

# Tüm Genom SNP Genotipleme ile Trizomi 21 ve Ebeveyn Etkisinin Tespiti

## Detecting Parental Origin of Trisomy 21 with Whole Genome SNP Genotyping

Fatma Yeşim Kesim<sup>1</sup>, Feyza Nur Tuncer<sup>1</sup>, Emrah Yücesan<sup>1</sup>,  
Özkan Özdemir<sup>1</sup>, Mustafa Çalık<sup>2</sup>, Uğur Özbek<sup>1</sup>, Sibel Aylin Uğur<sup>1</sup>

<sup>1</sup>İstanbul Üniversitesi, Deneysel Tıp Araştırma Enstitüsü,  
Genetik Anabilim Dalı, İstanbul, Türkiye.

<sup>2</sup>Harran Üniversitesi, Tıp Fakültesi, Çocuk Nörolojisi Bölümü, Şanlıurfa, Türkiye

### ÖZET

Tüm genom SNP (Single Nucleotide Polimorphism, SNP) genotipleme, bağlantı analizi ve ilişkilendirme çalışmalarında kullanılan etkin bir yöntemdir. Bu yöntem ile her bir SNP'e ait allel frekans dağılımı ve sinyal yoğunluğu elde edilebildiğinden, kopya sayısı değişikliklerinin (Copy Number Variation, CNV) bilgisayar destekli olarak belirlenmesinde de yaygın olarak kullanılmaya başlanmıştır. Bu çalışmada, akraba evliliği yapmış ve çekinik geçişli bir progresif epilepsi türüne sahip bir ailede bağlantı analizi amacıyla gerçekleştirilmiş olan tüm genom SNP genotipleme verisi CNV açısından da değerlendirilmiştir. Tüm genom SNP genotipleme aynı aileden 8 çocuk (6 etkilenmiş ve 2 etkilenmemiş) ve bu çocukların etkilenmemiş ebeveynlerinde HumanCytoSNP-12 BeadChip kiti (Illumina) kullanılarak gerçekleştirilmiştir. Bu bireylere CNV analizi uygulandığında sağlıklı olarak rapor edilen bireyin 21. kromozomunda trizomi olduğu tespit edilmiştir. Haplotip incelemesine dayalı olarak geliştirilen yeni teknik ile trizominin maternal kaynaklı olduğu gösterilmiştir. Bu çalışmamızda, SNP array veri analizi trizomilerde ebeveyn etkisinin gösterilebileceği yeni ve etkin bir yöntem olarak sunulmuştur.

**Anahtar Kelimeler:** Tüm genom SNP genotipleme, Trizomi 21, Kopya Sayısı Analizi, Ebeveyn Etkisi

### ABSTRACT

Whole genome SNP (Single Nucleotide Polymorphism, SNP) genotyping is a well-established tool for linkage studies and association analysis. This tool offers further advance in detecting copy number variations (CNVs), since allele frequencies and signal intensities can be analyzed simultaneously using well-established algorithms with graphical outputs. In this study, SNP based CNV analysis has been performed for a consanguineous pedigree afflicted with a recessive form of progressive epilepsy. 6 affected, 4 unaffected sibs and their unaffected parents have been genotyped using Illumina Human HumanCytoSNP-12 BeadChip (300K). Interestingly, in CNV analysis we have revealed trisomy of chromosome 21 who was reported to be unaffected for the neurological condition. A haplotype based strategy using genotyping data from parents and the trisomic child revealed maternal inheritance of the extra chromosomal copy. Herein, we present an effective novel approach for parent of origin detection in trisomies.

**Keywords:** Whole genome SNP genotyping, Trisomy 21, CNV analysis, Parent of Origin

## GİRİŞ

İnsan genomundaki kopya sayısı değişiklikleri (Copy Number Variations; CNVs) DNA segmentlerinde kayıp ve kazançlarla ifade edilen genomik yeniden düzenlenmelerdir (1), (2). CNV'ler hastalıklarla ilişkili olabildikleri gibi toplumda sık görülen ancak patolojik olmayan varyantlar olarak da genomda yer alırlar (1), (3). Uzunca bir süre CNV'lerin insan genomundaki varyasyonların çok küçük bir kısmını temsil ettiği düşünüldüğünden önemi anlaşılamamıştır (4).

Günümüzde CNV'leri de içeren birçok kromozomal aberasyon yüksek çözünürlüğe sahip tüm genom teknolojilerinden Array Karşılaştırmalı Genomik Hibridizasyon (Array Comparative Genomic Hybridization; aCGH) ve Tek Nükleotit Polimorfizm (Single Nucleotide Polymorphism; SNP) array yöntemleri ile tespit edilebilmektedir (5), (6).

SNP arrayleri çekici kılan özellik kullanıcıya 2 basamaklı analiz imkanı sağlamasıdır. İlk basamakta DNA'nın taşıdığı SNP alleli ile uyumlu prob dizisine bağlanması ve tek baz uzama prensibine dayalı uygulamasıyla allel dağılımının tespiti yapılır. Analize ekstra bir boyut veren diğer bir parametre de B Allel Frekansdır (BAF). Her bir SNP'e ait 2 allel baz yapısı ne olursa olsun A ve/veya B allel frekansı cinsinden ifade edilir. Böylece heterozigot olan (AB) ve olmayan (Loss of Heterozygosity,

LOH; AA veya BB) bölgeleri belirlenebilir. İkinci basamakta ise, hasta bireylerin hedef sinyal yoğunlukları referans sinyal yoğunlukları ile karşılaştırılarak her bir SNP'le ilişkili bir CNV değeri elde edilir. CNV değeri için kullanılan sinyal yoğunluğu Log R kesiri (Log R Ratio; LRR) ile ifade edilir ve bu değer toplam sinyal yoğunluğunun normalize bir ölçütüdür. CNV normalizasyonu aCGH'teki gibi 2 farklı örneğin array öncesi karıştırılıp, analiz edilmesi ile değil, çalışılan tek bir örneğin belirli sayıda farklı örnek ile bilgisayar ortamında karşılaştırılması ile yapılır. Elde edilen CNV değeri sayesinde LOH bölgelerinin gerçek homozigotluğu mu yoksa bir delesyon kaynaklı allel kaybı mı olduğu anlaşılabilir. BAF ve CNV değerlerini bir arada kullanmak, düşük seviyede mozaiklik hakkında da bilgi verebilir (7). Sonuç olarak SNP array ile genomdaki kopya sayısı değişiklikleri hem BAF hem de LRR birlikte değerlendirilerek incelenmektedir (8).

SNP array verisi ile allel frekans dağılımının belirlenebilmesi sayesinde bireylerin her bir SNP pozisyonu için genotipleri tespit edilebilmektedir (6). Bu sayede genomdaki kayıp ve kazanç bölgeleri için ve anne veya baba kaynaklı bir anomali olması durumunda genotip verilerinden faydalanılarak bu anomalinin ebeveyn etkisi belirlenebilir. Bu çalışmamızda Trizomi 21 örneği üzerinden geliştirdiğimiz yöntem ile tüm trizomilerde ebeveyn etkisinin araştırılabilmesi için kullanılacak hızlı ve etkili bir algoritmayı sunmaktayız.

## GEREÇ ve YÖNTEM

### GEREÇ

#### Örnekler

Çalışmaya katılan tüm bireylerden İstanbul Üniversitesi İstanbul Tıp Fakültesi Etik Kurulunun 2012/844-1076 numaralı etik onayına dayanarak bilgilendirilmiş onam formu toplanmıştır. Klinik inceleme Harran Üniversitesi Tıp Fakültesi Çocuk Nörolojisi Bölümü tarafından gerçekleştirilmiştir.

#### Örnek Hazırlanması ve SNP Array Analizi

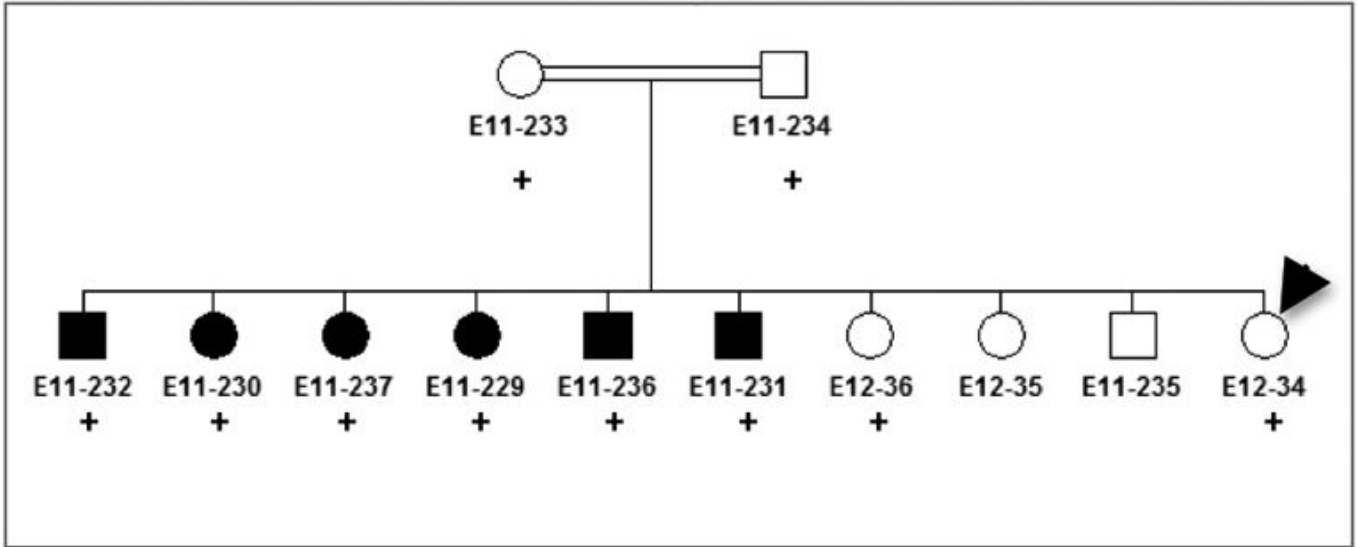
DNA periferik kan örneklerinden üreticinin protokolüne uygun olarak QIAamp DNA Blood Maxi Kiti (Qiagen, Hilden,

Germany) kullanılarak izole edilmiştir. DNA örneklerinin kalite ve miktar ölçümleri NanoDrop UV-Vis spektrofotometrede (Thermo Scientific) 260 ve 280 nm'de absorbans (A260/A280) değerleri karşılaştırılarak ve etidium bromür ile boyanmış %1'lik agaroz jel elektroforezi kullanılarak gerçekleştirilmiştir. A260/A280 değeri 1.8- 2.0 arasında olan ve jel görüntü degradasyona uğramamış DNA'yı işaret eden örnekler çalışmaya dahil edilmiştir. Bu kalite kontrol aşamasını geçen her örnek konsantrasyonu 50 ng/ul ve miktarı 200 ng olacak şekilde sabitlenmiştir. Bu çipteki prob aralığı ortalama 10 kb olarak hesaplanır ve genoma yayılmış 220.000 adet belirteci içerir. Bu belirteçler arasında 200.000 polimorfik tagSNP ile sadece CNV değeri veren bilinen sitogenetik anomalilere özgü 250 adet ve hastalıkla ilişkili genleri hedefleyen 400 adet prob sayılabilir.

Bu çip 12 örneğin birden çalışılabilmesine olanak sağlamaktadır.

Bu çalışmada akraba evliliği yapmış ve progresif epilepsinin segregasyonu ettiği geniş bir aile kullanılmıştır (Şekil 1). Toplam 9 kişiye ait SNP genotipleme tüm genom bağlantı ve CNV analizi yapılmak üzere elde edilmiştir. E12-34 numaralı bireyde Trizomi 21 tespit edilince, bu makalede detayları anlatıldığı üzere SNP array verisi üzerinden trizomilerde ebeveyn etkisini tespit edebilecek bir yöntem geliştirilmiştir. Örnekler standart protokole (Illumina Infinium HD Ultra Assay) uygun olarak sırasıyla tüm genom amplifikasyon, fragmentasyon, hibridizasyon ve tek baz uzama aşamalarından geçirilmiş, ortaya çıkan floresan ışımalar Illumina iScan Reader ile taranmıştır. Tarama sonrası oluşan görsel veri (idat dosyaları) Illumina'nın bir yazılımı olan

GenomeStudio programında Genotipleme Analizi Modülü kullanılarak incelenmiştir. Problemlere ait genomik koordinatlar Hg19 insan genom düzenlenmesine göre sıralanmıştır. BAF ve LogR değerleri ise Illumina, Inc. tarafından sağlanan ve yaklaşık 120 kontrol örneğine ait bilgiyi içeren HumanCytoSNP küme dosyası ile karşılaştırılarak elde edilmiştir. Kaliteli bir veri eldesinin kriteri her örnek için okuma oranının (call rate) %99'dan büyük olması olarak belirlenmiştir. Çipe yüklenen örneklerin sırası ile X ve Y kromozom sonuçları temel alınarak hesaplanan cinsiyetlerin uyumluluğu karşılaştırılarak aplikasyon sırasında ortaya çıkmış olabilecek herhangi bir karışıklık veya kros-kontaminasyon riski değerlendirilmiştir.



**Şekil.1.** Akraba evliliği yapmış ve çekinik geçişli bir progresif epilepsi türüne sahip aile ağacı. Genotip verisi olan bireyler '+' ile işaretlenmiştir. Ok ile işaretlenmiş bireyde Trizomi 21 gözlenmiştir.

### Kopya Sayısı Değişikliklerinin Tespiti ve Analizi

Kopya sayısı değişiklikleri sinyal yoğunluğu ile ilgili bir parametre olan LRR ve her bir SNP için allel dağılımını temsil eden BAF ile birlikte değerlendirilerek belirlenmiştir. 0 civarında LRR değeri için CNV beklenmez; artmış ve azalmış değerleri için ise sırasıyla insersiyon ve delesyon düşünülür. BAF profilindeki beklenti dışı sapmalar CNV tespitini kolaylaştırmaktadır. Bu çalışmada, tüm kopya sayısı değişiklikleri Illumina GenomeStudio yazılımı ile hem görsel olarak tespit edilmiş hem de ek bir yazılım olan cnvPartition ile tüm CNV değerleri excel tabanlı bir dosyada pozisyon, boy, aralık ve

türleri ile daha detaylı analiz edilmiştir (Tablo 1).

### Ebeveyn Etkisinin Tespiti ve Analizi

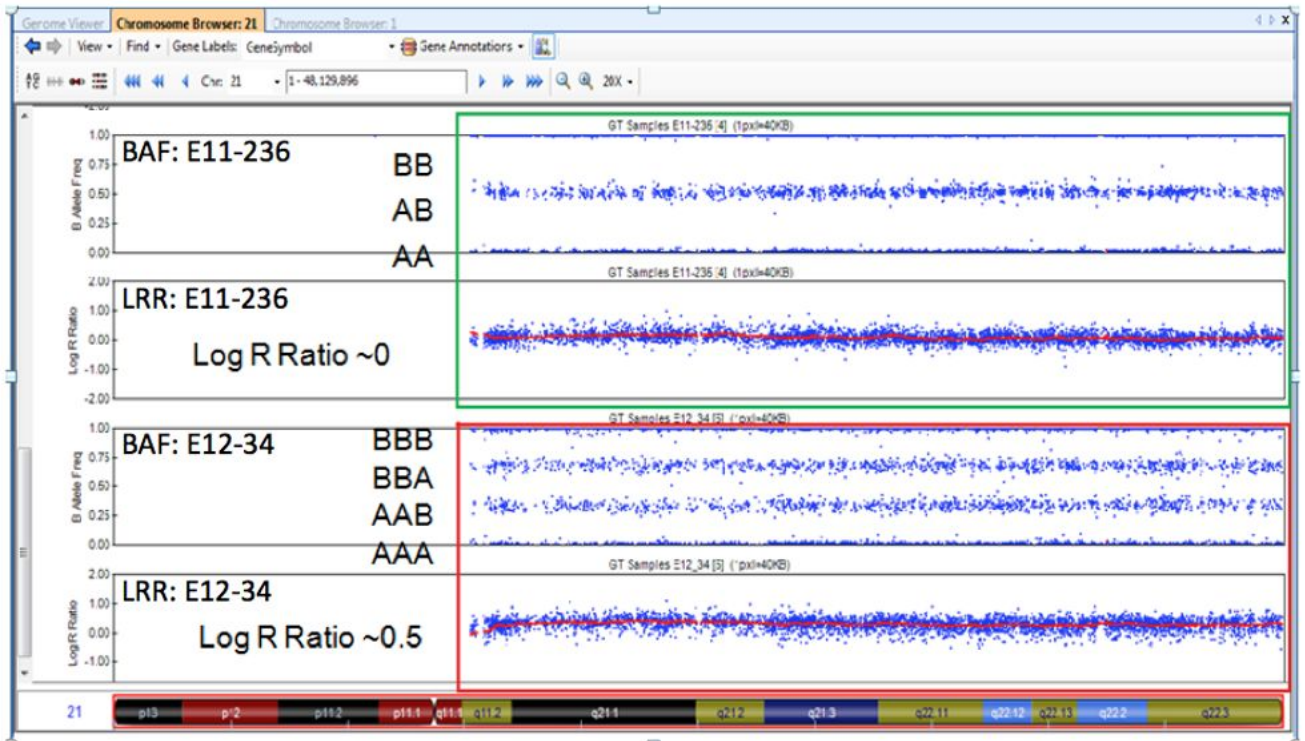
Ebeveyn etkisinin tespiti için GenomeStudio programından 21. Kromozoma ait SNP haritası ve ilgili allel frekansları listelenmiştir. Bu analiz için ebeveynler (E11-233 ve E11-234), trizomi 21'li çocuk (E12-34) ve iki kardeşi (E11-229 ve E11-235) kullanılmıştır. Bu liste Excel programında çeşitli filtrelenerek incelenmiştir.

## BULGULAR

### Trizomi 21 tespiti:

Şekil 1’de gösterilen ve orijinalinde bağlantı analizi için 10 kişinin tüm genom genotip verisinin elde edildiği aile olası kromozom anomalilerine karşı her kromozom için CNV açısından incelenmiştir. Bu analizde E12-34 kodlu bireyin kromozom 21 boyunca trizomik olduğu gözlenmiştir. Şekil 2’de E12-34 kodlu bireyin BAF ve LRR açısından kromozom

21 veri görüntüleri normal bir 21. kromozoma sahip kardeşi ile (E11-236) karşılaştırmalı olarak verilmiştir. LRR değerlerinde artış ve BAF frekans (f) grafiğinin 3 panel (fAA=0, fAB=0.5, fBB=1) halinde değil de farklılaşmış 4 panel halinde gözlenmesi (fAAA=0, fAAB=0.33, fABB=0.66, fBBB=1) bu bölgede 3 kopya artışı işaret etmektedir. Kromozom 21’in tamamıyla 3 kopya olması yani trizomi 21 Down Sendromu ile ilişkilidir. Klinik olarak da E12-34 Down sendromludur.

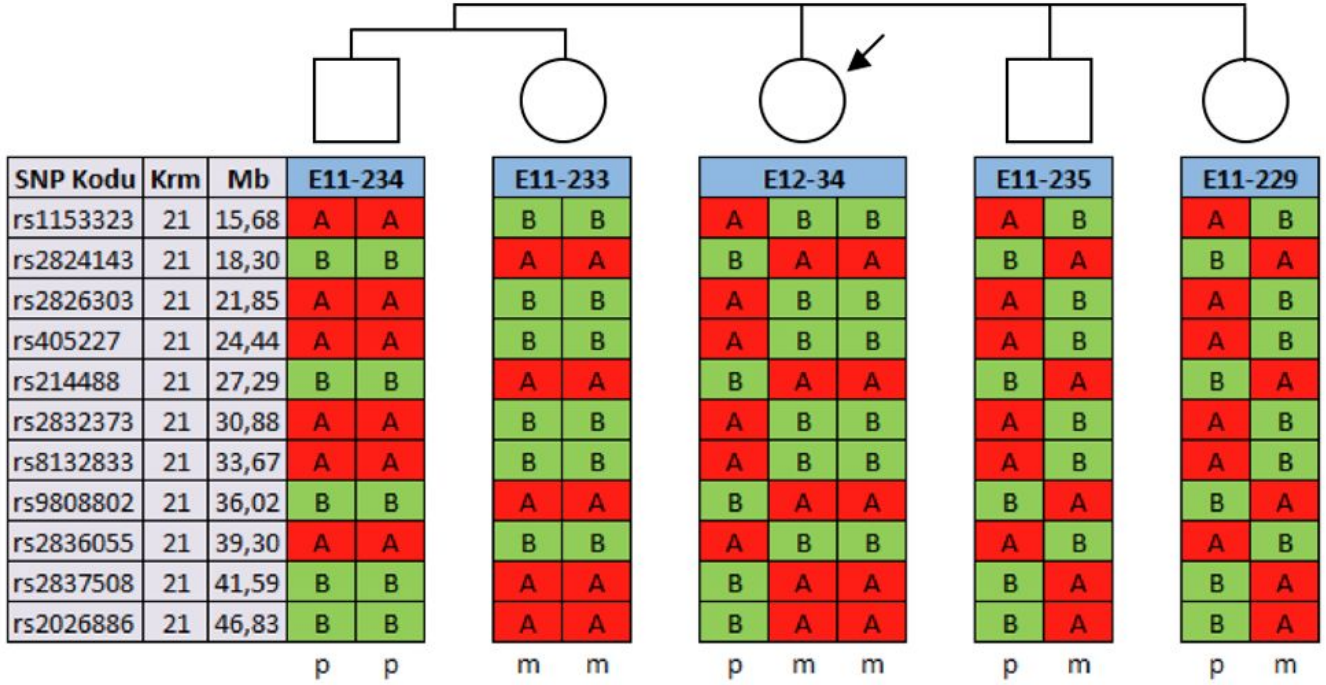


Şekil 2. Normal ve trizomiye sahip 21.kromozom için CNV analizi

### Ebeveyn Etkisinin Tespiti:

E12-34 kodlu bireyde trizomi 21 gözlenince bu çalışma kapsamında ebeveyn ve çocuk genotipleri kullanılarak ebeveyn etkisinin tespit edilebilmesi açısından yeni, hızlı ve efektif bir yöntem geliştirilmiştir. Bu amaçla öncelikle iki ebeveynin ve trizomi 21’li çocuğun genotipleri 21 kromozom açısından listelenmiş, daha sonra excel programı üzerinden harf ve renk koduna göre filtrelenerek sadece annede ve babada homozigot ancak birbirinden farklı SNP’lerin genotipleri bırakılmıştır. Bu analiz ile 21. Kromozom boyunca yer alan 4326 adet SNP’den sadece 11 tanesi kullanılmıştır. GenomeStudio Genotipleme

Modülüne ait genotip listesi sadece 2 alleli durumu hesaplayabildiğinden bu 11 SNP’e ait 3 allel E12-34 numaralı bireyde frekans değerleri üzerinden hesaplanmıştır. Yani, herhangi bir SNP’e ait genotip BAF değeri 0 ise AAA, 0,33 ise AAB, 0,66 ise ABB ve 1 ise BBB olarak hesaplanmıştır. Filtreleme ve frekansa göre allel dağılımı hesaplama yöntemiyle E12-34’te hangi allelin hangi ebeveyninden geldiği kolaylıkla anlaşılabilir (Şekil 3). Analizin daha kolay anlaşılabilmesi açısından kromozom 21’leri CNV açısından normal 2 kardeş de analize dahil edilmiştir. Sonuç olarak E12-34 numaralı bireyin ekstra kromozom kopyasının annesinden (E11-233) geldiği tespit edilmiştir.



**Şekil 3.** Ebeveyn etkisinin tespiti için kromozm 21 boyunca yer alan 11 SNP kullanılmıştır. Harf ve renk kodlarına göre filtrelenen bu SNP’lerde aranan özellik annede ve babada farklı homozigotluğa sahip olmasıdır. Bu özelliğe sahip 11 SNP için E12-34 kodlu bireyde frekansa dayalı genotipler manuel olarak belirlenmiştir. Böylece E12-34 için ekstra kromozomal kopyanın anneden geldiği açıkça gözlenmiştir. (Kısaltmalar; p: paternal; m: maternal; krm: kromozom)

İsim	Krm	Başlangıç (bç)	Bitiş (bç)	Aralık (Mb)	CNV	Yorum
Örnek	2	111.392.259	113.100.014	1,71	1	-
Örnek	1	49.260.879	50.589.798	1,33	2	LOH
Örnek	X	126.836.697	128.035.154	1,20	2	-
Örnek	1	147.806.652	149.400.252	1,59	3	-
Örnek	7	61.074.194	61.804.394	0,73	4	-

Tablo 1. CNV değerleri excel tabanlı olarak pozisyon, boy, aralık ve türleri açısından incelenebilir. CNV 1=tek allel kayıp, 2=nötr (LOH), 2=kazanç (X kromozomu, erkek birey), 3= tek allel kazanç ve 4=iki allel kazanç. (Kısaltmalar; Krm:Kromozom, bç: baz çifti; Mb: Megabaz)

## TARTIŞMA

Bu çalışmada orjinalinde bağlantı analizi yapılmak üzere tüm genom genotipleme yapılmış olan bir ailenin verisi kullanılmıştır. Progresif epilepsi fenotipi açısından sağlıklı olarak rapor edilen bir çocukta SNP verisinin çok yönlü kullanımı sayesinde öncelikle Trizomi 21 tespit edilmiştir. Daha sonra ise bu bireyin ve ebeveynlerinin kromozom 21'e ait genotip verisi özel olarak filtrelenerek ve frekans bazlı genotip analizi yapılarak

trizomi 21 için ebeveyn etkisi tespit edilebilmiştir.

Geliştirdiğimiz yöntem sadece trizomi 21 için değil tüm trizomiler için kolaylıkla uygulanabilir. Trizominin kaynağı etkin bir şekilde tespit edildiğinde yaşam koşulları, yaş gibi kriterlerin trizomi oluşumunda etkileri incelenebilir.

## SONUÇ

Bu çalışmada sonuç olarak trizomilere özgü ebeveyn etkisinin tespit edilebilmesi için SNP array yöntemine dayalı hızlı ve

kolay bir yöntem geliştirilmiştir.

## TEŞEKKÜRLER

Aile bireylerinin klinik incelemelerini yapan Doç. Dr. Mustafa Çalık' a ve çalışmaya katılan tüm aile bireyelerine teşekkür ederiz. Bu çalışma TÜBİTAK 1001 programı (Proje No: 113S331) ve İstanbul Üniversitesi Bilimsel Araştırma

Projeleri Birimi (Proje No: 43975) tarafından desteklenmiştir. FYK ve EY TÜBİTAK 113S331 projesi kapsamında bursiyer olarak görev almaktadırlar.

## KAYNAKLAR:

1. Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles, 2006, Global variation in copy number in the human genome: *Nature*, v. 444, p. 444-54.
2. Valsesia, A., B. J. Stevenson, D. Waterworth, V. Mooser, P. Vollenweider, G. Waeber, C. V. Jongeneel, J. S. Beckmann, Z. Kutalik, and S. Bergmann, 2012, Identification and validation of copy number variants using SNP genotyping arrays from a large clinical cohort: *BMC Genomics*, v. 13, p. 241.
3. Li Mura, I. E., B. Bauce, A. Nava, M. Fanciulli, G. Vazza, E. Mazzotti, I. Rigato, M. De Bortoli, G. Beffagna, A. Lorenzon, M. Calore, E. Dazzo, C. Nobile, M. Luisa Mostacciolo, D. Corrado, C. Basso, L. Daliendo, G. Thiene, and A. Rampazzo, 2013, Identification of a PKP2 gene deletion in a family with arrhythmogenic right ventricular cardiomyopathy: *Eur J Hum Genet*, v. 21, p. 1226-31.
4. Gökçümen, O., and C. Lee, 2009, Copy number variants (CNVs) in primate species using array-based comparative genomic hybridization: *Methods*, v. 49, p. 18-25.
5. Tsai, E. A., M. A. Berman, L. K. Conlin, H. L. Rehm, L. J. Francey, M. A. Deardorff, J. Holst, M. Kaur, E. Gallant, D. M. Clark, J. T. Glessner, S. T. Jensen, S. F. Grant, P. J. Gruber, H. Hakonarson, N. B. Spinner, and I. D. Krantz, 2013, PECONPI: A novel software for uncovering pathogenic copy number variations in non-syndromic sensorineural hearing loss and other genetically heterogeneous disorders: *Am J Med Genet A*, v. 161, p. 2134-47.
6. Scharpf, R. B., T. H. Beaty, H. Schwender, S. G. Younkin, A. F. Scott, and I. Ruczinski, 2012, Fast detection of de novo copy number variants from SNP arrays for case-parent trios: *BMC Bioinformatics*, v. 13, p. 330.
7. Peiffer, D. A., J. M. Le, F. J. Steemers, W. Chang, T. Jenniges, F. Garcia, K. Haden, J. Li, C. A. Shaw, J. Belmont, S. W. Cheung, R. M. Shen, D. L. Barker, and K. L. Gunderson, 2006, High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping: *Genome Res*, v. 16, p. 1136-48.
8. Teumer, A., F. D. Ernst, A. Wiechert, K. Uhr, M. Nauck, A. Petersmann, H. Völzke, U. Völker, and G. Homuth, 2013, Comparison of genotyping using pooled DNA samples (allelotyping) and individual genotyping using the affymetrix genome-wide human SNP array 6.0: *BMC Genomics*, v. 14, p. 50.