_HDIV, MutationTaster, LRT, MutationAssessor, FATHMM, phyloP7way Vertebrate, phyloP20way_mammalian, SiPhy_29way_logOdds, gerp++gt2, CADD in silico analiz araçlarından; gen fonksiyonları ve yolak analizleri için OMIM, GWAS_Pubmed_pValue, HGMD_ID_Disease(8, 9)me, HGMD_mutation, GO_BP, GO_CC, GO_MF, KEGG_PATHWAY,

PID_PATHWAY, BIOCARTA_PATHWAY, EACTOME_PATHWAY anotasyon veri tabanlarından faydalandıldı (10-15).

Transkriptom dizileme verilerinde hizalama (haritalama) işlemi GRCh38 – hg38 referans genomu doğrultusunda CLC Genomics Workbench, versiyon 20.0.4 (Qiagen, Almanya) biyoinformatik aracı kullanılarak yapıldı ve hizalama sonucunda BAM formatında hizalanmış veriler elde edildi. Analizlerde çalışmanın hedefleri ve amaçları doğrultusunda özel olarak tasarlanarak uygulanan iş akışında ClinVar, dbSNP Common, HapMap, PhastCons conservation scores, 1000 Genomes Project, ExAC, gnomAD, ESP veri tabanları kullanılarak anotasyonu öncesi varyant analizleri, gen ve transkript ekspresyonu ön verileri elde edildi. Sonrasında ise yine tez çalışması hedefleri ve amaçları doğrultusunda özel olarak tasarlanarak uygulanan farklı bir iş akışı (Şekil-1) kullanılarak anotasyon ve karşılaştırmalı gen ve transkript ekspresyon analizleri yapıldı (16, 17).

Şekil 1. Transkriptom dizileme sonrası uygulanan 1. İş akışı sonrası karşılaştırmalı ekspresyon analizleri için oluşturulan iş akışı şeması.



BULGULAR

Tüm genom dizileme işlemi sonunda sekans kalite kontrolleri sonrası yapılan varyant analizleri neticesinde elde edilen filtrelenmemiş verilerin sadece büyüklüğü bile çalışmanın ne denli geniş kapsamlı olduğunu gösterir niteliktedir. Bu analizlerde saptanan tek nükleotid değişimi şeklindeki varyantların toplam sayıları sırasıyla periferik kan lökositleri ile yapılan dizilemede 3.909.441 (üç milyon dokuz yüz dokuz bin dört yüz kırk bir) varyant, FFPE doku ile yapılan dizilemede 3.796.122 (üç milyon yedi yüz doksan altı bin yüz yirmi iki) varyant, periferik kandan elde edilen likit biyopsi materyalinden yapılan dizilemede ise 3.919.415 (üç milyon dokuz yüz on dokuz bin dört yüz on beş) varyanttır. Her üç örnekte de tespit edilen bu varyantların büyük bir kısmı intergenik bölgelerde gözlenmektedir. En az varyant tespit edilen bölge ise her üç örnek için de ncRNA "splice" bölgeleridir. Varyantların tespit edildiği bölgeler Tablo 1’te detaylı olarak sunulmuştur.

CDS: Kodlayan bölgelerdeki tek nükleotid değişimleri, Sinonim: amino asit değişimine sebep olmayan tek nükleotid değişimleri, Missense: amino asit değişimine sebep olan tek nükleotid değişimleri, Stop kazanımı: stop kodon oluşumuna sebep olan tek nükleotid değişimleri, Stop kaybı: stop kodon kaybına yol açan tek nükleotid değişimleri, Bilinmeyen: fonksiyonel değişimi veri tabanlarında gen ile ilgili yeterli veri olmamasından dolayı bilinmeyen tek nükleotid değişimleri, İntronik: intronik bölgelerdeki tek nükleotid değişimleri, UTR3: 3’ translyasyon edilmeyen bölgelerdeki tek nükleotid değişimleri, UTR5: 5’ translyasyon edilmeyen bölgelerdeki tek nükleotid değişimleri, Splicing: Ekzon-intron kavşaklarındaki tek nükleotid değişimleri, ncRNA_ekzonik: ncRNA’ların kodlayan bölgelerindeki tek nükleotid değişimleri, ncRNA_intronik: ncRNA’ların intronik bölgelerindeki tek nükleotid değişimleri, ncRNA_splicing: ncRNA’ların ekzon-intron kavşaklarındaki tek nükleotid değişimleri, Upstream: transkripsiyon başlangıcından 1 kb uzaklıktaki tek nükleotid değişimleri, Downstream: transkripsiyon bitişinden 1 kb uzaklıktaki tek nükleotid değişimleri, İntergenik: intergenik bölgelerdeki tek nükleotid değişimleri.

Tablo 1. Tek nükleotid değişimi varyantlarının, çalışılan örnek temelinde mutasyon tipi ve tespit edildiği bölgelere göre dağılımı

Örnek	Germ-line	FFPE doku	Likit biyopsi
CDS	24.017	13.814	24.217
Sinonim	12.213	6.596	12.297
Missense	11.518	6.991	11.619
Stop kazanımı	78	69	79
Stop kaybı	16	13	16
Bilinmeyen	203	152	217
İntronik	1.309.234	1.258.431	1.308.538
UTR3	25.858	20.662	25.912
UTR5	5.741	2.509	5.739
Splicing	71	51	72
ncRNA_ekzonik	12.288	9.055	12.256
ncRNA_intronik	201.365	195.229	200.962
ncRNA_splicing	57	40	59
Upstream	23.600	17.329	23.580
Downstream	24.454	21.346	24.530
İntergenik	2.281.644	2.229.921	2.292.436
Toplam	3.909.441	3.769.122	3.919.415

RNA üzerinden gerçekleştirilen transkriptom dizilemede tespit edilen toplam varyant sayıları incelendiğinde FFPE dokusunun, 5.279.773 (beşmilyon iki yüz yetmiş dokuz bin yedi yüz yetmiş üç) varyant ile en

çok varyanta sahip olan biyolojik materyal olduğu görülmüştür. Periferik kan lökositleri ile yapılan dizilemede ise 1.652.628 (bir milyon altı yüz elli iki bin altı yüz yirmi sekiz) varyant tespit edilirken en az varyant sayısının periferik kandan elde edilen likit biyopsi materyali ile yapılan dizileme sonucunda [92.075 (doksan iki bin yetmiş beş)] olduğu belirlenmiştir.

Transkriptom çalışmaları neticesinde çalışma kalite kontrolleri yapılmış, sonrasında elde edilen veriler, hem varyant tespitine yönelik hem de genlerin ekspresyonları bakımından değerlendirilmiştir. Herhangi bir kritere göre filtreleme yapılmadan gerçekleştirilen analizlerde, her üç örnek tipinde de en çok görülen varyant tipi olan tek nükleotid değişimlerinin toplam sayıları; periferik kan lökositleri ile yapılan dizilemede 3.909.441 (üç milyon dokuz yüz dokuz bin dört yüz kırk bir) varyant, FFPE doku ile yapılan dizilemede 4.373.601 (dört milyon üç yüz yetmiş üç bin altı yüz bir) varyant, periferik kandan elde edilen likit biyopsi materyalinden yapılan dizilemede ise 59.989 (elli dokuz bin dokuz yüz seksen dokuz) varyant olarak tespit edilmiştir. En az tespit edilen varyant tipi ise her üç örnek tipinde de delesyon ve insersiyonun birlikte görüldüğü indel tipi varyantların olduğu görülmüştür. Bunların sayıları ise; periferik kan lökositleri ile yapılan dizilemede 6.283 (altı bin iki yüz seksen üç) varyant, FFPE doku ile yapılan dizilemede 20.352 (yirmi bin üç yüz elli iki) varyant, periferik kandan elde edilen likit biyopsi materyalinden yapılan dizilemede ise 2.061 (iki bin dokuz altmış bir) varyanttır. Tespit edilen varyantların tipine göre dağılımı detaylı olarak Tablo 2’de sunulmuştur.

Tablo 2. Tespit edilen varyantların biyolojik örnek türüne göre ve varyant tipine göre dağılımı

Örnek	Germ-line	FFPE doku	Likit biyopsi
SNV	1.216.556	4.373.601	59.989
MNV	99.887	110.336	8.150
İnsersiyon	167.241	241.865	10.974
Delesyon	162.661	533.619	10.901
İndel	6.283	20.352	2.061
Toplam	1.652.628	5.279.773	92.075

SNV: Tek nükleotid değişimleri (Single Nucleotide Variation), MNV: Çok nükleotidli değişimler (Multi-Nucleotide Variation), İndel: İnsersiyon ve delesyon olan değişimler.

TARTIŞMA

Tüm genom dizileme sonucunda öncelikle tek nükleotid varyantları analiz edilmiştir. Her ne kadar her üç örnek tipinde de en çok varyant intronik bölgelerde tespit edilmiş olsa da FFPE doku ile yapılan dizileme çalışmalarında protein kodlayan bölgelerdeki varyantlar diğer iki örnek tipine oranla daha az sayıdadır. FFPE doku materyalinin hazırlık aşamalarında maruz kaldığı kimyasalların kodlayan bölgelerde hasara yol açabileceği ve bu nedenle bu kısımlarda oluşabilecek varyant kayıplarına sebebiyet verebileceği düşünülmüştür. En az tespit edilen varyant tipinin ise her üç örnekte de stop kodon oluşturan varyant olması, genel karakteristik anlamında beklenen bir durumdur. Bununla beraber elde edilen tek nükleotid varyantlarının tip ve bölgelere göre olan varyant sayıları bazı farklılıklar gösterse de varyantlar genel dağılımları bakımından benzerlik göstermektedir. Ayrıca, tümör heterojenitesi ve somatik varyantların farklı allel frekansı gösterebilmesi nedeniyle somatik varyantların biyoinformatik analizleri oldukça zorlayıcı olmakta ve germ-line genetik çalışmalara göre farklı bir yaklaşım gerektirmektedir. Bunlara ilaveten çok daha fazla bilgi birikimi ve tecrübeye sahip alanında uzman ve yetmişmiş kişilere olan ihtiyaç ile

bunların yüksek iş gücüne gereklilik duyulması bu alanın bir diğer zorlayıcı özelliğidir.

Günümüzde rutin uygulamalarda hem maliyet ve iş gücü kazanımı hem de test zamanlarını kısaltabilmek için küçük NGS panelleri kullanılmaktadır. Bu durum rutin uygulamalarda hastalara düşük maliyetli ve hızlı genetik test imkânı sağlasa da bu testlerden elde edilen genetik veri kanser patogenezinin aydınlatılması ya da kansere yönelik yeni biyobelirteçlerin tayin edilmesi için oldukça kısıtlı ve yetersiz kalmaktadır. Kansere yönelik rutin uygulamalarda yapılan yeni nesil dizileme çalışmalarının hedefe yönelik olarak sınırlandırılması bu sorunların daha kolay yönetilebilmesini sağladığından bu çalışmalar gerek maliyet gerek rutin uygulamalardaki etkinliğin artırılması adına belirli gen gruplarından oluşan hedefe yönelik yeni nesil dizileme temelli çoklu gen panelleri temelinde yürütülmektedir.

Likit biyopsi materyallerinden çalışmanın her ne kadar tümör heterojenitesi gibi problemleri olmasa da germ-line DNA kontaminasyonu ve potansiyel varyantların allel frekanslarının değişken olması gibi başka sorunlar, tıpkı FFPE doku örnekleri ile yapılan çalışmalarda da olduğu gibi çalışmaların gerçekleştirildiği merkezlerin ve çalışanlarının tecrübe, bilgi birikimi ve yoğun iş gücü kullanılması ile giderilebilmektedir. Dolayısıyla, dizilemedeki yüksek derinlik ihtiyacı, biyoinformatik analizlerin oldukça karmaşık ve zorlu olması ile yüksek olan maliyetin düşürülmeye çalışılması gibi sebeplerden dolayı likit biyopsi örnekleri ile yapılan rutin çalışmalar, sadece spesifik gen grupları ile yapılan yeni nesil dizilemeye yönelik çoklu gen panelleri ile sınırlı şekilde gerçekleştirilmektedir.

Rutin moleküler genetik test uygulamalarında yeni nesil dizileme tabanlı ekspresyona yönelik çalışmalar oldukça sınırlı sayıda olup genel olarak tüm dünyada olduğu gibi ülkemizde de gerçek zamanlı polimeraz zincir reaksiyonu (Real Time PCR) temelli testler kullanılmaktadır. Yeni nesil dizileme tabanlı olarak ise oldukça az sayıda merkez, hedefe yönelik çoklu gen panelleri kullanarak bu uygulamayı yapmaktadır. Ancak, sadece spesifik genler üzerinden yapılan çalışmaların bile biyoinformatik analizleri oldukça zorlayıcı olmaktadır. Rutin çalışmalarda kullanılan spesifik gen panelleri ile yapılan transkriptte yönelik çalışmalar kanser patogenezinin aydınlatılması ya da yeni biyobelirteçlerin tespit edilmesi için yetersiz kalmaktadır. Böylesine kapsamlı analizlerin RNA bazında gerçekleştirilmesi ancak tüm transkriptomun sekanslanması ile mümkün olabilmektedir.

Transkriptom dizileme verileri incelendiğinde periferik kandan elde edilen likit biyopsi materyaline ait verilerin diğer örneklerden elde edilen verilere göre daha düşük miktarda olduğu göze çarpmaktadır. Bunun sebebi olarak RNA'ların DNA'lara göre daha kırılabilir bir yapıya sahip olması ve RNA degradasyonunun çok daha hızlı gerçekleşmesi nedeniyle periferik kandan likit biyopsi izolasyonuna kadar olan zamanda materyalde RNA kaybının olması düşünülebilir. Ancak, elde edilen veri miktarı analizlerin yapılmasına engel teşkil etmeyecek yeterliliktedir.

Her üç tip biyolojik materyale çalışma öncesi rRNA depleyonu yapılmış olsa da dizileme sonunda elde edilen verilerin RNA tipine göre dağılımlarında FFPE doku örneğinde rRNA miktarının beklenenin üzerinde olması dikkat çekmektedir. Bununla beraber diğer iki örnek tipinde mRNA oranı diğer RNA tiplerine oranla belirgin şekilde yüksektir. Dikkat çeken bir nokta ise FFPE doku örneğinden yapılan çalışmaların sonucunda küçük RNA'ların oranının yaklaşık %20 gibi yüksek bir yüzdeye sahip olmasıdır. Diğer bir dikkate değer veri de likit biyopsi örneğinde ise ncRNA oranının %20'ye yakın yüksek bir değerdedir.

SONUÇ

Bu tez çalışmasıyla potansiyel biyobelirteçlerin belirlenmesine yönelik oluşturulan algoritmanın, multi-omik yaklaşımlarla yapılacak ileri düzey çalışmalar için hem önemli bir yol gösterici olması hem de bu alanda çalışacak araştırmacılara katkı sağlaması düşünülmüştür. Bununla beraber, elde edilen bulguların kanserin biyolojik temellerinin aydınlatılmasına yönelik çalışmalar açısından da geniş bir perspektif kazandırması ve yeni çalışmalara ilham vererek bilimsel literatüre katkı sağlaması beklenmektedir.

Çıkar Çatışması

Yazarlar arasında herhangi bir çıkar çatışması yoktur.

KAYNAKLAR

1. Boga I, Sonmezler O, Bisgin A. Clinical Validation of a Novel GeneReader Next Generation Sequencing System for Tumor Specific Mutations and Bioinformatics Variant Analysis. *Clin Lab*. 2020;66(11).
2. Bewicke-Copley F, Arjun Kumar E, Palladino G, Korfi K, Wang J. Applications and analysis of targeted genomic sequencing in cancer studies. *Computational and structural biotechnology journal*. 2019;17:1348-59.
3. Sonmezler O, Boga I, Bisgin A. Integration of Liquid Biopsies into Clinical Laboratory Applications via NGS in Cancer Diagnostics. *Clinical Laboratory*. 2020;66(5):763-9.
4. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer science*. 2018;109(3):513-22.
5. Paydas S, Zorludemir S, Ergin M. Granulocytic sarcoma: 32 cases and review of the literature. *Leuk Lymphoma*. 2006;47(12):2527-41.
6. Guermazi A, Feger C, Rousselot P, Merad M, Benchaib N, Bourrier P, et al. Granulocytic sarcoma (chloroma): imaging findings in adults and children. *AJR American journal of roentgenology*. 2002;178(2):319-25.
7. Almond LM, Charalampakis M, Ford SJ, Gourevitch D, Desai A. Myeloid Sarcoma: Presentation, Diagnosis, and Treatment. *Clinical lymphoma, myeloma & leukemia*. 2017;17(5):263-7.
8. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*. 2016;44(D1):D457-D62.
9. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers*. 2021;1(1):59.
10. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(D1):D840-d5.
11. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nature protocols*. 2016;11(1):1-9.
12. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*. 2013;Chapter 7:Unit7.20.
13. Schwarz JM, Rödelberger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods*. 2010;7(8):575-6.
14. Krawczak M, Ball EV, Stenson P, Cooper DN. HGMD: The Human Gene Mutation Database. In: Letovsky S, editor. *Bioinformatics: Databases and Systems*. Boston, MA: Springer US; 1999. p. 99-104.
15. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(Database issue):D514-7.
16. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*. 2018;46(D1):D1062-D7.
17. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-11.