



## **Tanısı Zor Tek Gen Hastalıklarında Hedefe Yönelik Yeni Nesil Dizileme Panel Tasarımı: Primer İmmün Yetersizlik Örneği**

### **Targeted Next Generation Sequencing Panel Design in Complex Monogenic Diseases: The Example of Primary Immunodeficiency**

Sinem Fırtına<sup>1,2</sup> , Özden Hatırnaz Ng<sup>3</sup> , Müge Sayitoğlu<sup>2</sup> , Yuk Yin Ng<sup>4</sup> 

<sup>1</sup> İstinye Üniversitesi, Fen-Edebiyat Fakültesi, Moleküler Biyoloji ve Genetik Bölümü, İstanbul, Türkiye

<sup>2</sup> İstanbul Üniversitesi Aziz Sancar Deneysel Tıp Araştırma Enstitüsü, Genetik Anabilim Dalı, İstanbul, Türkiye

<sup>3</sup> Acıbadem Üniversitesi, Tıp Fakültesi, Tıbbi Biyoloji Bölümü, İstanbul, Türkiye

<sup>4</sup> İstanbul Bilgi Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Genetik ve Biyomühendislik Bölümü, İstanbul, Türkiye

ORCID: S.F. 0000-0002-3370-8545;  
Ö.H.N. 0000-0001-7728-6527;  
M.S. 0000-0002-8648-213X;  
Y.Y.N. 0000-0001-9755-6045

#### **Corresponding author/Sorumlu yazar:**

Yuk Yin Ng,  
İstanbul Bilgi Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Genetik ve Biyomühendislik Bölümü, İstanbul, Türkiye  
E-mail: yuk.ng@bilgi.edu.tr

**Submitted/Başvuru:** 07.09.2020

**Accepted/Kabul:** 25.09.2020

**Citation/Atıf:** Firtina S, Hatırnaz NG O, Sayitoğlu M, Ng YY. Targeted Next Generation Sequencing Panel Design in Complex Monogenic Diseases: The Example of Primary Immunodeficiency. Sağlık Bilimlerinde İleri Araştırmalar Dergisi 2020; 3(3): 93-101.  
<https://doi.org/10.26650/JARHS2020-790469>

#### **ÖZ**

**Amaç:** Yeni nesil dizileme teknolojileri bugün çok sayıda aday genin, genomun tüm kodlayan bölgelerinin hatta tüm genomun analizini tek seferde ve kısa süre içerisinde düşük maliyet ve yüksek hassasiyette, güvenilir bir şekilde mümkün kılmaktadır. Hedefe yönelik yeni nesil dizileme sistemleri genomda sadece belirli bölgenin dizilenmesine imkan veren, tüm genom dizilemelere göre uygulaması ve analizi daha kolay, hızlı ve yüksek güvenilirlikte bir yöntem olarak pek çok rutin genetik tanı uygulamalarında yerini bulmuştur.

**Gereç ve Yöntem:** Çalışmamızda primer immün yetersizliklerin en yaygın grubu olan primer antikor yetersizlikleri (PAY) ve en ağır seyirli grubu ağır kombine immün yetersizlikler (AKİY) için hastalık ile ilişkili olduğu bilinen genleri kapsayan PZR temelli genetik tanı panelleri geliştirilmiş ve ortaya çıkan yüksek verinin yorumlanması için bir analiz akışı oluşturulmuştur.

**Bulgular:** Tasarlanan paneller ile toplam 112 hasta (PAY:64, AKİY:48) dizilenmiş ve AKİY hastalarının %58'i ve PAY hastalarının %14,2'sinde hastalık ile ilişkili varyantlar tespit edilmiştir. Tüm varyantlar Sanger dizileme ile doğrulanarak oluşturulan moleküler tanı panellerinin ve analiz algoritmasının doğruluğu kontrol edilmiştir.

**Sonuç:** Hedefe yönelik yeni nesil dizileme panellerinin hedeflenen bölgeye uygun olarak doğru yöntemle tasarlanması ve çıkan ham datanın doğru iş akışı ile analiz edilmesi panelin başarısını arttırmaktadır.

**Anahtar Kelimeler:** Yeni nesil dizileme, varyant yorumlama, hedefe yönelik panel tasarımı

#### **ABSTRACT**

**Objective:** Next-generation sequencing technologies can generate an analysis of a large number of candidate genes, all the coding regions of a genome, or whole genomes with a high degree of accuracy and within a short amount of time. Targeted next generation sequencing systems have been found in many routine genetic diagnosis applications that allow the sequencing of only the candidate regions of the genome.

**Materials and Methods:** In this study, we designed PCR-based targeted next generation sequencing (NGS) panels for severe combined immunodeficiency (SCID) and primary antibody deficiency (PAD) and created an algorithm for analysing high-throughput data.

**Results:** We screened 112 patients (48 SCID and 64 PAD) and we detected genetic variations in 58% of the SCID and in 14.2% of the PAD patients. All variants were validated by Sanger sequencing to validate the accuracy of the NGS panel and analysis algorithm.

**Conclusion:** Designing targeted next generation sequencing panels with an appropriate method, in accordance with the targeted region, and analysing the raw data with a suitable workflow, increases the success of the panel.

**Keywords:** Next generation sequencing, variant annotation, targeted-ngs panel design



## GİRİŞ

İnsan genom projesinin tamamlanmasından bu yana, genom dizileme teknolojileri hızla gelişmiş, maliyet azalmış ve daha uzun genomik bölgelerin dizilenmesini sağlayan yeni nesil dizileme (YND) teknolojileri ile olağanüstü ilerlemeler kaydedilmiştir (1,2). Yeni nesil dizileme teknolojileri sayesinde tüm genom, ekzom ve hedefe yönelik dizileme, epigenomik, RNA dizileme hatta tek hücre dizileme gibi birçok farklı alanda uygulama yapılabilmektedir (3). Son yıllarda yapılan yeni genom keşfi (de novo genome assembly), prenatal tanı, kanser ve virüs çalışmalarında, tek bir hücrenin genomunu amplifikasyon basamağı olmadan uzun okumalar yaparak dizileyebilen Tek Hücre Dizileme ya da diğer adıyla 3. Nesil Dizileme sistemleri (Pacbio ve Oxford Nanopore) oldukça sık kullanılmaktadır (4).

Yeni nesil dizileme sistemleri yöntemsel değişiklikler gösterse de temel olarak kütüphane hazırlanması, kütüphanenin işaretlenmesi ve tekrar çoğaltılması ("enrichment"), tüm kütüphanenin aynı anda dizilenmesi ve biyoinformatik araçlar ile bir referansa göre veya *de novo* olarak dizilenen parçaların tekrar birleştirilmesi basamaklarını takip eder (5). Yeni nesil dizileme yöntemleri sayesinde temel araştırma amaçlı oldukça büyük ve hızlı genomik bilgi elde edilmekle beraber, günümüzde bu sistemlerin klinik tanıda kullanımı da yaygınlaşmıştır. Bugün moleküller genetik uzmanlarının yanı sıra mikrobiyoloji, onkoloji, embriyoloji gibi birçok tıp dalında YND'nin farklı yaklaşımları kullanılmaktadır.

Hedefe yönelik yeni nesil dizileme ya da diğer bir adıyla amplikon dizileme (AD) genomda farklı lokuslardaki genlerin veya çok uzun bir genomik lokusa yayılmış bir genin, prob ya da primerler ile genomdan ayrılarak, sadece belirlenen hedef bölgenin birden fazla örnekte aynı anda dizilenmesi esasına dayanır. AD sayesinde aynı anda çok sayıda hasta örneği, yüksek hassasiyette ve tek seferde dizilenebilmektedir. Konvansiyonel Sanger dizilemenin aksine AD dizileme yönteminde her bir örneğin birden fazla kez dizilenmesi ("deep"=okuma derinliği) %1'den daha nadir gözlenen bir genetik değişimi bile yakalayabilmesine olanak sağladığından, bu yöntem tümör dokusu gibi

heterojen yapıya sahip dokulardaki farklı hücre klonlarının tespiti, hastalıkla ilişkili nadir alel bulma, HIV gibi virüslerin nadir varyantlarının tespiti veya evrim biyolojisi çalışmalarında sıklıkla kullanılmaktadır (6).

Genetik ve klinik heterojenitenin çok olduğu tek genli kalıtım gösteren hastalıklarda, hızlı genetik tanı için hedefe yönelik YND paneller oldukça sık kullanılmaktadır. Paneller hastalığın doğasına göre tek gen, birkaç gen, onlarca ya da yüzlerce genden oluşabilmektedir. Panel tasarımları panellerde taranan hastaların tanı başarısını doğrudan etkilemektedir. Rutin kullanıma uygun, maliyeti düşük ve analizi daha kolay panel tasarımları ticari olarak bir yarış haline gelmektedir. Panel tasarımlarında hastalık spesifik yaklaşımların kullanılması panel başarılarını artırmakta ve özellikle nadir çocukluk çağı hastalarının hızlı tanı alabilmesine ve tedavilerinin (kemik iliği nakli, replasman vb.) hızla planlanmasına yol açmaktadır (7).

Primer İmmün Yetersizlikler (PİY), immün sistemin gelişim ve fonksiyonel bozuklukları ile karakterize, çoğunlukla tek genli kalıtım gösteren fakat sorumlu 400'den fazla gen tanımlanmış hastalıklar grubudur (8). Primer immün yetersizlikler genetik ve klinik heterojenite gösteren bir hastalık grubu olduğu için hastaların çoğu erken dönemde teşhis edilememektedir. Moleküler tanı, hastalığın kesin tanısı ve tedavi seçeneklerinin belirlenmesi için gerekli en önemli faktörlerden biri olsa da hastalıkla ilişkili yüzlerce genin bildirilmiş olması bu süreci zorlaştırmaktadır. Genetik tanı ayrıca genetik danışma, prenatal tanı ve hastalığın ne kadar agresif seyredeceğinin önceden bilinmesine olanak verdiğinden büyük önem taşımaktadır (9,10).

Bu çalışmada primer immün yetersizlik hastalık modeli örneği ile karmaşık tek gen hastalıklarında hedefe yönelik yeni nesil dizileme yöntemi kullanılarak oluşturulan panel tasarımı, dikkat edilmesi gereken konular, analiz yaklaşımları ve tanı başarıları bildirilmektedir.

## GEREÇ VE YÖNTEM

### Panel Genlerinin Seçimi

PİY'lerin iki önemli alt grubu olan ağır kombine immün yetmezlik (AKİY) ve primer antikor yet-

mezliği (PAY) için hastalığa özgü genetik tanı paneli tasarlanmıştır. Çalışmada geniş bir literatür ve veri tabanı taraması sonucu sıklıkla mutasyona uğrayan ve klinik ilişkileri kanıtlanmış genler üzerine yoğunlaşmıştır. Paneller için Avrupa Klinik İmmünoloji Derneğinin PİYlerin genetik altyapısında tanımladığı genler arasından AKİY hastalarına özgü hastalıkta en sık görülen 18 gen, PAY hastalarına özgü hastalarda en sık tanımlanan 22 adet gen seçilmiştir (11). Hedef genlerin tüm ekzonları, 5'UTR (UnTranslated Region- kodlamayan bölgeler), 3'UTR ve promotör bölgelerini kapsayacak şekilde primerler tasarlanmıştır. Primer dizaynı için hg38/GRCh38 referans genom dizisi kullanılarak firma tarafından çip tasarımına uygun olarak dizayn edilmiştir. Primer dizileri genomdaki lokasyonu ve özgünlüğü, CLC Genomics (CLC Genomics Workbench, QIAGEN) ve çevrimiçi bir program olan BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) ile kontrol edilmiştir.

Hedef bölgelerin baz uzunluğu AKİY paneli için 167.092 baz çifti (bç), PAY paneli için ise 234.975 baz çifti olmuştur. AKİY paneli için belirlenen 18 aday gen, ortalama 386 baz çifti (en düşük 250 en yüksek 450 baz çifti) uzunluğundaki toplam 432 PZR ile PAY paneli için belirlenen 22 aday gen ise ortalama 423 baz çifti (en düşük 253 en yüksek 646 baz çifti) olmak

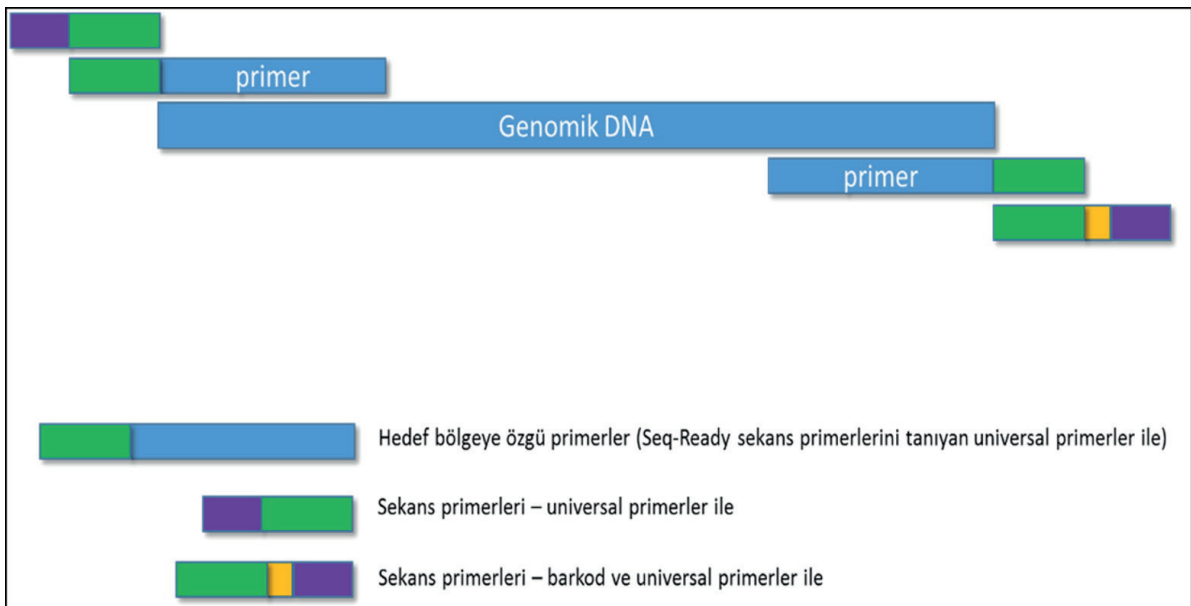
üzere toplam 555 PZR ile çoğaltılmıştır. Buna göre tasarlanan bir AKİY çipi ile aynı anda on iki, bir PAY çipi ile de sekiz hasta dizilenebilmiştir.

### Amplikon Dizileme Yaklaşımının Seçilmesi

Tekrar eden genomik bölgeler (dizilemede büyük sorun yaratmaktadır) ve psödogen sorunlarını bertaraf etmek amacı ile PZR (polimeraz zincir reaksiyonu) temelli amplikon dizileme metodu tercih edilmiştir. Amplikon dizileme metodu için hedef genleri yaklaşık 250-600 baz uzunluğunda tek tek PZR'ler ile ("singleplex") çoğaltan Seq-Ready (Seq-Ready™ TE MultiSample Custom DNA Panels, Wafergen, CA, ABD) çipleri kullanılmıştır. Seq-Ready çiplerinin çalışma prensibi hedef bölgeleri her bir kuyu içerisinde sekans primerleri ve hastaya özgü atanmış barkodlar ile birleştirilmiş primerler ile çoğaltarak amplikon kütüphaneleri oluşturmaktır (Şekil 1). Kullandığımız çiplerin 5184 ayrı nano kuyudan oluşması aynı anda 5184 farklı PZR'nin yapılmasına imkan vermekte ve bu da birden fazla örneğe birden fazla PZR'nin yapılmasını sağlayan esnek bir tasarım oluşturmaktadır.

### Yeni Nesil Amplikon Dizileme İş Akışı

Tasarlanan panellerinin başarılarını belirlemek için AKİY ve PAY şüphesi olan hastalarda YND



Şekil 1. PZR temelli (amplikon) hedefe yönelik yeni nesil dizileme yöntemi için primer tasarlama ve çalışma prensibi

gerçekleştirilmiştir. Yeni nesil dizilemede kullanılmak üzere amplikon kütüphanelerinin oluşturulması için öncelikle robotik sistemde çipteki her bir kuyuya DNA örneği ve barkodlu primer çiftleri eklenmiş ve bir kuyuda ayrı olarak PZR amplifikasyonu gerçekleştirilmiştir. Her bir primer çifti bir barkod taşıdığından amplifikasyon sonrası her hasta kendine ait bir barkod ile işaretli hale gelmektedir. Bu aşamadan sonra tüm amplikonlar tek bir tüpte toplanmış ve manyetik boncuklar kullanılarak kullanılmayan primerler ve PZR artıklarından temizlenmiştir. Manyetik boncuklar ile temizlenen ampikon kütüphanesinin Qubit florometre cihazı ile miktar tayini yapılmıştır (Qubit dsDNA HS Assay Kit, Thermo Fisher Scientific). Ampikon kütüphaneleri dizilemeye uygun hale getirildikten sonra illumina Miseq (Illumina, Kaliforniya, ABD) cihazında uygun dizileme protokolü kullanılarak dizilenmiştir. Çalışmamızda ampikonların uzunluğu 450-500 bp civarına kadar çıkabildiği için 500 dönü yapabilen V2 kitleri kullanılmıştır. Dizileme sonunda 7.5-8.5 GB data elde edilmiş ve dizileme yaklaşık 40 saat sürmüştür.

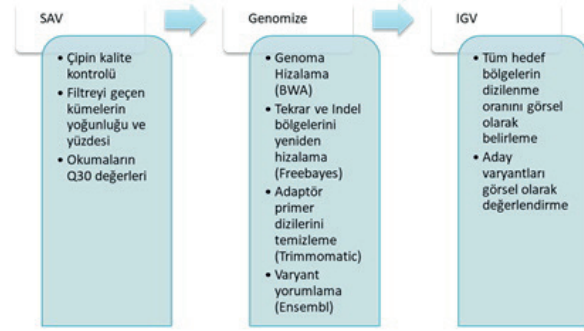
### Kalite/Kontrol Değerlendirilmesi

Ham datanın kalitesi ve dizileme metriklerini incelemek amacı ile Illumina'nın geliştirdiği "Sequence Analysis Viewer"-SAV programı kullanılmıştır. Çip üzerindeki ampikon yoğunluğunu ortalama 900-1300 K/mm<sup>2</sup> aralığında olan, çipteki okumaların en az %80'inde Q30 değer skoru ile dizilenmiş olması ve çip üzerinde filtreyi geçen kümelerin (cluster passing filter) yüzdesinin %80 ve üzeri olan dizilemeler çalışmaya dahil edilmiştir.

### Analiz İş Akışı ve Varyant Yorumlama

Miseq cihazından FASTQ formatında çıkan ham datanın kalite kontrolü yapıldıktan sonra Seq analiz programı (Genomize, İstanbul, Türkiye) ile ileri analizler yapılmıştır. Öncelikle Burrows-Wheeler Alignment (BWA) aracı ile ham datanın genomu hizalanması yapılmıştır (12). Panelin tüm tasarımları hg38'e göre yapılmış olduğundan genomu hizalama analizleri için hg38 genom versiyonu kullanılmıştır.

Genom hizalama işleminden sonra Freebayes ile varyantlar belirlenmiş, Trimmomatic ile adaptör primerlerin ampikon dizilerinden temizlenmesi işlemi yapılmıştır (13). Çoğaltılan bölgeler ve varyantları görsel olarak analiz etmek için IGV programı kullanılmıştır (14). Miseq cihazından alınan ham datadan varyantları belirlemeye kadar geçen tüm aşamalar özet olarak belirtilmiştir (Şekil 2).



Şekil 2. Ampikon dizileme analizinde kullanılan programlar ve takip edilen iş akışı

Yeni nesil dizileme ile tespit edilen varyantların yorumlanması için SIFT (15), Polyphen (16), Variant effect predictor (VEP) (17), Mutation Taster (18), SNPeff (19), Gene splicer (20) ve CADD (Combined Annotation Dependent Depletion) (21) programları kullanılmıştır. dbSNP veritabanı (<https://www.ncbi.nlm.nih.gov/SNP/>), 1000 Genom projesi örnekleri (<http://www.1000genomes.org/>), the Exome Variant Server (ESP) veritabanı (<http://evs.gs.washington.edu/EVS/>), The Human Gene Mutation Database (HGMD) veritabanı (<http://www.hgmd.cf.ac.uk/ac/index.php>) ve the Exome Aggregation Consortium (ExAC) (<http://exac.broadinstitute.org/>) veritabanı varyantların frekans verisinin belirlenmesi amacıyla kullanılmıştır.

## BULGULAR

Tasarlanan paneller ile 48 AKİY ve 64 PAY şüphesi olan hastalar dizilenmiştir. AKİY ve PAY paneli çiplerinin kümelenme yoğunluğu ortalama 800-1200K/mm<sup>2</sup>, filtreyi geçen küme yoğunluğu

ortalama %95 ve Q30 değerleri %80-90 aralığında görülmüştür. Tasarlanan YND panellerinde hedef genlerin tüm ekzonları, 5'UTR, 3'UTR bölgeleri ve ekzon-intron bağlantıları çalışmaya dahil edilmiştir.

AKİY panelinde tüm bölgelere primeri kapsayan primerler tasarlanmasına rağmen dizileme başarısı; hedef bölgelerde ortalama okuma kapsamı (coverage) %96,35, genlerin kodlayan bölgelerinin okuma kapsamı ise ortalama %97,89 olarak görülmüştür. Çalışmaya dahil edilen 18 aday genin 11'i yüksek kalitede dizilenirken *CORO1A* (ekzon 1, 2, 11), *ADA* (ekzon 1), *JAK3* (ekzon 1,13) ve *IL7R* (ekzon 7) bölgeleri hastalarda çoğaltılmamış ya da düşük güvenilirlikte tespit edilmiştir. *PRKDC*, *LIG4* ve *ZBTB24* genlerinin tüm kodlayan bölgeleri dizilenirken kodlama-yan bölgelerinde kaçırılan kısımlar olmuştur.

PAY panelinde ise tüm hedef bölgelerin ortalama okuma kapsamı %92,04 olarak görülmüşken, bu oran

kodlayan bölgelerde %92,96 olarak tespit edilmiştir. *TNFRSF12*, *CD19* ve *BTK* genlerindeki hedef bölgelerin tamamı dizilenmiş, *TNFRSF13C*, *CD40*, *AICDA*, *TCF3*, *TRNT1* ve *CD40L* genlerinin kodlamayan bölgelerinde dizilenmeyen bölgeler olsa da tüm kodlayan bölgeleri dizilenmiştir. Ayrıca *CR2* (ekzon 8,10 ve 11'in bir kısmı), *MS4A1* (ekzon 6), *PIK3R1* (ekzon 6,7), *TTC37* (ekzon 7,12), *CD79A* (ekzon 4), *BLNK* (ekzon 8,11), *NFKB2* (ekzon15 ve 18'in bir kısmı), *ICOS* (ekzon 2), *CD81* (ekzon 1,7), *UNG* (ekzon 7'nin bir kısmı), *IGLL1* (ekzon 2), *TNFRSF13B* (ekzon 4), *LRBA* (ekzon 10,17,31,33,48 ve 57) ve *CD79B* (ekzon 2) bölgeleri hastalarda dizilenememiş ya da düşük güvenilirlikte dizilenememiştir (Tablo 1).

### Aday Varyantların Belirlenmesi

Varyantlar öncelikle güvenilirliklerine göre filtrelenmiş ve en az 20 okuma derinliği olan (20X), hete-

**Tablo 1.** Tasarlanan AKİY ve PAY panellerinin okuma derinliği analizleri

Panel	Gen	Transkript	Cov (%)	Trans. cov (%)	Panel	Gen	Transkript	Cov (%)	Trans. cov (%)
Pay Paneli	TNFRSF12	NM_003809	100	100	Akiy Paneli	CD3E	NM_000733	100	100
	CD19	NM_001178098	100	100		NHEJ1	NM_024782	100	100
	BTK	NM_001287344	100	100		RAG2	NM_000536	100	100
	IGHM		100	100		CD3D	NM_000732	100	100
	TNFRSF13C	NM_052945	99,89	100		RAG1	NM_000448	100	100
	CD40	NM_001250	98,27	100		DCL-RE1C	NM_001033855	100	100
	AICDA	NM_020661	94,98	100		PNP	NM_000270	100	100
	TCF3	NM_003200	93,89	100		CD247	NM_198053	100	100
	TRNT1	NM_182916	88,57	100		PTPRC	NM_002838	100	100
	CD40LG	NM_000074	82,99	100		AK2	NM_001625	100	100
	CR2	NM_001006658	97,08	96,29		IL2RG	NM_000206	100	100
	MS4A1	NM_021950	96,84	96,18		PRKCD	NM_006254	87,64	100
	PIK3R1	NM_181523	92,83	95,79		ZBTB24	NM_014797	97,51	100
	TTC37	NM_014639	89,61	92,63		LIG4	NM_206937	98,74	100
	CD79A	NM_001783	95,53	91,73		CORO1A	NM_001193333	73,24	85,55
	BLNK	NM_013314	93,63	91,51		JAK3	NM_000215	90,87	89,5
	NFKB2	NM_001077494	91,2	90,83		IL7R	NM_002185	96,99	89,95
	CD81	NM_004356	79,25	90,48		ADA	NM_000022	89,47	97,04
	ICOS	NM_012092	91,55	89,43					
	UNG	NM_080911	88,31	86,97					
	IGLL1	NM_020070	87,18	82,03					
	TNFRSF13B	NM_012452	87,95	81,43					
	LRBA	NM_006726	83,07	80,54					
	CD79B	NM_001039933	84,5	72,38					

Kap: Kapsam, Trans Kap: Transkript Kapsamı



rozigot varyantlar için en az %20, homozigot varyantlar için ise en az %95 kapsamda görülen varyantlar yüksek güvenilirlikli varyantlar, bunun dışında kalan varyantlar ise düşük güvenilirlikli varyantlar olarak değerlendirilmiştir. İlk aşamada yüksek güvenilirlikli varyantlar analize dahil edilmiş fakat aday varyant bulunamayan hastalarda düşük güvenilirlikli varyantlar da Q30 skorlarına bakılarak analize dahil edilmiştir. İkinci aşama olarak, varyantlar öncelikle literatürdeki ve aynı dizileme panelinde dizilenen hastalar arasındaki alel frekanslarına göre değerlendirilmiş ve minör alel frekansı 0.05'in altındaki varyantlar ile analize devam edilmiştir. İkinci filtreleme işleminden sonra kalan varyantlar proteine etkisine göre tekrar değerlendirilmiş ve patojenik olduğu düşünülen varyantların hastanın kliniği ile uyumu araştırılmış ve genotip-fenotip ilişkisi uygun varyantlar aday varyant olarak değerlendirilmiştir. Çalışmada takip edilen iş akışının özeti aşağıdaki şekilde gibidir (Şekil 3).



Şekil 3. Hastalığa sebep olan aday varyantları belirlemede kullanılan analiz iş akışı ve çevrimiçi programlar

Analizler sonucunda, AKİY paneli ile hastaların %58'inde, PAY paneli ile de hastaların %14,2'sinde hastalığa sebep olan varyantlar tespit edilmiş, Sanger dizileme ile varyantların doğruluğu gösterilmiştir. Panel ile bulunan varyantların 11 tanesi ilk defa bu çalışma ile gösterilmiştir.

## TARTIŞMA

Yeni nesil dizileme teknolojileri aynı anda birçok genin incelenmesine imkan verdiği için PİY gibi yüksek genetik heterojenite gösteren hastalık gruplarının kesin moleküler tanısı için sıklıkla tercih edilen güncel bir yaklaşımdır (7). Hedefe yönelik yeni nesil dizileme yöntemleri aday genlerin aynı anda

incelenmesine olanak veren, hızlı, analizi tüm genom yaklaşımlarına göre daha basit, maliyeti düşük ve güvenilir bir yöntemdir. PİY'lerde hedefe yönelik yeni nesil dizileme ile hastalarda kesin genetik tanıyı belirleme oranı hastalık spesifik panellerde %25-40 civarındadır (10, 22-26). Bu farklılık hastalığın etiolojisinde var olan heterojenite ile orantılı gözükmektedir. Bu çalışmada, aday genlere özgü primerler tasarlanarak, AKİY ve PAY özgü iki ana amplikon dizileme paneli geliştirilmiş ve tasarlanan paneller PİY şüphesi olan hastalarda valide edilmiştir.

AKİY paneli ile hastaların %58'inde hastalığa sebep olan varyantlar tespit edilmiştir. Panel ile bulunan varyantların sekiz tanesi ile defa bu çalışma ile gösterilmiştir (27-29). AKİYlerin monogenik arka planı diğer immün yetersizliklere oranla daha iyi karakterize edilmiş ve sorumlu genleri büyük ölçüde tanımlanmış bir hastalık grubudur. Bu nedenle tasarlanan panelin tanı başarısı %60 civarındadır. Literatürde AKİY hastalarında yapılan panel ya da ekzom dizileme çalışmalarında başarı oranları %90 hatta %100'lere kadar çıkabilmektedir. AKİY, diğer PİY gruplarına göre genetik altyapısı bilinen, genetik ve klinik açıdan tanımlanabilen bir gruptur, bizim panel başarımızın daha düşük olmasının sebebi kısıtlı gen sayısına bağlıdır. Ayrıca, AKİY hastaları için, panel sonrası yapılan tüm ekzom dizileme analizlerinde hastaların bir kısmında primer immün yetersizliğe yol açan genetik sendromların (Ataksi telenjipektazi ve FOXN1 eksikliği gibi) olduğu görülmesi tasarlanan genetik tanı panellerinin daha kapsamlı olması gerektiği savını doğrulamaktadır.

Primer antikor yetersizlikleri (PAY) paneli ile 64 yeni tanı hasta dizilenmiş, hastaların %14,2'inde hastalık ile ilişkili varyantlar tespit edilmiştir. Literatürde PAY hastalarındaki genetik tanı başarısının %2-10 seviyesinde olduğu düşünülürse, tasarladığımız genetik tanı paneli bu hastalar için uygun başarı oranını yakalamıştır. Fakat bu başarı oranları PAY için tasarlanan genetik tanı panellerinin hastalığın tanısında kullanılabilmesi için yeterli bir seviyede değildir. Her geçen gün sayıları artan hastalıkla ilişkilendirilmiş gen/varyasyon bulunmaktadır. Ancak hala çok sayıda genetik sebebi açıklanamamış PAY fenotipi

tipleri bulunmaktadır. Genetik etiolojinin tam çözümlenememiş olması panel başarılarını doğrudan etkilemektedir. İlave olarak Tablo 1’de yer aldığı gibi her iki panelde de tüm genleri %100 kapsamayan dizileme sonuçları bu bölgelerde var olan varyasyonların tanımlanamamasına ve düşük panel başarısına sebebiyet vermektedir.

Amplikon dizileme yönteminde doğru kütüphane hazırlama yöntemini belirlemek, çalışmanın kalitesi ve doğruluğunu da belirlemektedir. Prob tabanlı sistemlerde, kullanıcılar tek tek prob tasarlamak güç olduğu için problemler genellikle çevrimiçi programlar ile otomatik olarak tasarlanmakta ve çalışma başladıktan sonra amplifiye olmayan bölgelere tekrar prob tasarlamak mümkün olamamaktadır. Ayrıca problemler sistemlerde genomik tekrarlar belirlenememekte ve hedef bölgenin genomda psödogeni varsa problemler psödogen bölgelerinden kaçamamaktadır. PZR tabanlı metotlarda ise tasarlanan primer çiftlerini yenileyerek çok kolay optimizasyon yapılabilmekte ve hedef dizileme kalitesini arttırılabilmek mümkündür. Bunun yanında kopya sayısı değişiklikleri analizleri her zaman yapılamamaktadır. Ayrıca hedef bölgeler çoğaldıkça primer maliyeti ve optimizasyon güçlüğü yaşanabileceğinden, geniş panellerde (200-300 gen) prob tabanlı, ufak panellerde ise PZR tabanlı yaklaşımlar tercih edilmektedir (30). Hedef bölgelerin tekrar bölgeleri veya psödogenler ile ortak dizileri olduğu takdirde PZR ve prob kombinasyonu ile hazırlanan kütüphaneler de tercih edilmektedir (10).

Bu çalışmada PZR temelli bir YND panelinin tercih edilmesi tekrar bölgeler ve psödogenler açısından zengin olan immün sistem genlerinin doğru olarak dizilenebilmesi içindir. Hedef bölgelerin psödogenler ile ortak dizileri sadece kodlayan gen bölgesine özgün tasarlanan primer ile dışlanabilmiştir. Örnek vermek gerekirse AKİY panelinde bulunan *DCLRE1C* geninin moleküler analizleri psödogeni ile yüksek benzerlik göstermesi sebebiyle oldukça zordur (31). Çalışmamızda tespit edilen p.L187\* varyantı genin psödogeni ile yüksek oranda benzerlik gösterdiği ekzon 8 bölgesinde yer almaktadır. Panel oluşturulduğunda tercih ettiğimiz amplikon dizileme metodu ile psödogen dizilerini hedeflemeyen primerler tasar-

lanmış ve kodlayan bölgedeki varyantları doğru tespit edebilmemiz mümkün olmuştur.

Bu çalışma ile ikinci olarak PİY’lerin genetik tanısında kullanılacak bir analiz algoritması oluşturulmuştur. Belirlenen iş akışı tüm tek gen hastalıklarının moleküler tanı sonucu ortaya çıkan varyantlarının yorumlanmasında kullanılacak bir akıştır. YNDnin ürettiği ham verinin yaklaşık %1,3’ünün yalancı negatiflikler olduğu bilinmektedir (32). Yalancı pozitiflik/negatifliklerin önüne geçilmesi için öncelikle aday varyantın ‘yüksek güvenilirlikte’ dizilenmiş olduğu kontrol edilmelidir. Yeni nesil verisi ile ortaya çıkan varyantların analizinde öncelikle aday varyantın kalite/kontrol parametrelerine göre en az 20 kere (20X) ve %99,9 (Q30) güvenilirlikle dizilenmiş olması çok büyük oranda yalancı negatiflik veya pozitiflik oranını düşürmek için bakılması gereken ilk parametrelerdendir. Bizim çalışmamızda aday tüm varyantlar Sanger dizileme ile tekrar doğrulanmıştır. Varyant yorumlamada ikinci aşama olarak minör alel frekansı (MAF) <0.01 olan varyantlar ile devam edilmiştir. Bugün patojenik varyantların frekanslarının %0,1’den küçük olduğu bilinmektedir (33). Varyantların frekansları literatürde düşük olsa da aday varyantın çalışmada aynı teknik ile dizilenen tüm hastalarda da frekansının düşük olması gerekmektedir. Zira olası PZR ve dizileme hataları yalancı pozitifliğe sebep olabilmektedir. Son olarak aday varyantın proteine etkisi araştırılmalı ve varyantın hastalık ile genotip-fenotip ilişkisini açıklıyor olması gereklidir.

## SONUÇ

Sonuç olarak, zaman ve maliyet açısından en doğru yaklaşım öncelikle sık görüldüğü bildirilen genler açısından hastayı taramak ve sonrasında daha ileri tüm genom yaklaşımlarına geçmektir. Hedefe yönelik yeni nesil dizileme sistemleri bu amaçla etkin olarak kullanılacak bir yaklaşımdır ve pek çok rutin genetik tanı uygulamalarında yerini bulmuştur. Çok sayıda aday genin tek seferde ve kısa süre içerisinde analizini uygun fiyatla, yüksek hassasiyette ve güvenilir bir şekilde mümkün kılmaktadır. Fakat yYND ile ortaya çıkan verinin yorumlanması oldukça zordur. Bunun için hastanın öyküsü ve aile ağacı ayrıntılı alınarak

hastalığın kalıtım modeli doğru belirlenmeli, hastalığa uygun analiz iş akışı belirlendikten sonra moleküller tanı mutlaka ayrıntılı muayene, immünotipleme ve basit fonksiyonel testler ile desteklenmelidir.

**Teşekkür:** Çalışmaya dahil edilen hasta örneklerini sağlayan Doç. Dr. Selda Hançerli Torun, Prof. Dr. Elif Aydın, Doç. Dr. Ayca Kiykim, Prof. Dr. Yıldız Camcioglu, Prof. Dr. Safa Baris ve Prof. Dr. Ahmet Ozen'e, Çalışmaya bilimsel katkıları için Prof. Dr. Ugur Ozbek'e teşekkürlerimizi sunarız.

**Acknowledgement:** We would like to thank to Dr. Selda Hancerli Torun, Dr. Elif Aydin, Dr. Ayca Kiykim, Dr. Yıldız Camcioglu, Dr. Safa Baris and Dr. Ahmet Ozen for providing patient samples and thank Dr. Ugur Ozbek for his scientific contributions for the study.

**Hakem Değerlendirmesi:** Dış bağımsız.

**Peer Review:** Externally peer-reviewed.

**Yazar Katkıları:** Çalışma Konsepti/Tasarım- S.F., Y.Y.N. M.S., Ö.H.N.; Veri Toplama- S.F., Y.Y.N.; Veri Analizi/Yorumlama- S.F., M.S., Y.Y.N.; Yazı Taslağı- S.F., M.S.; İçeriğin Eleştirel İncelemesi- Y.Y.N., M.S., Ö.H.N.; Son Onay ve Sorumluluk- S.F., Ö.H.N., M.S., Y.Y.N.

**Author Contributions:** Conception/Design of Study- S.F., Y.Y.N. M.S., Ö.H.N.; Data Acquisition- S.F., Y.Y.N.; Data Analysis/Interpretation- S.F., M.S., Y.Y.N.; Drafting Manuscript- S.F., M.S.; Critical Revision of Manuscript- Y.Y.N., M.S., Ö.H.N.; Final Approval and Accountability- S.F., Ö.H.N., M.S., Y.Y.N.

**Çıkar Çatışması:** Yazarlar çıkar çatışması beyan etmemişlerdir

**Conflict of Interest:** Authors declared no conflict of interest.

**Finansal Destek:** Bu çalışma, İstanbul Üniversitesi Bilimsel Araştırma Projeleri Birimi (Proje numarası: 52575 ve 20499) ve İstanbul Bilgi Üniversitesi araştırma fonu (NGYY-2018.01.0006) tarafından desteklenmiştir.

**Financial Disclosure:** This project was supported by Istanbul University Research Fund (Project no: 52575 and project no: 20499) and Istanbul Bilgi University Research Fund (NGYY-2018.01.0006).

## KAYNAKLAR/REFERENCES

1. Goodwin S, JD, McPherson, WR McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016; 17(6):333-51.
2. Shendure J, H Ji. Next-generation DNA sequencing. *Nat Biotechnol*, 2008; 26(10):1135-45.
3. Mardis ER. DNA sequencing technologies: 2006-2016. *Nat Protoc* 2017; 12(2):213-8.
4. Giani AM, Gallo LGR, Gianfranceschi G, Formenti, Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J* 2020; 18:9-19.
5. Pettersson E, Lundeberg J, Ahmadian A. Generations of sequencing technologies. *Genomics* 2009; 93(2):105-11.
6. Cifaldi C, Brigida I, Barzaghi F, Zoccolillo M, Ferradini V, Petricone D et al. Targeted NGS Platforms for Genetic Screening and Gene Discovery in Primary Immunodeficiencies. *Front Immunol* 2019;10:316.
7. Raje N S, Soden D, Swanson CE, Ciaccio SE, Kingsmore DL, Dinwiddie. Utility of next generation sequencing in clinical primary immunodeficiencies. *Curr Allergy Asthma Rep* 2014; 14(10):468.
8. Notarangelo LD. Primary immunodeficiencies. *J Allergy Clin Immunol* 2010; 125(2 Suppl 2):S182-94.
9. Parvaneh N, Casanova JL, Notarangelo LD, Conley ME. Primary immunodeficiencies: a rapidly evolving story. *J Allergy Clin Immunol* 2013; 131(2):314-23.
10. Stoddard JL, Niemela JE, Fleisher TA, Rosenzweig SD. Targeted NGS: A Cost-Effective Approach to Molecular Diagnosis of PIDs. *Front Immunol* 2014;5:531.
11. Al-Herz, Bousfiha WA, Casanova JL, Chatila T, Conley ME, Cunningham-Rundles C, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front Immunol* 2014; 5:162.
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25(14):1754-60.



13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30(15):2114-20.
14. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011; 29(1):24-6.
15. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4(7):1073-81.
16. Adzhubei IA, Schmidt S, L. Peshkin, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; 7(4):248-9.
17. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* 2016; 17(1):122.
18. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014; 11(4):361-2.
19. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; 6(2):80-92.
20. Perteau M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; 29(5):1185-90.
21. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46(3):310-5.
22. Nijman IJ, Montfrans JM van, Hoogstraat M, Boes ML, van de Corput L, Renner ED, et al. Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies. *J Allergy Clin Immunol* 2014; 133(2):529-34.
23. Al-Mousa H, Abouelhoda M, Monies DM, Al-Tassan N, Al-Ghoniaim A, Al-Saud B, et al. Unbiased targeted next-generation sequencing molecular approach for primary immunodeficiency diseases. *J Allergy Clin Immunol* 2016; 137(6):1780-7.
24. Yu H, Zhang VW, Stray-Pedersen A, Hanson IC, Forbes LR, de la Morena MT, et al. Rapid molecular diagnostics of severe primary immunodeficiency determined by using targeted next-generation sequencing. *J Allergy Clin Immunol* 2016; 138(4):1142-1151 e2.
25. Fang M, Abolhassani H, Lim CK, Zhang J, Hammarstrom L. Next Generation Sequencing Data Analysis in Primary Immunodeficiency Disorders - Future Directions. *J Clin Immunol* 2016; 36 Suppl 1:68-75.
26. Erman B, Bilic I, Hirschmugl T, Salzer E, Boztug H, Sanal O, et al. Investigation of Genetic Defects in Severe Combined Immunodeficiency Patients from Turkey by Targeted Sequencing. *Scand J Immunol* 2017; 85(3):227-34.
27. Firtina S, Yin Ng Y, Hatirnaz Ng O, Kiykim A, Aydiner E, Nepesov S, et al. Mutational landscape of severe combined immunodeficiency patients from Turkey. *Int J Immunogenet* 2020;00:1-10.
28. Firtina S, Ng YY, Ng OH, Nepesov S, Yesilbas O, Kilercik M, et al. A novel pathogenic frameshift variant of CD3E gene in two T-B+ NK+ SCID patients from Turkey. *Immunogenetics* 2017; 69: 653-9.
29. Firtina S, Cipe F, Ng YY, Kiykim A, Ng OH, Sudutan T, et al. A Novel FOXP1 Variant Is Identified in Two Siblings with Nude Severe Combined Immunodeficiency. *J Clin Immunol* 2019; 39(2):144-147.
30. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods* 2010; 7(2):111-8.
31. Pannicke U, Honig M, Schulze I, Rohr J, Heinz GA, Braun S, et al. The most frequent DCLRE1C (ARTEMIS) mutations are based on homologous recombination events. *Hum Mutat* 2010; 31(2):197-207.
32. Mu W, Lu HM, Chen J, Li S, Elliott AM. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J Mol Diagn* 2016; 18(6):923-32.
33. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536(7616):285-91.