Tekil hücre genom analizi yoluyla sağlıklı ve hasta bireylerin beyinlerinin incelenmesi

Program Kodu: Uluslararası ERANET-RUS

Proje No: 215Z495

Proje Yürütücüsü: Doç. Dr. Mehmet SOMEL

Bursiyerler:

Poorya Parvizi

Zeliha Gözde Turan

Etka Yapar

OCAK 2019

ANKARA

ÖNSÖZ

Final raporu sunulmakta olan bu çalışmanın ana amacı, insan beyninde somatik mutasyon kökenli kromozom kopya sayısı varyasyonu (KR-KSV) yaygınlığını kapsamlı biçimde ölçmek ve bu olgunun beyin yaşlanması ve Alzheimer Hastalığı'nda (AH) rolü olduğu hipotezini test etmektir. Proje, Leipzig Üniversitesi koordinatörlüğünde yürütülen "SIGNIFICANS" isimli ERANET-RUS PLUS projesinin parçasıdır.

Çalışmanın grubumuza ait olan ve TÜBİTAK tarafından desteklenen kısmının amacı, sinir hücrelerinde KR-KSV sıklığını tekil hücrelerden genom dizileme verisi kullanarak tespit etmektir. Bu kapsamda 17 bireyden alınmış 1190 hücreye ait genom verisi üretilmiştir. Çalışmanın deneysel aşamaları (örnek hazırlanması, kütüphane hazırlama ve dizileme) için Leipzig Ü. araştırmacılarıyla ortaklık ve hizmet alımı kullanılmış, üretilen geniş kapsamlı verinin hesaplamalı ve istatistiksel analizi grubumuzca yürütülmüştür.

Bilgimiz dahilinde ilk defa bu projede, genom dizileme yoluyla KR-KSV analizinde spesifik hücre tiplerinin incelenmesine izin verebilen piramidal nöronlara odaklanılabilmiştir.

Elde edilen verinin analizi, sinir hücreleri arasında somatik KR-KSV'nin düşük frekansta (muhtemelen %10'un altında) olduğuna, ayrıca Alzheimer Hastalığı'nda KR-KSV'nin bariz bir rolü olmadığına işaret etmiştir. Bu sonuç, geçmişte yapılan sitometri çalışmalarından farklı, ancak yakın zamanlı tekil hücre genom dizileme çalışmalarıyla tutarlıdır.

Projenin ana bulgusu negatif bir sonuç ise de, önemli bir hipotezin kapsamlı şekilde test edilmesini içerdiği için sinirbilim ve AH araştırmaları açısından, ayrıca lazerle yakalama işleminin tekil hücre genom analizinde kullanıldığı bir örnek olarak da genombilim metodolojisi açısından değerlidir. Çalışmanın yakın zamanda yayımlanması beklenmektedir.

Proje sayesinde proje ekibinde yer alan üç genç araştırmacı ve ayrıca bir lisans öğrencisi yetişmiştir. Yürütülen çalışmalara ve ekte sunulan rapora çeşitli biçimlerde katkıda bulunmuş olan araştırmacılar şöyledir: Sinan Can Açan, Thomas Arendt, Ulaş Işıldak, Poorya Parvizi, Mehmet Somel, Zeliha Gözde Turan, Uwe Ueberham, Etka Yapar, İdil Erte Yet.

Türkiye'de insan tekil hücre genom çalışmalarının ciddi anlamda hızlanmasına izin veren bu projeyi TÜBİTAK ERANET-RUS PLUS Programı çerçevesinde destekleyen Türkiye Bilimsel ve Teknolojik Araştırma Kurumu'na ve personeline, ayrıca dikkatli ve yapıcı yönlendirmeleri için hakemlere teşekkür etmek isteriz.

Doç. Dr. Mehmet SOMEL

Proje Yürütücüsü

İÇİNDEKİLER

ÖZET 1
ABSTRACT 2
1. GİRİŞ 3
1.1 Alzheimer Hastalığı 3
1.1.1 Alzheimer Hastalığı görülme sıklığı 3
1.1.2 Alzheimer Hastalığının bilinen sebepleri 4
1.1.3 Sağlıklı beyinde somatik anöploidi ve KSV 5
1.1.3.1 Tekil hücrelerde ikinci nesil dizileme yöntemiyle (scWGS) somatik anöploidi ve KSV analizi 6
1.1.3.2 scWGS yöntemiyle gerçekleştirilen somatik anöploidi ve KSV ölçüm çalışmalarının özeti 7
1.1.4 Alzheimer Hastalığında somatik anöploidi ve KSV 10
1.1.4.1 Alzheimer Hastalığında somatik anöploidi ve KSV ölçüm çalışmalarının özeti 11
1.1.4.2 AH'de somatik anöploidi ve KSV: Olası sebepleri ve önemi 12
2. GEREÇ VE YÖNTEM 14
2.1 Örneklem 14
2.2. Laboratuvar çalışmaları 14
2.2.1. Hücrelerin ayrıştırılması 14
2.2.1.1. Lazerle kesme (LCM) 14
2.2.1.2. Floresan ayırma (FACS) 15
2.2.2. Dizileme örneklerinin hazırlanması 15
2.3 Hesaplamalı iş hattının oluşturulması 17
2.3.1 Verilerin ön işlenmesi ve iş hattının optimizasyonu 17
2.3.1.1 Dizileme verisinin ön kalite kontrolü 17
2.3.1.2 Düşük dizilemenin etkisinin kontrolü 20
2.3.2. Üretilen verilerin analizi 22
2.3.2.1 Okumaların hazırlanması 22
2.3.2.2 Hizalama ve filtreleme işlemleri 22
2.3.2.3 HMMcopy ile kopya sayısı tahmini 23
2.3.2.4 Değişkenlik değeri filtresi 24
2.3.2.5 Dairesel ikili segmentasyon 25
2.3.2.6 KR-KSV tahminlerini kesiştirme ve ilgili segmentleri belirleme 25
2.3.2.6 İstatistiksel analizler 27
3. BULGULAR 29
3.1 Tekil hücre dizileme sonuçları ve deneysel yöntem etkisinin analizi 29
3.2 KSV ve anöploidi oranı tahminleri 31
3.2.1 Tüm hücreler arasında KSV ve anöploidi oranı 31
3.2.2 Kromozom başına KSV ve anöploidi sıklıkları 31
3.2.3 KSV'lerin uzunluk dağılımları 34
3.2.4 Biyolojik faktörlerin KSV ve anöploidi sıklığı üzerinde etkisi 35
3.2.4.1 Sıfır yığılmalı negatif binom regresyon modeli 35
3.2.4.1.1 Bütün KSV'leri açıklayan faktörler 36

3.2.4.1.1.1 Yalnızca kayıp sinyallerinin değerlendirilmesi 36
3.2.4.1.1.2 Yalnızca kazanım sinyallerinin değerlendirilmesi 36
3.2.4.1.1.3 Kayıp ve kazanım sinyallerinin birlikte değerlendirilmesi..... 36
3.2.4.1.2 Tüm kromozom anöploidi sinyallerini açıklayan faktörler 36
3.2.4.1.2.1. Yalnızca kayıp ve yalnızca kazanım sinyallerinin değerlendirilmesi..... 36
3.2.4.1.2.2 Kayıp ve kazanım sinyallerinin birlikte değerlendirilmesi..... 37
3.2.4.2 Parametrik olmayan testler..... 37
3.2.4.2.1 Bütün KSV'ler ile yapılan testler 37
3.2.4.2.1.1 Kayıp ve kazanım KSV olaylarının sıklıklarının karşılaştırılması..... 37
3.2.4.2.1.2 AH ve kontrol bireyler arasında KSV sıklıklığının karşılaştırılması..... 37
3.2.4.2.2 Tüm kromozom anöploidiler ile yapılan testler 38
3.2.4.2.2.1 Kayıp ve kazanım anöploidi olaylarının sıklıklarının karşılaştırılması..... 38
3.2.4.2.2.2 AH ve kontrol bireyler arasında anöploidi sıklıklarının karşılaştırılması 38
3.2.5 Teknik gürültünün anöploidi tahminleri üzerinde etkisi 39
3.2.5.1 Gürültü filtresi sonrası bağımlı değişkenlerin test edilmesi..... 40
3.2.5.1.1 Bütün KSV'ler ile yapılan testler 40
3.2.5.1.1.1 Kayıp ve kazanım KSV olaylarının sıklıklarının karşılaştırılması..... 40
3.2.5.1.1.2 AH ve kontrol bireyler arasında KSV sıklıklığının karşılaştırılması..... 41
4. TARTIŞMA VE SONUÇ..... 42
4.1 Bulguların özeti ve değerlendirmesi 42
4.1.1 Düşük dizileme etkisi..... 42
4.1.2 LCM ve FACS yöntemlerinin karşılaştırması..... 42
4.1.3 KSV ve anöploidi sıklıkları..... 43
4.1.4 KSV ve anöploidi sıklığı dağılımını etkileyen faktörler..... 44
4.2 Gelecek çalışmalar..... 44
KAYNAKÇA..... 46

ŞEKİL DİZİNİ

Şekil 1. Metilen mavi ile boyanmış insan beyin kesitlerinden PAML ile yapılan örnekleme. 15
Şekil 2: Okumaların uzunlukları boyunca kalite değerleri. 19
Şekil 3. Proje kapsamında dizilenen tekil sinir hücrelerinin dizileme derinlikleri..... 20
Şekil 4. McConnell vd. (2013) veri setinde düşük dizileme derinliği simülasyonu. 21
Şekil 5: McConnell vd. (2013) veri setinde düşük ve yüksek derinlikte verilerle anöploidi tahmini. 21
Şekil 6: Dizilenen 995 hücrede değişkenlik değeri (değişkenlik puanı) dağılımları. 30
Şekil 7: FACS ve LCM ile üretilmiş hücrelerin değişkenlik değerleri. 30
Şekil 8: FACS ve LCM ile üretilmiş hücrelerin dizileme derinliği değerleri. 31
Şekil 9: Üç farklı hücrede KSV ve anöploidi tahmini. 32
Şekil 10: AH ve kontrol örneklerinde kromozom başına tüm kromozom anöploidi olaylarının sayısı..... 33
Şekil 11: AH ve kontrol örneklerinde kromozom başına KSV olaylarının sayısı. 34
Şekil 12: KSV uzunluklarının hastalık durumu ve kayıp / kazanım sinyalleri açısından incelenmesi 35
Şekil 13: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen KSV olayı sıklığı farkları. 38
Şekil 14: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen anöploidi olayı sıklığı farkları..... 39
Şekil 15: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen KSV olayı sıklığı farkları (DD<0.45) 41

TABLO DİZİNİ

Tablo 1: Çalışmada kullanılan bireylerin ve örneklerin özellikleri. 16
Tablo 2: Proje dahilinde hücreleri dizilenen bireylerin cinsiyet dağılımları, kullanılan beyin bölgeleri kullanıldığı ve durumları (kontrol/ AH). 17
Tablo 3: Dizilenen kütüphanelerin ortalama (medyan) okuma sayıları 18
Tablo 4: En az bir KSV sinyali gösteren hücrelerin sıklığı. 32
Tablo 5: En az bir anöploidi sinyali gösteren hücrelerin sıklığı 33
Tablo 6: Değişkenlik değeri üst eşiği 0.45 olacak şekilde verinin yeniden filtrelenmesi sonrası en az bir KSV gösteren hücrelerin dağılımı 40
Tablo 7: Değişkenlik değeri üst eşiği 0.45 olacak şekilde verinin yeniden filtrelenmesi sonrası en az bir tüm kromozom anöploidi gösteren hücrelerin dağılımı 40

ÖZET

Alzheimer Hastalığı (AH) genetik, çevresel ve rastgele etmenlere bağlı olarak gelişen karmaşık ve ölümcül bir hastalıktır. AH'nin sebepleri ve gelişim süreci halen çok sınırlı ölçüde anlaşılmıştır.

Geçtiğimiz onyıllarda geleneksel sitometri ve FISH gibi yöntemler kullanan çalışmalar, AH vakaları ve sağlıklı kontrol bireyleri karşılaştırmış, AH vakalarının nöronlarında anlamlı ölçüde yüksek seviyede anöploidi ve megabaz büyüklüğünde kopya sayısı varyasyonu (KSV) bulmuştur. Ancak bu yöntemlerin teknik gürültüye açık olması sonuçların güvenirliğini sınırlamıştır.

Bu çalışmanın amacı da AH vakalarında somatik KSV yükünün rolü olabileceği hipotezini, tekil hücre tüm genom dizilemesi yöntemi kullanarak test etmektir. Ayrıca bilgimiz dahilinde ilk defa bu çalışmada genom dizileme işlemi için nöron çekirdeklerini toplarken, özgün hücrelerin seçilmesine izin veren lazerle kesme ve yakalama (LCM) yöntemi kullanılmıştır.

Çalışmamız kapsamında 10 AH vakası ile aynı yaş aralığında ve sağlıklı 7 kontrol bireyin temporal korteks ve hipokampüs örneklerinden elde edilen nöronal çekirdeklerden, iki farklı yöntemle toplam 1190 kütüphane oluşturulmuş ve düşük seviyede Illumina platformunda dizilenmiştir. Bunlar arasından 472 hücreye ait veriler kalite ve gürültü filtrelerini geçebilmiştir.

Analizimizin sonuçları, literatürde son yıllarda yayımlanan benzer verilerle de beraber değerlendirilerek, birkaç başlıkta özetlenebilir: Birincisi, megabaz uzunluğunda somatik KSV'ler insan beyin nöronları arasında %10 veya daha düşük orandadır. Tüm kromozom anöploidiler ise %2 veya daha düşük oranda görülmektedir. İkincisi, yakın zamanlı benzer bir çalışmayla paralel olarak, AH ve kontrol bireyleri arasında anlamlı bir fark bulunmamaktadır. Teknik açıdan ise, çekirdekleri ayrıştırma işleminde LCM, FACS'a göre anlamlı ölçüde daha gürültülü bir yöntemdir. Ayrıca teknik gürültü tekil genom analizinde de (muhtelemen dizileme derinliğinden bağımsız olarak) ciddi bir sorundur.

Tekil nöronların genom varyasyonunun ve somatik çeşitliliğin AH'de rolünün incelenmesini hedefleyen gelecek çalışmalar, gürültüyü minimize eden yöntemlerle makro değil mikro mutasyonları incelemeyi hedefleyebilir.

ABSTRACT

Alzheimer's Disease (AD) is a complex and deadly pathology that develops through the influence of genetic, environmental and stochastic factors. Its etiology is still only partly understood. Within the last decades, a number of studies using classical methods such as cytometry or FISH have compared AD patients and healthy controls, to find that AD patients' neurons carry a significantly higher proportion of aneuploidy and megabase-scale copy number variation (CNV). However the vulnerability of these methods to technical noise has limited their credibility.

The aim of this study has been to test the hypothesis that somatic CNV load may have a role in AD using single cell whole genome sequencing. In addition, to our knowledge for the first time in this study, the laser-capture microdissection (LCM) method has been used for isolating nuclei of specific neurons for genome sequencing.

We have generated a total of 1190 libraries from neuronal nuclei, derived from temporal cortex and hippocampus samples of 10 AD patients and 7 healthy age-matched control individuals, and performed low coverage Illumina sequencing. In total 472 cells data have passed our data quality and noise filters.

In conjunction with recent reports, our results can be summarised as follows: First, megabase-scale CNV's are most likely at 10% or lower frequency among human neurons. Whole chromosome aneuploidies are 2% or lower. Second, in line with a recent similar study, we cannot find any indication of higher CNV among AD patients' neurons than healthy control neurons. On the technical side, LCM appears to be a significantly more noisy method than FACS for isolation of nuclei. Further, technical noise haunts single cell genome analyses, most likely independent of sequencing depth.

Future studies aiming to investigate the possible roles of single cell genomic variation and somatic diversity in AD may prefer to use methods that particularly minimise noise and to focus on micro, instead of macro mutations.

1. GİRİŞ

1.1 Alzheimer Hastalığı

Alzheimer Hastalığı (AH) dünyada en sık görülen nörodejeneratif hastalıklardan biridir ve demansın en yaygın sebebidir (Bishop vd., 2010; Ballard vd., 2011). AH genellikle ileri yaşta ortaya çıkar ve önce hafıza kaybı, konuşma ve bilişsel aktivitelerde zayıflama, ileri safhalarında ise günlük hayatı idame ettirememe şeklinde kendini gösterir (Qiu vd., 2009). AH'nin şu anda tedavisi yoktur ve yalnızca semptomları hafifletici müdahaleler mevcuttur; günümüzde AH teşhisinden itibaren ortalama 10 yıl içinde ölümle sonuçlanmaktadır (Masters vd., 2015).

1.1.1 Alzheimer Hastalığı görülme sıklığı

Hastalığın görülme sıklığı yaşa bağlıdır ve 65 yaşından sonra AH sıklığı **her 5 yılda bir ikiye katlanmaktadır** (Qiu vd., 2009). 2005 yılına ait bir çalışmanın verilerine göre AH'nin dünya genelinde **60 yaş üzeri nüfusta sıklığı %3.9'dur**; gelişmiş ülkelerde ise bu genel ortalamanın üzerinde oranlarda görülmektedir (Ballard vd., 2011). Gelişmiş ülkelerde oranın yüksek olmasının sebebinin, ölüm sebeplerinde farklılaşmanın yanısıra, daha az gelişmiş ülkelerde AH'nın daha seyrek teşhisi olabileceği tahmin edilmektedir.

Türkiye'de AH sıklığı konusunda bilgimiz dahilinde şimdiye kadar yerel çalışmalar yürütülmüştür. Eskişehir'de yapılan bir çalışma 55 yaş üzeri nüfusta AH sıklığını %4.1 olarak tespit etmiştir (Arslantaş vd., 2009). İstanbul'da bir çalışma ise **70 yaş üzerinde AH sıklığını %11.0** olarak belirlemiştir (Gurvit vd., 2008). Hastalık sıklığının 5 yıllık müteakip yaş dilimleri arasında iki katı arttığı hesaba katılırsa, bu veriler, Türkiye'de AH'nin **dünya ortalamasına benzer oranlarda** ortaya çıktığını göstermektedir.

Önümüzdeki onyıllarda hem Türkiye'de hem dünya çapında yaşam süresinin uzaması ve nüfus yaşlanmasının artması öngörülmektedir. Bu nedenle demans ve özel olarak AH'nin sıklığının artacağı, toplumsal ve maddi maliyetlerinin katlanacağı beklenmektedir. Örneğin 2040 yılında 60 yaş üzerinde AH hastalarının sayısının 2000 yılına kıyasla dünya genelinde yaklaşık dört kat, Ortadoğu bölgesinde ise yaklaşık **beş kat** artacağı tahmin edilmektedir (Ballard vd., 2011). Bu rakamlar ışığında Türkiye'de de AH'nin **giderek daha ağır bir sorun** olacağı aşikardır. Bu durum da AH'nin sebeplerini tespit ederek etkili tedavi yöntemleri geliştirmenin önemini artırmaktadır. Ancak ne yazık ki AH son derece karmaşık bir hastalıktır ve sebepleri ile gelişim süreci halen çok sınırlı düzeyde anlaşılmıştır.

1.1.2 Alzheimer Hastalığının bilinen sebepleri

Ender durumlarda (%1) AH nadir ve güçlü genetik faktörler nedeniyle erken yaşta ortaya çıkar, ancak AH vakalarının %95'ten fazlası çok sayıda genetik, çevresel ve rastgele etmenin bir araya gelmesiyle gelişen ve **geç yaşta ortaya çıkan sporadik AH** vakalarıdır (Masters vd., 2015).

AH gelişiminin gerisinde yatan fizyolojik değişimler, merkezi sinir sistemine ait sinir hücrelerinde moleküler ve fizyolojik stres birikimi sonucu yaşanan sinaps kayıpları ve hücre ölümleridir (Ballard vd., 2011; Masters vd., 2015).

AH'nin moleküler düzeyde başlıca iki kaynağı tespit edilmiştir:

(1) Çözülemeyen **amiloid beta (Aβ) plaklarının** beyinde sinir hücreleri arasındaki bölgelerde birikimi,

(2) Mikrotübül proteini **tau'nun fosforile olmuş tiplerinden oluşan nörofibril düğümlerin** sinir hücreleri içinde birikimi (Ballard vd., 2011; Masters vd., 2015).

Aβ plakların APP proteininin yanlış kesilmesi nedeniyle oluşup biriktiği düşünülmektedir. Aβ plaklarının AH beyinlerinde aşırı birikimi, erken AH vakalarına sebep olan mutasyonların Aβ plaklarının oluşumundan sorumlu ***APP*** geni ve başka ilgili genlerde yer aldıklarının tespit edilmesi, ayrıca ileri yaşta AH'ye karşı koruma sağlayan alellerin yine *APP* geni içinde bulunması (Jonsson vd., 2012), **Aβ plakların hastalıkta temel rol oynadığı** hipotezinin temelidir. Aβ plak ve nörofibril düğüm birikiminin beyinde genellikle temporal ve frontal kortekslerde başlayıp beynin diğer bölgelerine zaman içinde yayıldığı gözlemlenmiştir (Masters vd., 2015).

Aβ plakların bilişsel aktivite yitiminde tetikleyici olabileceği fikrinden yola çıkarak, **monoklonal anti-Aβ antikorlar** yoluyla bu plakları beyinden temizlemeyi amaçlayan bir dizi potansiyel tedavi yöntemi hem hayvan modellerinde hem de klinik araştırmalarda incelenmiştir (van Dyck 2018). Ancak şu ana kadar plakları temizlemeye dönük bu girişimler ya tamamen etkisiz olmuş (örneğin hayvanlarda nörodejenerasyonu engelleyememiş) ya da çok sınırlı bir iyileşmeyle sınırlı kalmış, bu durum da AH'de **Aβ plakların rolünün sorgulanmasına** yol açmıştır (Herrup 2015). Aynı şekilde, AH teşhisi konmamış bazı bireylerin postmortem beyin incelemelerinde yoğun Aβ plak birikimi görülmesi, Aβ plakların tek başlarına AH'yi açıklamadığı fikrini desteklemektedir (Herrup 2015). AH'nin gerisinde yatan mekanizmaların karmaşıklığına işaret eden çevreler, Aβ plak ve nörofibril düğüm birikiminin yanısıra artan **mitokondriyel stres** ve **kalsiyum dengesizliği** gibi etmenlerin de sinaps kaybı ve hücre ölümü sonuçlarına katkıda bulunduğunu vurgulamaktadır (Herrup 2015; Masters vd., 2015).

Öte yandan tüm bu değişimlerin genetik ve çevresel kaynakları büyük ölçüde belirsizliğini korumaktadır. Geç yaşta ortaya çıkan AH vakalarının **kalıtsallığı** İsveç'te yürütülen bir ikiz çalışmasında %79 olarak tahmin edilmiştir (Barber 2012), ancak ilgili genetik arkaplana dair şu ana kadar bilgimiz sınırlıdır. Geçtiğimiz yıllarda yürütülen **genom çapında ilişkilenme araştırmaları** (GWAS) AH sıklığını açıklayan bazı yaygın genetik polimorfizmler tespit edebilmiştir; ancak bunlar iki düzineyi geçmemektedir. Örneğin 2013 yılında yürütülen ve 74 bini aşkın bireyi içeren bir meta-analiz yalnızca **20 lokus** tespit edilmiştir. Bunlardan başlıcaları AH riskini 3 ila 10 kat artıran ve Aβ transferinde rol alan *APOE* geninin alelleri (ör. *APOE4*) iken, diğer lokuslardaki polimorfizmlerin etkisi genelde riski %10 civarında artırmakla sınırlıdır (Lambert vd., 2013). Toplamda varyasyonun ancak çok düşük bir miktarı yaygın polimorfizmle açıklanabilmektedir (Barber 2012) ve eldeki modeller, neden bazı bireylerde ileri yaşta AH geliştiğini yeterince açıklayamamaktadır.

Bu sonuçları değerlendiren araştırmacılar, geç yaşta ortaya çıkan sporadik AH'nin tek boyutlu ve tek etmenli bir hastalık olmadığını, yaşlanma sırasında yaşanan doğal hasar birikimi ve ortaya çıkan genetik ve fizyolojik sorunların kombinasyonunun özel bir dışa vurumu olduğunu vurgulamaktadır. AH gelişiminde nadir genetik varyantların ya da **gen-gen veya gen-çevre etkileşiminin**, özellikle de **somatik hasar ve mutasyon birikiminin** rol alması mümkündür.

1.1.3 Sağlıklı beyinde somatik anöploidi ve KSV

Hem AH hem de normal yaşlanmada hücre içi hasar ve mutasyon birikiminin rolü olduğu bir süredir kabul edilmektedir (Lopez-Otin vd., 2013). Bu tip hasar ve mutasyonun bir sınıfı da **anöploidi**, yani fazla veya eksik kromozomlar, ayrıca megabaz düzeyinde **kopya sayısı varyasyonu (KSV)**, yani büyük delesyon (silinme/kayıp) ve duplikasyon (çiftlenme/kazanım) vakalarıdır. Özellikle de insan merkezi sinir sisteminde nöronlarda, erken gelişim sırasında ya da sonrasında ortaya çıkan somatik mutasyonlar nedeniyle düşük seviyede ve hücreye özgü **anöploidi ve KSV** gözlemlenebildiği bugün yaygın olarak kabul görmektedir (Rosenkrantz ve Carbone 2017; Adriani vd., 2017). Ancak bu somatik KSV yükünün seviyesi, sebepleri ve olası sonuçları halen çok az anlaşılabilmiştir.

2000-2012 yılları arasında klasik kromozom sayma yöntemleriyle yürütülen toplam dokuz karyotipleme çalışmasının sonuçları Faggioli vd., (2012) tarafından derlenmiştir: Bu çalışmalar sağlıklı bireylerin korteks sinir hücrelerinde anöploidi sıklığını %3 ila %30 arasında bir aralıkta tahmin etmiştir. Bu çalışmalarda **lam temelli sitometri**, kromozoma özgü ve interfaz kromozomları işaretleyebilen **floresan in situ hibridizasyon (FISH)**, ayrıca **spektral karyotipleme (SKY)** gibi yöntemler kullanılmıştır (ör. Rehen vd., 2005; Iurov vd.,

2009; Fischer vd., 2012). Ancak bu yöntemlerin kısıtları, örneğin FISH'de sınırlı sayıda kromozomun incelenebilmesi ya da sitometride yüksek gürültü seviyesi, bu alandaki araştırmacıları farklı yöntem arayışına sevketmiştir.

1.1.3.1 Tekil hücrelerde ikinci nesil dizileme yöntemiyle (scWGS) somatik anöploidi ve KSV analizi

Son on yıl içinde farklı bağlamlarda anöploidi ve KSV ikinci (yeni) nesil dizileme yöntemleriyle incelenmeye başlamıştır (Baslan vd., 2012; Garvin vd., 2015). Aynı şekilde bir dizi çalışma, beyin hücrelerinde anöploidi ve KSV sıklığını **tekil hücrelerde tüm genom dizileme** (İng. "single cell whole genome sequencing", **scWGS**) kullanılarak incelenmeye girişmiştir. Bu scWGS çalışmaları günümüzde özetle şu şekilde yürümektedir:

- Doku örneğindeki hücreler birbirinden ayrıştırılır; bunun için Floresan Aktive Hücre Ayırma (İng. "Fluorescence-activated cell sorting", FACS) ya da lazerle yakalama ve mikrodiseksiyon ("laser capture microdissection", LCM) gibi bir yöntem, ayrıca örneğin sinir hücrelerinin seçilmesi için NeuN belirteci gibi belirteçler kullanılır;
- Her bir hücrenin tüm genomik DNA'sı saflaştırılır;
- DNA, Sigma'nın GenomePlex/WGA kitinde olduğu gibi rastgele kısa oligonükleotidleri primer olarak kullanan bir polimeraz zincir tepkimesi (PCR, İng. "polymerase chain reaction") veya "multiple displacement amplification" (MDA) gibi bir yöntemle çoğaltılır;
- Yükseltgenen bu DNA ile yeni nesil dizileme (genelde Illumina) kütüphanesi hazırlanır;
- Birden fazla kütüphane aynı şeritte birleştirilerek düşük seviyede (genelde 0.5x ila 0.01x arası) dizileme yapılır;
- Üretilen veriler referans genoma hizalanır ve genelde 500 kb uzunluğunda pencereler içinde okuma yoğunluğu hesaplanır;
- Hücre başına bir gürültü katsayısı (örneğin her bir kromozom içindeki okuma yoğunluğunun standart sapması) hesaplanır ve önceden belirlenen gürültü eşiğini aşan hücreler filtrelenir;
- Okuma yoğunluğu verileri kullanılarak, her bir hücrede kopya sayısı ortalamanın üstünde veya altında olan kromozomlar (anöploidi) ya da kromozom parçaları (KSV), KSV tespiti için geliştirilmiş (daha sonra ayrıntılandıracağımız) algoritmalar yardımıyla tahmin edilir (Rosenkrantz ve Carbone 2017).

Hücreler arası DNA kopya sayısı varyasyonunun incelenmesi çekici bir analiz olsa da ciddi zorluklar da taşımaktadır. Özellikle de çok az miktarda DNA'nın yükseltgenmesi sırasında ortaya çıkan teknik gürültü, büyük kromozom segmentlerinin silinmesi gibi yapay sonuçlar doğurabilmektedir (Baslan vd., 2012; Garvin vd., 2015). Ayrıca yüksek kapsamda genom dizileme verilerinde KSV tespiti için geliştirilmiş olan hesaplamalı araçlar bu kadar düşük kapsamdaki verilere doğrudan uygulanamadığı için özgün iş hatları ve algoritmalar geliştirme zorunluluğu doğmuştur.

1.1.3.2 scWGS yöntemiyle gerçekleştirilen somatik anöploidi ve KSV ölçüm çalışmalarının özeti

Son beş yıl içinde yayımlanan scWGS çalışmaları, beyinde anöploidi (tüm kromozom kaybı ya da kazanımı) oranını %1 ila %10 arasında, KSV (megabaz düzeyindeki kayıp ya da kazanımları) oranını ise %10 ila %70 arasında geniş bir yelpazede rapor etmiştir. Aşağıda bu çalışmaları kısaca özetliyoruz:

- 2013 yılında McConnell ve arkadaşları 3 genç bireyin frontal cerebral korteksinden elde ettikleri toplam 110 hücreyi (birey başı ~30 hücre, ~0.1x derinlik) dizilememiş, bu **hücrelerin %2.7'sinde tüm kromozom anöploidi** bulmuştur. Ayrıca hücrelerde **%41 oranında en az bir KSV vakası** (çoğu 3 ila 75 megabaz ölçeğinde) tespit etmişlerdir. KSV'lerin çoğu telomerlere yakın çıkmıştır. Kayıplar (silinmeler) kazanımlardan iki kat daha sıktır (McConnell vd., 2013). Bu çalışmada GenomePlex ve hücre başına ~0.1x dizileme kullanılmıştır.
- Cai ve arkadaşları 3 erişkin bireyin korteksinden elde ettikleri toplam 97 hücreyi (birey başı ~30 hücre, ~0.1x derinlik) dizilememiş, bu hücrelerin **hiçbirinde tüm kromozom anöploidi bulamamıştır**. Ancak yazarlar, daha ayrıntılı inceledikleri bir bireye ait 19 sinir hücresinden **%68'inde KSV vakaları** (hücre başına 1.7-17 Mb büyüklüğünde ortalama 3.4 vaka) tespit etmişlerdir (Cai vd., 2014). Kayıpların kazanımlardan çok olduğu görülmüştür. Ayrıca iki hücrede aynı KSV'nin görülmesi, bu mutasyonların erken gelişim kökenli olabileceği fikrini desteklemiştir. Bu çalışmada GenomePlex ve hücre başına ~0.1x dizileme kullanılmıştır. Yazarlar GenomePlex'in MDA'dan daha az gürültülü olduğunu karşılaştırmalı olarak göstermiştir.
- 2014 yılında Knouse ve arkadaşları farklı örneklerde scWGS kullanarak anöploidi incelemiştir. Grup, öncelikle yöntemleri ve algoritmalarının anöploidiyi yakalamak için etkili olduğu test etmek için trizomik farelerde tekil hücre dizilemesi yapmış ve beklenen trizomiyi tespit edebilmiştir (Knouse vd., 2014). Ardından 4 erişkin bireyin

frontal lobundan elde edilen toplam 89 hücreyi (birey başı ~20 hücre, ~0.1x derinlik) dizilemiştir. Bu hücrelerin **%2.2'sinde tüm kromozom anöploidi** rapor etmişlerdir. Çalışmanın yazarları kromozom altı vakaları rapor etmemiştir. Yazarlar erişkin fare beyninde de 19 hücre dizilemiş ancak anöploidi bulamamışlardır. Çalışmanın yazarları ayrıca NeuN ile hücre işaretlemenin DNA örneği hazırlamayı zorlaştırarak gürültü yarattığı sonucuna ulaşmışlardır. Yazarlar, **fazla gürültülü** buldukları (hücre içinde pencereler arasında varyansın yüksek olduğu) yaklaşık %10 hücreyi analizden çıkardıklarını not etmişlerdir. Bu çalışmada GenomePlex ve hücre başına ~0.1x dizileme kullanılmıştır.

- Knouse ve arkadaşları 2016'da yayınladıkları çalışmada ise dikkatlice tasarlanmış pozitif ve negatif kontroller yoluyla kullandıkları yöntem ve algoritmaların hassasiyetini ve spesifisitesini (özgüllüğünü) ölçmüş, beş megabaz altı vakaların genellikle doğru tahmin edilemeyeceği sonucuna ulaşmışlardır (Knouse vd., 2016). Buradan yola çıkarak 0.1x dizileme ile yine aynı 4 erişkin bireye ait 105 beyin hücresinde (birey başına ~25 hücre) anöploidi ve KSV incelemiştir. Hücrelerde **yalnızca bir tane tüm kromozom anöploidi (<%1) vakası** rapor edilmiştir. Ayrıca hücrelerin **%9'unda KSV vakaları** (≥3 Mb) tespit edilmiştir. Bu KSV'lerin %90'ı kayıp vakalarıdır ve telomerlerde yoğunlaşmışlardır. Bu çalışmanın yazarları daha önceki scWGS çalışmalarında (sırasıyla McConnell vd., 2013 ve Cai vd., 2014) rapor edilen %40 ve %60'ı aşkın KSV oranlarının abartılı olduğunu iddia etmiş, daha özgül olduğuna inandıkları kendi algoritmaları ve gürültü filtreleri ile aynı veri analiz edildiğinde KSV oranının sırasıyla %17 ve %10 olduğunu belirtmişlerdir. Yazarlar ayrıca dizilenen 55 keratinositle karşılaştırıldığında, beyin hücrelerinde gürültünün çok daha yüksek olduğunu bulmuştur. Yazarlar hücre tipleri arasındaki bu farkın kromatin yapısı farklarıyla ilgili olabileceğini tahmin etmiştir. Bu çalışmada GenomePlex ve hücre başına ~0.1x dizileme kullanılmıştır.
- 2016 yılında van den Bos ve arkadaşları ise 6 kontrol bireyin frontal serebral korteksinden alınan toplam 589 hücreyi dizilemiş (birey başına ~100 hücre), hücrelerin yalnızca **%0.7'sinde tüm kromozom anöploidi** rapor etmiştir (van den Bos vd., 2016). Kromozom altı KSV vakaları rapor edilmemiştir. Çalışmanın yazarları, yöntem ve algoritmalarının verimliliğini kanıtlamak için Down Sendromlu bir bireyin 36 beyin hücresini aynı şekilde incelemiş, tümünde trizomi 21 kromozom bulduklarını belirtmiştir. Yazarlar, kütüphane hazırlığı öncesinde PCR yürütmediklerini (GenomePlex kullanmadıklarını) ve bu yaklaşımlarının gürültüyü azaltabileceğini belirtmiştir (ancak bu konuda karşılaştırmalı veri sunulmamıştır). Yazarlar ayrıca,

diziledikleri hücrelerin yaklaşık %40'ını düşük kalite sebebiyle analizden çıkardıklarını not etmiştir.

- 2017 yılında ise Vitak ve arkadaşları çok sayıda hücreyi paralel olarak ve özgün biçimde işaretleyerek düşük seviyede dizileme ile anöploidi tahmini yapmayı hedefleyen yeni bir deneysel yöntem (SCI-seq) rapor etmiştir (Vitak vd., 2017). Genel olarak yeni yöntemin klasik dizilemeden daha gürültülü olduğu görülmektedir, ancak görece ucuz olması ve hücre ayırma işlemi gerektirmemesi nedeniyle tercihi mümkündür. Bu yazarlar da iki erişkin makak maymununun frontal korteksinden sinir hücreleri dizileyerek toplam 305 hücrede **%10 civarında tüm kromozom anöploidi** tespit etmiş, KSV tahmini ise yapmamıştır. Ancak yazarlar standart scWGS yöntemleriyle aynı bireylere ait hücreler işlendiğinde, bu oranın %3'e kadar düştüğünü de rapor etmiştir.
- 2018 yılında Rohrback ve arkadaşları fare beyin gelişimi sırasında anöploidi ve KSV sıklığında değişimleri incelemiştir (Rohrback vd., 2018). Bu çalışmanın yazarları GenomePlex yerine tagmantasyon denilen yöntemi, genom çapında daha homojen dizileme sağladığı gerekçesiyle tercih etmişlerdir (ancak bu hususta karşılaştırmalı veri sunulmamıştır). Yazarlar, yöntemlerinin gerçek KSV'lere karşı hassasiyetinin göstergesi olarak somatik V(D)J rekombinasyonu geçirmiş lemfositlerde, bu lokusta 0.25 Mb büyüklüğünde KSV'leri gerçekçi biçimde tahmin edebildiğini rapor etmiştir. Ayrıca yazarlar, erkeklere ait sinir hücrelerinde gürültü seviyesinin daha yüksek seviyede olduğunu not etmiş, bunun X monozomisi nedeniyle ortaya çıktığını ve dolayısıyla gürültü filtrelerinin gerçek vakaların elenmesine yol açabileceğini not etmiştir. Bu çalışmada incelenen bir erişkin bireye ait 55 sinir hücresinde **%2 oranında tüm kromozom anöploidi** rapor edilmiştir (kromozomun %50'sinden fazlasının etkilendiği vakalar yazarlarca anöploidi olarak sınıflandırılmıştır). Aynı zamanda hücrelerin >%80'inde KSV rapor edilmiştir (çoğu 5 Mb altında hücre başına ~9 vaka). Başka çalışmalardaki sonuçlara benzer şekilde, kayıpların kazanımlardan yaklaşık 5 defa daha sık olduğu belirtilmiştir.

Görüldüğü üzere bu tip analizlerde halen standart bir deneysel ve hesaplamalı iş hattı geliştirilememiştir. Çalışmaların vardıkları sonuçlar arasındaki **tutarsızlıkların çalışmalar arasındaki teknik farklardan** kaynaklanıyor olması muhtemeldir. FISH ve sitometri gibi yöntemlerin gürültüye karşı hassasiyeti çokça vurgulansa da (van den Bos vd., 2016; Knouse vd., 2016), scWGS çalışmaları arasında da yüksek tutarlılık olmadığı görülmektedir. Burada sorun yalnızca gürültülü verilerin filtrelenmesiyle ve sahte pozitiflerin elenmesiyle de çözülememektedir, çünkü filtrelemeler sahte pozitif oranını düşürürken sahte negatif (yani tespit edilemeyen) vakaların oranını artırmaktadır. Kullanılan yöntemlerde karşılaşılan ciddi

teknik gürültü (PCR'den kaynaklı sapmalar ve/veya hücreler arası kontaminasyon) ve hesaplamalı yöntemlerin gürültüye karşı hassasiyeti, aşılması gereken sorunlar olarak durmaktadır.

Bu belirsizliklere rağmen yukarıda özetlediğimiz yedi dizileme deneyinin sonuçları şöyle derlenebilir:

- Genel olarak sağlıklı erişkin beyinlerinin nöronlarında somatik **tüm kromozom anöploidiler** çok seyrek, **muhtemelen %5'in altında**, belki de %1'in altında sıklıkta ortaya çıkmaktadır.
- Megabaz boylu somatik **KSV'ler** ise hücrelerin **en az %10'unda**, belki de daha sık görülmektedir.
- Somatik KSV'ler arasında **kayıplar** kazanımlardan daha sık olabilir.
- Somatik KSV'ler **kromozom uçlarına doğru** daha sık ortaya çıkmaktadır.

Bu sonuçlar ışığında beyin dokusunun genetik açıdan **mozaik bir karakter** sergilediği söylenebilir. Bu mozaisizme yol açan başlıca unsurun **embriyonik gelişim sırasında yaşanan kromozom ayrılma sorunları** (ör. meritolik yapışmalar; kromozomların iki taraftan gelen mikrotübüllere yapışması) olduğu düşünülmektedir (Faggioli vd., 2012; Adriani vd., 2017). Nitekim fare beyninde hücreye özel KSV'lerin embriyonik gelişim sırasında yoğunlaştığı, ancak sonradan sayılarının azaldığı, yani bir tür **negatif seçilime** uğradıkları, tekil hücre dizileme yöntemiyle gösterilmiştir (Rohrback vd., 2018). Bu çalışmada hücre başına KSV oranının farenin geç embriyonik döneminde (E14.5) zirve yaparak erişkin seviyesinin iki katına ulaştığı, doğuma doğru azaldığı ve doğum ile erişkinlik arasında sabit kaldığı ortaya konmuştur. Negatif seçilim modeli doğruysa erişkinlerde gözlemlenen anöploidi ve KSV vakaları yaşlanma sırasında yaşanan sorunlardan değil, **nörojenez** sırasında ortaya çıkan ve negatif seçilime uğramamış (hafif etkili) KSV'lerden kaynaklanıyor olabilir. Ayrıca insanda, anöpoidik hücrelerin sayısının erişkin yaşam süresi içinde azalma eğiliminde olduğu, serebral korteks sinir hücrelerinde sitometri yoluyla gösterilmiştir (Fischer vd., 2012). Bu sonuç da anöploidi ve KSV'lerin erken dönemde ortaya çıkıp daha sonra eleniyor olması fikriyle tutarlıdır.

1.1.4 Alzheimer Hastalığında somatik anöploidi ve KSV

Bir dizi farklı bulgu somatik mutasyonların, özel olarak da somatik anöploidi ve KSV'lerin sporadik AH'de rolü olabileceğine işaret etmektedir (Adriani vd., 2017). Bunlardan biri, **Down Sendromlu** bireylerde AH'de görülen tipte plak ve düğümlerin sıklıkla gözlemlenmesidir

(Mann 1988). ***APP* geninin 21. kromozomda bulunması** bu durumun olası sebeplerinden en az biridir ve dengesiz gen ifadesinin hastalığın gelişimine katkıda bulunabileceğine işaret etmektedir. Daha yakın zamanlı çalışmalar da sporadik AH bireylerin sinir hücrelerinde ***APP* geninin onlarca anormal rekombinant formunu** bulmuşlar, bunların *APP* mRNA'sının anormal uçbirleştirme sonrası DNA'ya geri yazılması sonucu genomda biriktiğini, ilerleyen yaşla bu anormal ve toksik *APP* formlarının çoğaldığını da tespit etmişlerdir (Bushman vd., 2015; Lee vd., 2018).

1.1.4.1 Alzheimer Hastalığında somatik anöploidi ve KSV ölçüm çalışmalarının özeti

AH ve KSV arasında ilişki bulunduğuna dair bir ikinci gözlem ise doğrudan karyotip analizi ve sitometriden gelmektedir. Kromozoma spesifik ve interfaz kromozomları işaretleyebilen FISH yöntemiyle **21. kromozom** ve **X kromozomunun** sıklığını inceleyen çalışmalar, Alzheimer hastalarında, aynı yaştaki kontrollere kıyasla bazen iki kat daha fazla anöploidi olduğunu rapor etmiştir (Iourov vd., 2009; Yurov vd., 2014). Buna ek olarak, çok sayıda hücrede kromozom sayısı dağılımını basit ve hızlı bir şekilde kestirmeye yarayan sitometri çalışmalarında, aşırı anöploid (örneğin tetraploid) hücrelerin aynı yaştaki kontrollere göre AH'de daha fazla olduğunu, ayrıca hücre başına DNA miktarında artışın entorinal korteks gibi AH'den yoğun etkilenen bölgelerde yoğunlaştığı ama örneğin oksipital kortekste görülmediği rapor edilmiştir (Mosch vd., 2007). Beriki çalışmayı yapan grubun bir başka çalışmasında ise **AH'nin ileri aşamalarında anöploidi taşıyan nöronların azaldığı** bulunmuş, bu azalma söz konusu hücrelerin apoptoza girmesiyle açıklanmıştır (Arendt vd., 2010).

Bilgimiz dahilinde, bu soruyu tekil hücre tüm genom dizileme (scWGS) kullanarak inceleyen yalnızca bir çalışma yayımlanmıştır. van den Bos ve arkadaşlarının yukarıda (1.1.2.2.) özetlediğimiz çalışmasında 6 sağlıklı kontrol bireye ek olarak 10 Alzheimer hastasının beyinlerinden toplam 1500'e yakın sinir hücresi dizilenmiştir. Sağlıklı bireylerde tespit edilen %0.7 sıklığındaki tüm kromozom anöploidiye benzer biçimde, AH bireylerin sinir hücrelerinde de %0.6 sıklığında anöploidi bulunmuştur (van den Bos vd., 2016). Kısacası FISH ve sitometri çalışmalarının aksine, sağlıklı bireylere kıyasla **AH'de anöploidi eğilimi tespit edilememiştir**. Ayrıca FISH ile daha önce görülmüş olan kromozom 21 kazanımı da bulunamamıştır. Bu çalışmanın yazarları, kendi sonuçları ile daha önce yayımlananlar arasındaki tutarsızlığı, FISH ve sitometrinin slayt hazırlama sürecindeki **teknik varyasyon** nedeniyle gürültülü olmasıyla açıklamışlar, ancak sonuçlarının daha geniş örneklemlerle tekrarının gerekli olduğunu da belirtmişlerdir.

1.1.4.2 AH'de somatik anöploidi ve KSV: Olası sebepleri ve önemi

Özetlemek gerekirse, Alzheimer hastaları ve sağlıklı bireylerin sinir hücreleri arasında anöploidi ve KSV sıklığı farkı olup olmadığı şu anda belirsizliğini korumaktadır. Sağlıklı erişkin bireylerde anöploidi ve KSV sıklığı konusundaki belirsizlik gibi, AH vakaları ve kontroller arasındaki farka dair belirsizlik de en azından kısmen deneysel standardizasyon eksikliğinden kaynaklanmaktadır.

Olası bir farkın sebepleri de şimdilik yalnızca spekülasyon konusudur. Bir olası açıklama, AH'de yüksek KSV seviyesinin aslında embriyonik dönemde nörogenez sırasında ortaya çıkmasıdır. Bu durumda bazı bireylerde aşırı somatik mutasyon yükü ileri yaşta kendisini göstererek AH riskini artıyor olabilir. Alternatif olarak, anöploididekı bu artışı AH vakalarında sinir hücrelerinin özelleşmişliklerini kaybederek yeniden hücre döngüsüne girmeleri eğilimine bağlayan araştırmacılar da bulunmaktadır (Arendt vd., 2000; Mosch vd., 2007). Bu ikinci modele göre anöploidi ve KSV yalnızca bir sonuçtur.

AH'de anöploidi ve KSV sıklığının rolünün ve olası sebebinin doğru anlaşılması, AH'ye karşı alınabilecek etkili önlemlerin belirlenmesi açısından gereklidir. Birincisi, bir anöploidi/KSV etkisinin olduğunu ve bunun da çevresel stres veya stokastik etmenler sebebiyle **embriyonik dönemde ortaya çıkan somatik mutasyon yükünden** kaynaklandığını varsayacak olursak, bu bilgi hem risk grubundaki bireylerin erken tespitinin, hem anöploidi/KSV'lerin hücre üzerindeki uzun vadeli etkilerinin incelenmesinin, hem de nörogenez sırasında mutasyon yükünü düşürebilecek önlemlerin araştırılmasının önünü açacaktır.

İkincisi, bir anöploidi/KSV etkisi varsa ve bu olgu, strese giren erişkin **sinir hücrelerinin anormal biçimde hücre döngüsüne girmesinden** kaynaklanıyorsa (Mosch vd., 2007), bu bilgi sayesinde anöploidi/KSV'lerin hücre üzerindeki kısa vadeli etkilerinin incelenmesi ve olası etki noktalarına karşı müdahaleler araştırılması söz konusu olabilecektir.

Bir üçüncü olasılık ise van den Bos ve arkadaşlarının (2016) sonuçlarının işaret ettiği gibi **anöploidi ve KSV sıklığının AH'de rolü olmadığıdır**. Toplanacak yeni verilerle bu ihtimal güçlenecek olursa, hastalığı önleme çalışmalarının odağı başka yönlere (örneğin anormal APP izoformlarının birikimi ve ifadesinin engellenmesi veya proteostasisin korunması gibi alanlara) kaydırılabilecektir.

Almanya, Fransa, Rusya ve Türkiye'den dört araştırma grubunun katıldığı bu geniş kapsamlı proje de söz konusu sorunu daha yakından araştırmayı hedeflemektedir. Bu kapsamda Almanya ve Rusya'daki ekipler sağlıklı ve AH bireylerde anöploidi/KSV'lerin sıklığını FISH ve sitometri yöntemleriyle araştırırken, grubumuz da Almanya'daki ekiple beraber aynı konuyu

scWGS ile araştırmayı planlamıştır. Bu çalışma kapsamında şu temel soruların cevaplanması hedeflenmektedir:

- Sağlıklı erişkin ve AH bireylerin beyinlerinde tüm kromozom anöploidilerin ve KSV'lerin sıklıkları fark göstermekte midir?
- Anöploidi/KSV vakaları belli kromozomlarda yoğunlaşmakta mıdır?
- Kazanım ve kayıpların sıklıkları arasında fark var mıdır?
- Anöploidi/KSV sıklığı beyin bölgeleri arasında fark göstermekte midir?
- LCM (lazerle kesme) ve FACS (floresanla ayırma) yöntemleri arasında deneysel gürültü açısından fark var mıdır?

2. GEREÇ VE YÖNTEM

2.1 Örneklem

Bu çalışmada Fransa'nın Paris kentinde bulunan Fransa Ulusal Beyin Bankası'nda depolanmış postmortem beyin örnekleri kullanılmıştır. Bunlar donmuş temporal korteks ve/veya hipokampüs örnekleridir. Çalışmada şu ana kadar 17 bireyden alınmış **1190 örneğe ait kütüphane** dizilenmiştir. Bireyler ile ilgili detaylı bilgi **Tablo 1**'de gösterilmiştir. **Tablo 2**'de ilgili bireylerin hastalık durumları ve incelenen beyin bölgelerine dair bir özet sunulmaktadır. Kontrol bireyler AH bireylerle aynı yaş aralığından seçilmiştir.

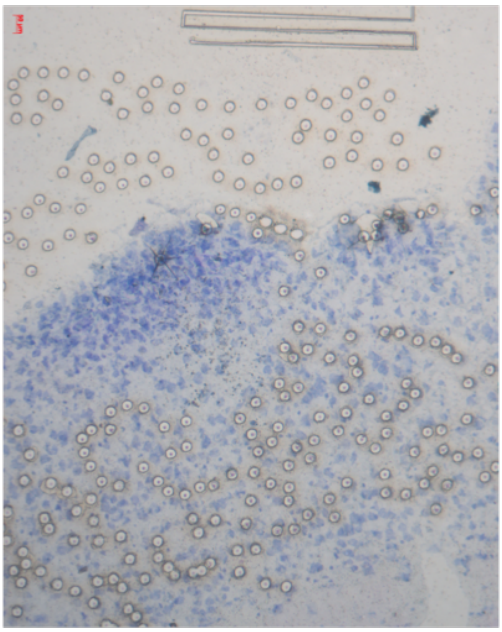
2.2. Laboratuvar çalışmaları

Çalışma için seçilen beyin örnekleri laboratuvar çalışmaları için Almanya'nın Leipzig Üniversitesi'ne bağlı Paul Fleissig Enstitüsü'ne gönderilmiştir. Çalışmanın laboratuvar aşamaları Paul Fleissig Enstitüsü'nde yürütülmüştür.

2.2.1. Hücrelerin ayrıştırılması

2.2.1.1. Lazerle kesme (LCM)

Donmuş postmortem beyin temporal korteks ve hipokampüs örneklerinden kesitler alınmış, bunlar metilen mavi ve NeuN sinir hücresi boyasıyla işaretledikten sonra hücre şekillerine göre uzman tarafından belirlenen **piramidal nöron çekirdekleri** lazerle kesme ve yakalama (İng. "laser-capture-microdissection", LCM) ile kesilerek toplanmıştır. Bu işlem için **PALM MicroBeam sistemi** (Carl Zeiss, Jena, Almanya) kullanılmıştır. PALM ile hücre toplama sırasında olası kontaminasyonu tespit etmek için slaytlarda hücre içermeyen bölgelerden negatif kontroller de alınmıştır (**Şekil 1**).



Şekil 1. Metilen mavi ile boyanmış insan beyin kesitlerinden PAML ile yapılan örnekleme. Şeklin alt tarafındaki delikler seçilen sinir hücreleri kesildikten sonra kalan boşlukları, şeklin üst tarafındaki delikler ise doku dışındaki bölgede, olası DNA kontaminasyon düzeyini belirlemek üzere seçilmiş negatif kontrol kesitlere ait deliklerini göstermektedir.

2.2.1.2. Floresan ayırma (FACS)

FACS ile örneklerin ayrılması Matevossian ve Akbarian'ın (2008) önerdikleri protokole göre yapılmıştır. Özetle, donmuş beyin örnekleri hipotonik bir solüsyonda çözüldükten sonra çekirdekler toplanmış, **NeuN** antikorları kullanılarak sinir hücrelerine ait çekirdekler işaretlenmiş, ardından BD FACSAria II SORP (BD Biosciences) tipi floresan ayırma cihazında PBS tamponu kullanılarak 96'lık kuyulu tabaklara ayrılmıştır.

2.2.2. Dizileme örneklerinin hazırlanması

Seçilen hücrelerden her birinin DNA'sı Sigma'nın **WGA4** (GenomePlex® Single Cell Whole Genome Amplification Kit, Sigma-Aldrich) protokolüne göre çoğaltılmıştır. Bu kit ve protokol kapsamında hücreler parçalandıktan, DNA yaklaşık 400 bç uzunluğunda parçalanmış, ardından polimeraz zincir reaksiyonu ile 25 çevrim çoğaltılmıştır. Bu yöntemin sinir hücreleriyle denenen başka protokollere göre daha iyi sonuç verdiği, örneğin DNA uzunluğu dağılımının az varyans gösterdiği daha önce rapor edilmiştir (Ning et al. 2015). Kalite kontrolünden geçirilen örneklere, TruSeq Nano adaptörleri eklenmiştir. Daha sonra bu

örnekler **Illumina HiSeq2500** cihazında 100 veya 150 PE (iki taraftan 100 veya 150 baz) yüksek çıktı (İng. “high output”) modunda tek bir şeritte (İng. “lane”) yürütülmüştür.

Toplam **10 farklı dizileme işlemi** farklı zamanlarda ve farklı kurumlarca (BM firması aracılığıyla Macrogen, Kore ve Leipzig Üniversitesi, Almanya) yürütülmüştür. Bu 10 işlemin ürünü olan dizileme veri setleri tarafımızdan UL_0-UL_9 şeklinde etiketlenmiştir (**Tablo 1**). Üretilen veriler ODTÜ'deki sunucumuza aktarılmıştır. Sonuçların yayımlanması sürecinde veriler ENA veri tabanına yüklenecektir.

Tablo 1: Çalışmada kullanılan bireylerin ve örneklerin özellikleri.

Proje kapsamında üretilen verilerin özellikleri tabloda özetlenmiştir. Pilot deney için kullanılan bireyler yıldız ile işaretlenmiştir. 5433 ve 5433_H numaralı örnekler aynı bireyin sırasıyla temporal korteks ve hipokampüs örnekleridir. 5603 ve 5603_F numaralı örnekler aynı bireyin sırasıyla LCM ve FACS ile işlenmiş örnekleridir.

Örnek no.	**Beyin bölgesi**	**Hücre tipi**	**Durum**	**Cinsiyet**	**Braak seviyesi**	**Örnek sayisi**	**Veri seti adı**
106*	temporal korteks	nöron ve negatif kontrol	AH	erkek	IV	17	UL_0
108*	temporal korteks	nöron ve negatif kontrol	AH	erkek	III	17	UL_0
107	temporal korteks	nöron	kontrol	erkek	-	64	UL_7
5433	temporal korteks	nöron	kontrol	dişi	-	64	UL_7
7654	temporal korteks	nöron	AH	erkek	VI	68	UL_7
6203	temporal korteks	nöron	kontrol	erkek	-	64	UL_1
3549	temporal korteks	nöron	kontrol	erkek	-	64	UL_2
3862	temporal korteks	nöron	AH	erkek	III	64	UL_2
6563	temporal korteks	nöron	AH	erkek	V	63	UL_3
7753	temporal korteks	nöron	AH	erkek	IV	64	UL_3
8401	temporal korteks	nöron	kontrol	dişi	-	64	UL_3
5603	temporal korteks	nöron	kontrol	dişi	-	64	UL_4
5603_F	temporal korteks	nöron	kontrol	dişi	-	12	UL_6
5946	temporal korteks	nöron	AH	dişi	VI	64	UL_4
7903	temporal korteks	nöron	AH	dişi	IV	64	UL_4
8730	hipokampüs	nöron	kontrol	erkek	-	100	UL_8

5433_H	hipokampüs	nöron	kontrol	dişi	-	100	UL_8
8326	hipokampüs	nöron	AH	erkek	VI	86	UL_9
5138	hipokampüs	nöron	AH	dişi	VI	87	UL_9

2.3 Hesaplamalı iş hattının oluşturulması

2.3.1 Verilerin ön işlenmesi ve iş hattının optimizasyonu

Kendi ürettiğimiz veriler, **17 bireyden** alınmış toplam **1190 kütüphaneden** oluşmaktadır. Örneklerimiz **tekil nöronal çekirdek (n= 1175) ya da negatif kontrollerden (n= 15)** oluşmaktadır**.** On yedi bireyden ikisinin hücreleri iki farklı deneyde yeniden kullanılmıştır: 5603 numaralı bireyin hücreleri hem FACS hem de LCM deneyi için kullanılmıştır. 5433 numaralı bireyin ise hem hipokampüs hem de temporal korteksinden alınan hücreler dizilenmiştir.

Hipokampüse ait toplamda 373 hücre dizilenirken, bu sayı **temporal kortekste** 817'dir. Cinsiyet dağılımları ve bireylerin durumları **Tablo 2**'de özetlenmiştir.

Tablo 2: Proje dahilinde hücreleri dizilenen bireylerin cinsiyet dağılımları, kullanılan beyin bölgeleri kullanıldığı ve durumları (kontrol/ AH).

Beyin Bölgesi	**Durum**	**Cinsiyet**
Temporal korteks	7 kontrol / 8 AH	9 erkek / 6 dişi
Hipokampüs	2 kontrol / 2 AH	2 erkek / 2 dişi

2.3.1.1 Dizileme verisinin ön kalite kontrolü

Veriler analize dahil edilmeden önce çesitli kalite kontrol testlerine tabi tutulmuştur. Öncelikle her bir örnek (hücre) için üretilen kütüphanelerin okuma sayıları incelenmiştir. Dizileme deneyleri arasında hücre başına okuma değerleri fark göstermektedir. On farklı dizileme deneyinin her birinde hücre (kütüphane) başına ortalama (medyan) **80 bin ila 2.1 milyon okuma** üretilmiştir (**Tablo 3**).

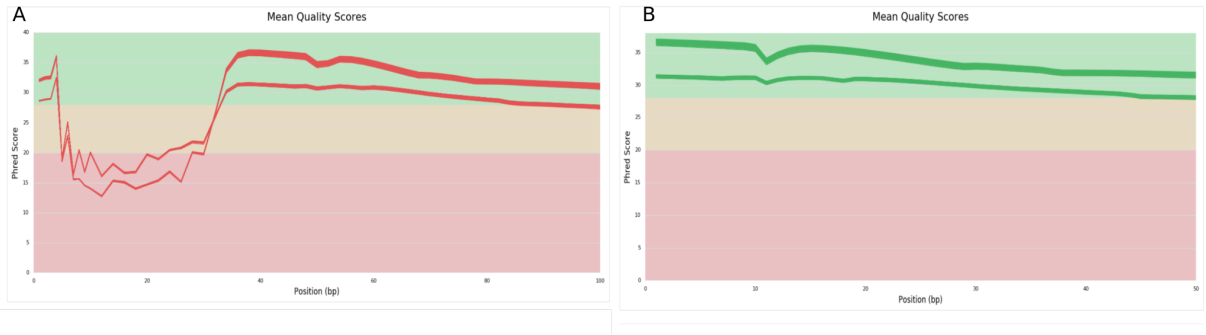
Tablo 3: Dizilenen kütüphanelerin ortalama (medyan) okuma sayıları

Veri seti	Beyin bölgesi	Kütüphane başına okuma sayısı (medyan)	Hücre sayısı	Deneysel method
UL_0	temporal korteks	505,392	34	LCM
UL_1	temporal korteks	506,449	64	LCM
UL_2	temporal korteks	918,594	128	LCM
UL_3	temporal korteks	1,040,857	191	LCM
UL_4	temporal korteks	1,628,523	158	LCM
UL_6	temporal korteks	2,067.378	15	FACS
UL_7	temporal korteks	1,482,247	12	LCM
UL_8	hipokampüs	332,810	232	LCM
UL_9	hipokampüs	79,948	148	LCM

UL_8 ve UL_9 veri setleri hipokampüs bölgesine ait olup, 2 AH (hücre sayısı = 173) ve 2 kontrol (hücre sayısı = 200) bireyi içermektedir. **Tablo 3**'te de görüldüğü üzere bu veri setlerinde kütüphane başına ortalama okuma değerleri diğerlerine göre daha düşüktür. Bu iki deneye ait hücreler (kütüphaneler) arasından, okuma sayısı medyan değerinden düşük olan hücreler (her veri setinin kendi medyan değeri baz alınmıştır) veri setinden çıkarılmıştır. Bu işlem sonucunda hipokampüse ait toplam örnek sayısı 184'e düşmüştür.

Daha sonra ham örnekler **FastQC** programına (Andrews vd., 2010) sokularak çeşitli açılardan incelenmiştir. FastQC programı (Andrews vd., 2010) her bir örneği GC oranı, okuma uzunluğu, klonalite, örnekte bulunabilecek adaptör ve aşırı temsil edilen diziler hakkında da istatistikler sunmaktadır. Örneğin **Şekil 2**'de, UL_2 veri seti için tüm okumalardan elde edilmis 'phred' kalite değerinin ortalamasını gösterilmektedir. 'Phred' kalite

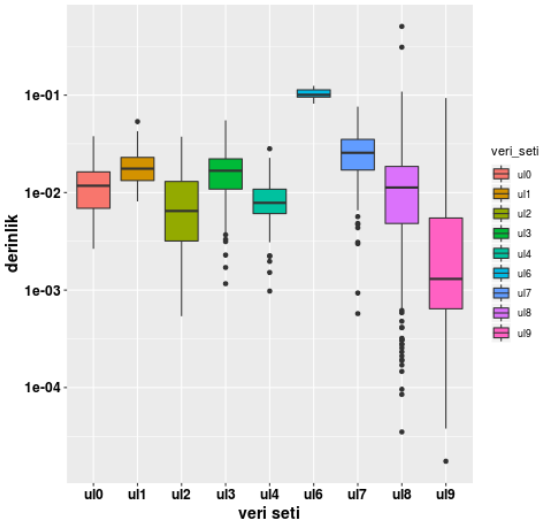
değeri, log10 bazında hata ihtimalinin -10 katıdır; örneğin ‘phred’ kalite değerinin 30 olması, 1000’de 1 hata ihtimalini ifade etmektedir. **Şekil 2**’de görülen yeşil bölge baz kalitesinin kabul edilir seviyede olduğu aralığı göstermektedir. Şekil 2A panelinde görüldüğü üzere ham okumalarda yaklaşık ilk 40 baz çiftinin (bç) ‘phred’ kalite değerleri kabul edilebilir seviyenin altındadır. Bu durum dikkate alınarak okumalardan ilk 40 bç Trimmomatic (Bolger vd., 2014) programı kullanılarak kırpılmıştır. Bu şekilde okumaların baz kalite değerleri asgari seviyeyi sağlamaktadır (Şekil 2B). Bu işlem tüm kütüphaneler için yürütülmüştür.



Şekil 2: Okumaların uzunlukları boyunca kalite değerleri.

Y-ekseni UL_2 veri seti içindeki tüm kütüphanelere (n= 128) ait okumalardan elde edilmiş ‘phred’ kalite değerinin (1 ila 40 arası değişmektedir) ortalamasını, X-ekseni ise okumalardaki her bir bazın okumanın 5’ ucundan itibaren pozisyonunu göstermektedir. A ve B panelleri sırasıyla aynı kütüphaneye ait ham sonuçları ve ilk 40 bç kırpıldıktan sonraki sonuçları göstermektedir.

Ayrıca her veri seti için dizileme derinliği (İng. “depth”), yani genomda ortalama her bir bazın kaç defa okunduğu belirlenmiştir. *samtools*’un ‘*depth -a’* argümanı ile belirlenen derinlikler **Şekil 3’**te gösterilmiştir. Burada görüldüğü üzere hücrelere ait kütüphanelerin çoğu 0.01X civarında (yani her 100 bazdan birinin bir defa okunduğu sıklıkta) dizilenmiştir.



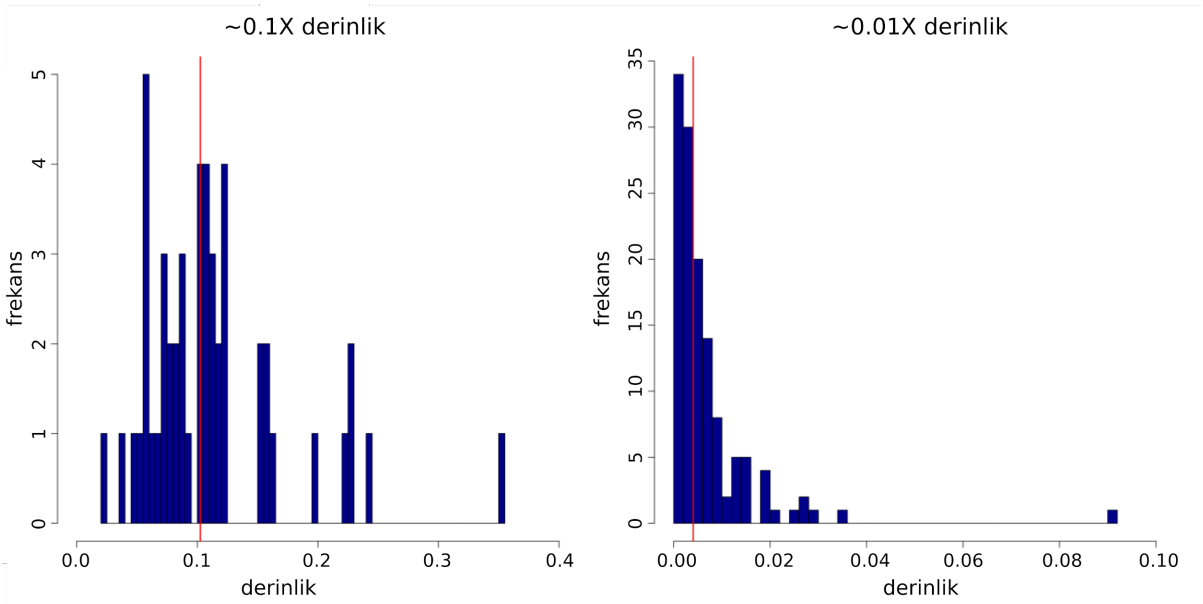
Şekil 3. Proje kapsamında dizilenen tekil sinir hücrelerinin dizileme derinlikleri.

X-ekseni dizileme işleminin etiketini (veri seti ismini), Y-ekseni ise log10 tabanında kütüphane başına dizileme derinliği değerlerini göstermektedir. Veri setlerinin medyan derinlik değerleri sırasıyla: 0.01, 0.02, 0.01, 0.02, 0.01, 0.1, 0.03, 0.01, 0.001 olarak belirlenmiştir.

2.3.1.2 Düşük dizilemenin etkisinin kontrolü

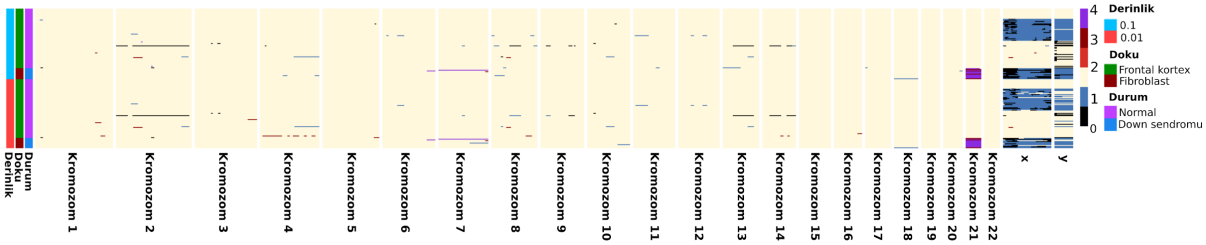
Çalışmamızda mali kısıtlar nedeniyle sınırlı miktarda dizileme yapılmıştır. Yakın zamanda yayımlanan van den Bos ve arkadaşlarının (van den Bos vd., 2016) çalışmasında bizim çalışmamıza benzer şekilde milyon okumadan daha düşük derinlikte dizilenmiş kütüphaneler kullanılmışsa da, birçok diğer çalışmada dizileme derinliği 0.1X civarında, yani bizim çalışmamızdan 10 kat daha yüksektir. Söz konusu **dizileme derinliği farkının** sonuçlarımızı etkileyip etkilemeyeceğini test etmek için, yayınlanmış tek hücre genom dizileme verileri kullanılarak bir **simülasyon deneyi** yürütülmüştür. Söz konusu veriler arasında en yüksek derinliğe sahip olan veri seti (~0.1X; McConnell vd.) seçilmiştir. Buradaki örneklerin derinlikleri, *samtools* (v. 1.3) programı (Li vd., 2009) kullanılarak, her bir kütüphanedeki okumaların bir kısmının program tarafından rastgele seçilmesiyle **~0.01X'e düşürülmüştür** (**Şekil 4**). Yapay olarak derinliğini düşürdüğümüz bu veriler ile orijinal (~0.1X) veriler, paralel olarak, bir sonraki bölümde ayrıntılandırılan yöntemler ve *HMMcopy* algoritması (Lai vd., 2016) kullanılarak anöploidi tahmini analizine tabi tutulmuştur. Bu paralel analiz sonucunda, derinliğin azaltılmasından önce %7.3 olan kayıp veya kazanç vakalarının, sonrasında da çok yakın değerde kaldığı ve **aynı vakaların bulunduğu** tespit edilmiştir (**Şekil 5**). Örneğin şekilden görülebileceği üzere 0.1X veri ile sistematik olarak tespit edilebilen Trizomi 21,

0.01X veri ile de tespit edilebilmektedir. Bu karşılaştırmadan, hücre başına **0.01X dizileme derinliğinin anöploidi tespitini zorlaştırmadığı** sonucuna varılmıştır.



Şekil 4. McConnell vd. (2013) veri setinde düşük dizileme derinliği simülasyonu.

McConnell vd. (2013) veri setinden, sağlıklı iki dişi ve bir erkek bireye ait toplam n=110 kortikal nöron ile Trizomi 21 taşıyan bir erkek bireye ait n=13 fibroblast hücresinin verileri elde edilmiştir. Yaklaşık 0.1X derinlikteki (sol panel) bu tüm genom dizileme verileri *samtools* (v. 1.3) programı kullanılarak ~0.01X derinliğe (sağ panel) indirilmiştir. Histogramlarda X-ekseni her bir kütüphanenin dizileme derinliğini, Y-ekseni de frekansı göstermektedir.



Şekil 5: McConnell vd. (2013) veri setinde düşük ve yüksek derinlikte verilerle anöploidi tahmini.

Şekil 4'te gösterildiği şekilde 0.1X (gerçek veri) ve 0.01X (simülasyonla seyreltilmiş veri) ortalama derinlikteki veri setleri, aynı algoritma (*HMMcopy*) kullanılarak analiz edilmiş, düşük derinliğin anöploidi tahminini etkileme seviyesi ölçülmüştür. Figürdeki her bir satır bir hücreyi temsil etmektedir. Matriste 2 kopya (dizomi) tahmin edilen bölgeler sarı renk, 1 kopya

lacivert, 0 kopya siyah, 3 kopya kırmızı, 4 kopya koyu kırmızı, 5 kopya mor ile gösterilmiştir. Bir kromozomun tümünü kaplayan yatay renkli çizgiler anöploidiyi, kısmi renkli çizgilerse diğer KSV'leri göstermektedir. En soldaki "Derinlik" anahtarında gösterilen açık mavi ve pembe bloklar sırasıyla yüksek ve düşük derinlikteki verileri temsil etmektedir. "Doku" anahtarında yeşil frontal korteksi, kırmızı ise fibroblastları göstermektedir. "Durum" anahtarında mor sağlıklı bireyleri, koyu mavi ise Trizomi 21 (Down sendromlu) bireyleri göstermektedir. 0.1X ve 0.01X derinliklerine ait sonuçlar yukarıdan aşağıya aynı sırayla gösterilmiştir.

2.3.2. Üretilen verilerin analizi

2.3.2.1 Okumaların hazırlanması

Yukarıdaki anlatılan ön analizlerin ardından, hazırladığımız iş hattı deneylerimiz kapsamında üretilen tek hücre genom verilerine uygulanmıştır. **Bölüm 2.3.1**'de de anlatıldığı üzere, FastQC programı sonucunda elde edilen bilgilerden yola çıkarak dizi uçlarında bulunan adaptör dizileri 40 bç'ye kadar **kırpılmıştır**. Ayrıca ihtiyatlı olmak açısından FastQ Screen (Andrews vd., 2011) isimli program ile örneklerde kontaminasyon olup olmadığı ve varsa kontaminasyona hangi canlının DNA'sının (ör. *E. coli*) sebep olduğu incelenmiştir.

KR-KSV hesaplamak için belli okuma uzunluklarına özgü **hizalanabilirlik (İng. "mappability") dosyalarına** ihtiyaç duyulmaktadır. WIG formatındaki bu dosyalar genom boyunca özgün olarak hizalanmaya açık pozisyonların istatistiğini içerir. Çalışmamızda bu dosyalardan 50 bç veya 100 bç için hazırlanmış olanları UCSC Genom Tarayıcısından (genome.ucsc.edu) indirilerek kullanılmıştır. Bununla tutarlı olması için dizilenen kütüphanelerin okuma uzunlukları da **50 bç** veya **100 bç** olacak şekilde kesilmiştir. Bu değerlerin altında kalan uzunluklara sahip okumalar atılmıştır.

2.3.2.2 Hizalama ve filtreleme işlemleri

Hizalama işlemi şöyle yürütülmüştür: 50 bç'ye kesilen diziler insan referans genomuna (hg19) ***bwa*** (Li ve Durbin, 2009) programının "*aln*" algoritması kullanılarak, 100 bç'ye kesilenler ise yine *bwa*'nın "*mem*" algoritması kullanılarak hizalanmıştır. Bunun sebebi ise "aln" algoritmasının 50 bç ve altında, "mem" algoritmasının ise daha uzun dizileri hizalarken doğruluk payının ve verimliliğinin yüksek olmasıdır. İki algoritmanın da standart parametreleri kullanılmış ve 96% eşleşme şartı aranmıştır.

Bu aşamada oluşturulan SAM dosyaları, *samtools* programı yardımıyla BAM indeks ve sıralanmış BAM formatına çevrilmiştir (Li vd., 2011). Bu sırada üç filtreleme işlemi uygulanmıştır:

- Dosyalar BAM formatına çevrilirken *samtools*'un '*view -F 256*' *argümanı* eklenerek, birden fazla lokusa hizalanan okumalar atılmıştır.
- Buna ek olarak *samtools*'un '*view -q 60*' argümanı kullanılarak da hizalanma kalitesi düşük olan okumalar uzaklaştırılmıştır.
- Polimeraz zincir tepkimesi (PCR) sapma nedeniyle meydana gelen tekrarlayan (duplike) kopyalar da *samtools rmdup* kullanılarak atılmıştır.

Böylece yalnızca genomda **özgün tek bir yere hizalanan ve yüksek kalitede olan okumalar** analizin geri kalanında kullanılmıştır.

2.3.2.3 HMMcopy ile kopya sayısı tahmini

Hizalanan ve filtrelenen okumalar ilk olarak *HMMcopy* (Lai vd., 2016) programı kullanılarak KR-KSV tahmininde kullanılmıştır. *HMMcopy* algoritması, bir örneğe ait düşük seviye genom dizileme verisinin her bir kromozom boyunca oluşturulan geniş (genelde **500 bin bç** [500 kb] uzunluğundaki) pencerelerde okuma yoğunluğunu kullanır. *HMMcopy* bir Gizli Markov Modeli (HMM, İng. "Hidden Markov Model") yardımıyla **ardıl pencerelerdeki okuma yoğunluğunun** genom çapındaki okuma yoğunluğundan farkını kullanarak, kromozom kopya sayısı tahmin eder. *HMMcopy*, aynı zamanda pencere başına okuma yoğunluğu değerlerini, aynı pencerelerdeki GC yoğunluğu ve hizalanabilirlik özelliklerine göre de **normalize eder**.

HMMcopy kullanarak yapılmış çalışmalarda, normalize edilmiş verilerden üretilen **pencere başına gözlenen/ beklenen okuma oranları** (log2 tabanına çevrilmiş halde) KR-KSV yaygınlığını belirlemede kullanılan ana istatistiktir. Buna **KR-KSV değeri** de denebilir. Verilerin işlenmesi sürecinde, her bir pencere için hesaplanan KR-KSV değerleri HMM metoduyla sekmelere ayrılmıştır.

Yazılımın bir örnekte çalıştırılmasının ardından, her bir kromozom için, kopya sayısı bakımından **kendi içinde tutarlı ama genom ortalamasından farklı genomik segmentler** çıktı olarak elde edilir. Yazılım, rapor ettiği her segment için bir kopya sayısı kategorisi atamaktadır:

- Kategori 1 veya *HOMD*: Homozigot delesyon olayı [~0 Kopya]
- Kategori 2 veya *HETD*: Heterozigot delesyon olayı [~1 Kopya]

- Kategori 3 veya *NEUT*: Nötral KSV [~2 Kopya]
- Kategori 4 veya *GAIN*: Kromozom kazancı [~3 Kopya]
- Kategori 5 veya *AMPL*: Amplifikasyon olayı [~4 Kopya]
- Kategori 6 veya *HLAMP*: İleri derece amplifikasyon olayı [~5 Kopya]

Çalışmamızda hizalanan okuma verileriyle 500 kb'lik pencereler başına okuma yoğunluğu hesaplanmış, bu sonuçlar WIG formatına dönüştürülmüş, sonra da GC oranı ve hizalanabilirlik için normalize edilmiştir. Optimizasyon çalışmalarında 250 kb ve 1 Mb'lik pencereler de denenmiş olup, okuma sayısına uygunluğu ve tahmini anöploidiyi daha güvenilir yakalaması nedeniyle 500 kb'lik pencerelere karar verilmiştir (veri gösterilmemiştir). Ayrıca 500 kb'lik pencereler literatürde yaygın olarak tek hücre genom analizlerinde de kullanılmıştır (Knouse vd., 2014).

2.3.2.4 Değişkenlik değeri filtresi

Yakın zamanda yayımlanan birçok tekil hücre genom dizileme çalışması, dizilenen kütüphanelerde teknik gürültüyü ölçmek üzere çeşitli istatistikler kullanmışlardır. Bunlardan biri **değişkenlik değeri** (DD, İng. "variability score") adında bir istatistiktir (Knouse vd., 2014; Knouse vd., 2016). Değişkenlik değeri şu şekilde hesaplanır: Her bir tekil hücre için, kromozomların okuma pencerelerine denk gelen okuma sayılarının standart sapması kaydedilip, en yüksek standart sapmaya sahip üç kromozomun değerlerinin aritmetik ortalaması, o hücre örneği için değişkenlik değeri olarak rapor edilir. Bu değer özünde, bir tekil hücreye ait okuma yoğunluğu verisinin her bir kromozom içinde ne kadar değişkenlik içerdiğinin bir ölçüsüdür. Kromozom içinde aşırı varyasyonun biyolojik değil teknik gürültü kaynaklı olması muhtemeldir. Dolayısıyla yüksek değişkenlik değeri, bir hücreye ait veride teknik gürültünün baskın olduğunun ve bu verinin KR-KSV tahmini açısından kullanışsız olabileceğinin bir göstergesidir.

Knouse vd., (2016) hem ampirik veri hem de benzetimler kullanarak DD'nin gerçek anöploidi tarafından etkilenmediğini, ayrıca bir hücre için hesaplanan DD arttıkça bu hücrede tespit edilen KR-KSV vakalarının güvenirliklerinin azaldığını göstermişlerdir. Söz konusu çalışmanın yazarları genel olarak DD'nin 0.3'ün altında olmasını önermektedir. Söz konusu makalenin sonuçlarına bakılırsa, DD'nin 0.5'in üzerinde olduğu hücrelerde KR-KSV tayini fiilen mümkün değildir. Çünkü bu seviyede **hem sahte pozitif, hem de sahte negatif** vakaların sıklığı çok yükselmektedir.

Bu nedenlerle çalışmamızda yalnızca **DD değeri 0.5'in altında olan hücreler** kullanılmıştır. Dizilediğimiz hücrelerin 472'si bu kriteri karşılamaktadır.

2.3.2.5 Dairesel ikili segmentasyon

Yakın zamanda Knouse ve arkadaşları tarafından KSV içerdiği rapor edilen genomik segmentlerin doğruluğunu artırmak amacıyla, yaygın olarak kullanılan *HMMcopy* algoritmasının yanı sıra ikinci bir algoritma daha kullanılması önerilmiştir (Knouse vd., 2016; Knouse vd., 2017). Dairesel ikili segmentasyon (İng. "circular binary segmentation") adında bu algoritma *DNAcopy* isimli yazılım (bioconductor.org/packages/release/bioc/html/DNAcopy.html) kapsamında genom verisine uygulanabilmektedir. Bu yöntem Markov Modeli temelli olmayıp, özellikle biyolojik verilerin analizi için geliştirilmiş ve mikrodizin analizinde yaygın olarak kullanılmış bir algoritmadır (Olshen vd., 2004).

DNAcopy yazılımı için girdi hazırlama ve veri ön işleme basamakları *HMMcopy* girdileri için yukarıda anlatılan işlemlerden kısmen farklılık gösterir. Çalışmamızda, söz konusu işlemler sırasında yine Knouse ve arkadaşları (Knouse vd., 2017) tarafından sağlanan betikler kullanılmıştır. Bu basamaklar şu sırayla gerçekleşir:

- PCR duplikelerinden arındırılmış ve tekrar sıralanmış BAM dosyaları, UCSC Genom Tarayıcısı (genome.ucsc.edu) tarafından sağlanan ve hg19 insan genomundaki dinamik 500 kb uzunluğundaki hizalanabilir pencerelerin konumunu tanımlayan BED dosyası bir referans olarak kullanılarak, *bedtools* yazılımının *coverage* modülü ile okuma sayısı verisine çevrilir.
- Bir önceki basamaktan elde edilen okuma sayma verisi, ilgili 500 kb pencerelerin her biri için GC içeriği bilgisi içeren bir dosya ile birlikte girdi olarak kullanılarak, GC içeriğine göre normalize edilmiş pencere başına okuma sayıları elde edilir.
- Sonuçta her bir tekil hücre verisi için elde edilen normalize edilmiş pencere başına okuma sayısı verisi, *DNAcopy* R kütüphanesi bünyesinde bir dizi zincir fonksiyon için girdi olarak kullanılarak, *HMMcopy* için alınan çıktı benzeri bir çıktı elde edilir.

2.3.2.6 KR-KSV tahminlerini kesiştirme ve ilgili segmentleri belirleme

Knouse ve arkadaşlarının (Knouse vd., 2016; Knouse vd., 2017) önerileri doğrultusunda, sahte pozitif oranını asgariye indirmek üzere hem *HMMcopy* hem *DNAcopy* tarafından da kazanç veya kayıp olarak rapor edilen ve genomik lokasyonları örtüşen, yani **kesişen**

segmentler gerçek vakalar olarak kabul edilmiştir. İş hattının ileri aşamalarında yalnızca bu segmentlerle devam edilmiştir.

Kesiştirme öncesinde, her iki yazılımın çıktıları içerisinden de dizomi kopya sayısı profiline sahip, yani **normal segmentler**, Knouse ve arkadaşları (Knouse vd., 2016) tarafından ampirik olarak belirlenmiş eşik değerleri kullanılarak elenmiştir. Bu işlem *HMMcopy* çıktıları için, segment log2 medyan kopya sayısı değeri **0.4** ile **-0.35** değerleri arasında olan segmentlere, *DNAcopy* çıktıları için ise **1.32** ile **0.6** değerleri arasındaki segmentlere karşılık gelmektedir. Kesiştirme sırasında kopya sayısı profili ataması için *HMMcopy* tarafından atanan kategoriler, kopya sayısı üzerinden yapılan filtreler için ise *DNAcopy* tarafından rapor edilen segment medyan log2 kopya sayısı değerleri kullanılmıştır.

Kesiştirme işlemi şu şekilde yapılmıştır:

- Her bir örnek için, o örneğe ait rapor edilen *HMMcopy* ve *DNAcopy* çıktı segmentleri kayıp veya kazanç segmentleri ayrı dosya çiftlerinde yer alacak şekilde diske yazılır.
- *bedtools* yazılımının *intersect* modülü kullanılarak, kazanç ve kayıp çıktılarının işlemleri birbirlerinden ayrı tutulacak şekilde iki çıktının genomik lokasyonları kesiştirilir.
- Bir önceki basamak sonucunda elde edilen genomik segmentlerin kopya sayılarının, rapor edildikleri HMM kopya sayısı kategorilerince **beklenen kopya sayılarına uzaklıkları** hesaplanır.
- Ait olduğu kopya sayısı kategorisi uyarınca beklenen kopya sayısına uzaklığı **0.5** değerini aşan segmentler filtrelenir.

Kesiştirme sonunda elde edilen segmentlerin beklenen kopya sayısına olan uzaklıklarına göre filtrelenmesi işleminde Rohrback ve arkadaşlarının yaklaşımından ilham alınmıştır (Rohrback vd., 2018). Bu işlemde eşik değeri olarak 0.5 seçilmesinin sebebi, kesiştirme öncesi dizomi profiline sahip segmentler veri setinden elenirken kullanılan eşikle tutarlılık gözetmektir (nitekim *DNAcopy* çıktıları için önerilen logaritmik ölçekteki üst eşik değeri 1.32, gerçek kopya sayısı cinsinden **2.5**'e, logaritmik ölçekteki alt eşik olan 0.6 değeri ise gerçek kopya sayısı cinsinden **1.5**'a karşılık gelmektedir. Dolayısıyla dizomi olmadığı düşünülen segmentlerin dizomi için beklenen kopya sayısına uzaklığı 0.5 olacaktır.) Bu ek filtreleme işleminin, hem verimizdeki biyolojik olarak bir anlam ifade etmeyen teknik varyasyonu temizlemek, hem de kopya sayısı profilleri için kategori ataması yaparken her bir kategori için beklenen ortalama kopya sayısı değerlerini bir anlamda veriden öğrenme yaklaşımına sahip olan *HMMcopy* yazılımının çıktılarını doğrulamak gibi avantajları olacağı düşünülmektedir.

Gerekli filtrelemeler yapıldıktan sonra toplam 472 tekil hücre verisi KR-KSV tahmini için analiz edilmiştir. Eldeki segmentler, segment uzunlukları ve rapor edildikleri kromozomun uzunlukları kullanılarak iki ayrı tipte veriye ayrılmıştır: (a) uzunluklarından bağımsız **bütün KSV** sinyalleri, (b) yalnızca **tüm kromozom anöplöidi** sinyalleri. Tüm kromozom anöplöidi sinyalleri belirlenirken rapor edilen segmentin uzunluğunun segmentin rapor edildiği kromozomun uzunluğunun **en az %90'ını** kapsaması kıstas olarak kullanılmıştır.

2.3.2.6 İstatistiksel analizler

Çalışmanın son aşamasında KR-KSV sinyallerinin hangi faktörlerden etkilendiği test edilmiştir. Test için dizilenen ve filtreleri geçen tüm hücrelere ait her bir otozomal kromozoma ait sonuçlar bir araya getirilmiştir. Her bir kromozom için dizomi hali 0, kayıp veya kazanımlar 1 olarak kodlanmıştır. Yürütülen testlerde toplam 472x22 = 10,384 kromozoma ait bu binom kodlu değerler **bağımlı değişken** olarak kullanılmıştır.

Kayıpları ve kazanımları ayrı ayrı analiz etmek için de, yalnızca ilgilenilen vakaların (yalnızca kayıpların veya yalnızca kazanımların) 1, diğer tüm durumların 0 olarak kodlandığı veriler kullanılmıştır.

Bu tip bir verinin analizinde çok faktörlü regresyon kullanılmasının uygun olduğuna karar verilmiştir. Elimizdeki verinin sıfır yığılmalı bir doğası olması nedeniyle, sayma verisi için geliştirilmiş **sıfır-yığılmalı negatif binom regresyonu** modeli kullanılmıştır. Bu yöntem, veriyi iki dağılımın bileşimi olarak modeller: Hiç tespit edilmeyen vakalar ve tespit edilip negatif binom dağılım gösteren vakalar. Negatif binom dağılım, Poisson'a benzer ancak daha esnek olması (ortalama ve standart sapmanın tek bir parametre değeriyle belirlenmemesi) açısından tercih edilmiştir. İlgili istatistiksel modelin kurulması ve test amaçlı kullanılmasında, *glmmADMB R* kütüphanesi *glmmADMB* fonksiyonu (Fournier vd., 2012) kullanılmıştır.

Regresyon modelinde kullanılan **tahmin edici faktörler** şunlardır:

- Sinyalin hangi **kromozomdan** geldiği,
- Sinyal alınan hücrenin **AH veya kontrol** bir birey olması,
- Sinyal alınan hücrenin **temporal korteks veya hipokampüs** bölgelerinden gelmesi,
- Sinyal alınan hücrenin **değişkenlik değeri**,
- Sinyal alınan hücrenin hangi **bireye** ait olduğu.

Birey faktörü **rastgele değişken**, diğer değişkenler sabit değişken olarak modele katılmıştır.

Ayrıca parametrik olmayan yöntemlerde de sonuçların tekrar edilmesi tercih edilmiştir. Bunun için (a) verinin bir kısmının geri eklemeyle çok defa örneklenmesi ("sampling with replacement") ve aynı istatistiğin çok sayıda iterasyonda (tekrar işleminde) hesaplanması biçiminde yürütülen **bootstrap testi** kullanılmıştır. (b) İlgili faktörün çok sayıda iterasyonda rastgele karıştırılarak sıfır dağılımının oluşturulduğu **permütasyon testleri** kullanılmıştır.

Bu analizlerde bireyler arasında hücre sayısının farklı olması durumunu da telafi edecek bir tasarım uygulanmıştır. Özel olarak hazırlanan *R* betikleri ile iki spesifik hipotez incelenmiştir: AH ve kontrol grupları arasında fark olması, ayrıca kayıpların kazanımlardan fazla olması.

Boostrap analizinde, her bir iterasyonda, bireyler geri eklemeyle rastgele örneklenmiştir (yani toplam örneklem büyüklüğü n=17 kalırken, bir bireyin örneklemde birden fazla defa temsili veya hiç temsil edilmemesi mümkündür). Her bir iterasyonda, o sefer seçilen bireylere ait hücreler içinden yine rastgele 5 hücre seçilmiştir. Burada n=5 eşiğine, asgari sayıda filtreleri geçebilen hücreye sahip birey baz alınarak karar verilmiştir. Bu şekilde seçilerek oluşturulan alt kümelerde, kayıp ve kazanım vaka sayıları arasındaki fark hesaplanmıştır. Bu işlem 1000 defa tekrar edilmiş, dağılımın alt %2.5 ve üst %97.5 değerleri %95 güven aralığı hesaplamak için kullanılmıştır.

Permütasyon testinde ise bu iki kategori arasındaki farkın şansa bağlı olarak nasıl değişebileceği modellenmiştir. Burada her bir iterasyonda AH ve kontrol gruplarına ait bireyler rastgele karıştırılmıştır. Oluşturulan iki yapay grubun örneklem büyüklükleri sabittir (10 ve 7). Her bir gruptaki bireyler için, kendilerine ait rastgele n=5 hücre seçilmiş, bu hücreler kullanılarak iki yapay grup arasında fark hesaplanmıştır. Yapay gruplar arasındaki farkın 1000 iterasyonda elde edilen dağılımı sıfır dağılımı olarak kullanılmıştır. Bu dağılımın gerçek veride hesaplanan fark ile karşılaştırılması ile **ampirik p-değeri** hesaplanmıştır. Burada hücre sayısının minimumda tutulmasının testimizi tutucu kılacağı da belirtilmelidir.

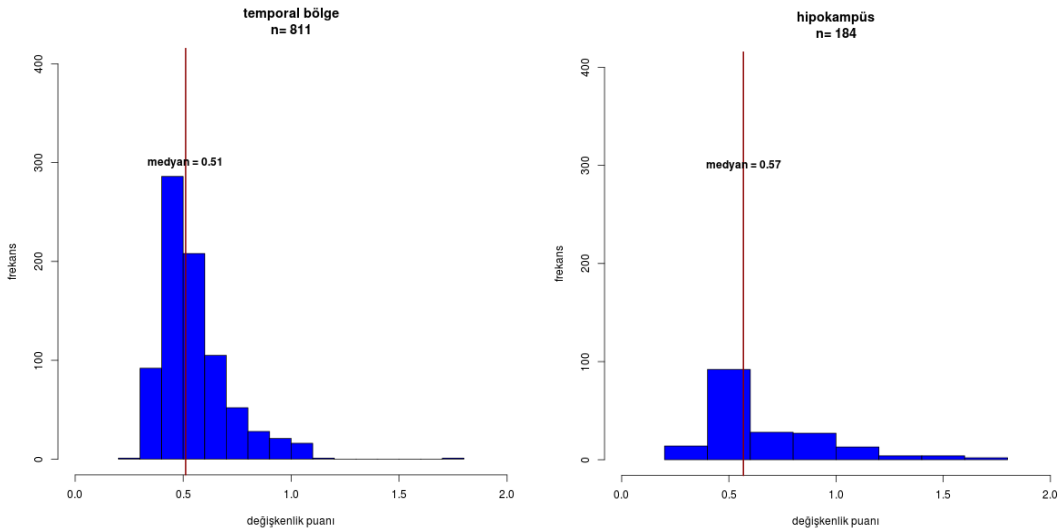
3. BULGULAR

3.1 Tekil hücre dizileme sonuçları ve deneysel yöntem etkisinin analizi

Çalışmamızda **17 bireyden alınmış 1190 kütüphane** dizilenmiştir. Bunlardan **995'i** okuma derinliği filtrelerini geçen nöronal çekirdeklere aittir (2.3.1.1). Çalışmamızın bugüne dek yapılan ve 1. Bölüm'de özetlenen yayımlanmış nöronal sgWGS çalışmalarından başlıca farkı deneyseldir. Deneysel çalışmaları yürüten Leipzig ekibinin tercihi üzerine çalışmamızdaki hücrelerin çoğu, FACS yerine, spesifik bir hücre tipinin (piramidal nöronların) mikroskop altında belirlenip çekirdeklerinin toplanmasına izin veren lazerle kesme ve yakalama (LCM) yöntemiyle elde edilmiştir.

2. Bölüm'de (Gereç ve Yöntem) ayrıntılandırıldığı üzere, dizilenen hücrelere ait veriler, kalite filtrelerinden geçirilmiş, 500 kb'lik pencelerde her bir hücre için okuma yoğunluğu hesaplanmıştır. Ayrıca her bir hücre başına, teknik gürültüyü ölçen **değişkenlik değeri** (DD) hesaplanmıştır (2.3.2.4). Bu değer incelenen hücreler arasında 0.28 ila 1.72 arasında değişmektedir (**Şekil 6**). Öte yandan çalışmamızda elde edilen temporal korteks ve hipokampüs hücrelerinin DD değerleri (sırasıyla 0.51 ve 0.57) anlamlı fark göstermemektedir (Mann-Whitney U testi p = 0.1).

Bu DD değeri dağılımın literatürdeki benzer çalışmalardakinden daha yüksek olduğu not dikkat çekmektedir (Knouse vd., 2014; Knouse vd., 2016). Örneğin Knouse ve arkadaşları 2016 yılındaki çalışmalarında inceledikleri FACS ve scWGS kullanarak inceledikleri insan nöronlarında DD'yi 0.14 ila 1.03 arasında, medyan değerini de 0.2 olarak bulmuşlardır (Knouse vd., 2016). Dizilediğimiz hücrelerin DD değerleri bundan anlamlı ölçüde yüksektir (Wilcoxon testi p < 2.2e-16).

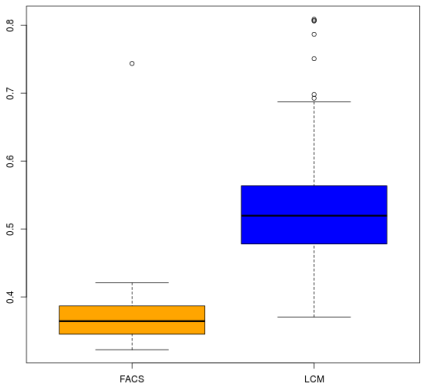


Şekil 6: Dizilenen 995 hücrede değişkenlik değeri (değişkenlik puanı) dağılımları.
Soldan sağa iki panelde sırasıyla temporal bölge ve hipokampüs bölgelerine ait hücreler için hesaplanan DD dağılımı gösterilmektedir.

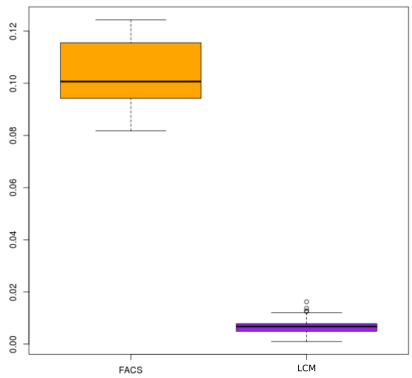
Tartışma bölümünde ele alacağımız üzere bu sonuçlar, çalışmamızda kullanılan protokolün bir veya birden fazla aşamasının **teknik gürültüye karşı hassasiyet** yarattığına işaret etmektedir. Olası bir etmen, çalışmamızda FACS yerine ağırlıkla LCM metodunun kullanılmış olmasıdır.

FACS ile LCM arasında teknik gürültü açısından fark olup olmadığı, çalışmamız kapsamında bu iki yöntemle üretilmiş verilerin özelliklerinin karşılaştırılmasıyla incelenmiştir. İlk olarak, 5603 numaralı bireye ait ve FACS ile üretilmiş n=12 hücrenin (UL_6 veri seti) değişkenlik değerleri, yine aynı bireye ait ve LCM ile üretilmiş n=12 hücrenin (UL_4 veri seti) değişkenlik değerleriyle karşılaştırılmıştır (**Şekil 7**). FACS ile üretilmiş hücrelerin değişkenlik değerlerinin anlamlı ölçüde **düşük** olduğu tespit edilmiştir (Mann-Whitney U testi $p < 7e\text{-}06$).

İkinci olarak, deneysel yöntemin dizileme derinliğine etkisi olup olmadığını anlamak için LCM ve FACS ile elde edilen hücrelerin dizileme derinlikleri karşılaştırılmıştır (**Şekil 8**). FACS ile toplanan örneklerin derinliğinin (medyan = 0.10), LCM'ye (medyan = 0.006) göre daha yüksek olduğu bulunmuştur. Yapılan bu analizler, belli özellikteki hücreleri yakalamadaki potansiyel üstünlüğüne rağmen LCM'nin FACS'a göre daha gürültülü olduğunu göstermektedir.



Şekil 7: FACS ve LCM ile üretilmiş hücrelerin değişkenlik değerleri.
Kutu grafikleri aynı bireye ait sırasıyla n=12'şer değişkenlik değerlerinin (Y-ekseni) dağılımını karşılaştırmaktadır.



Şekil 8: FACS ve LCM ile üretilmiş hücrelerin dizileme derinliği değerleri.
Kutu grafikleri UL_4 ve UL_6 veri setlerine ait sırasıyla n=64 ve n=12 hücrenin dizileme derinliği değerlerinin (Y-ekseni) dağılımını karşılaştırmaktadır.

Yayımlanmış ampirik veriler DD'nin 0.5'ten yüksek olduğu durumlarda sonuçların güvenirliğinin çok düşük olduğuna işaret etmektedir (bkz. 2.3.2.4). Bu durum göz önüne alınarak, çalışmada yalnızca DD<0.5 filtresini geçebilen 472 hücre kullanılmıştır. Bu hücrelerin 302'si temporal korteks, 70'i hipokampüsten elde edilmiştir.

3.2 KSV ve anöploidi oranı tahminleri

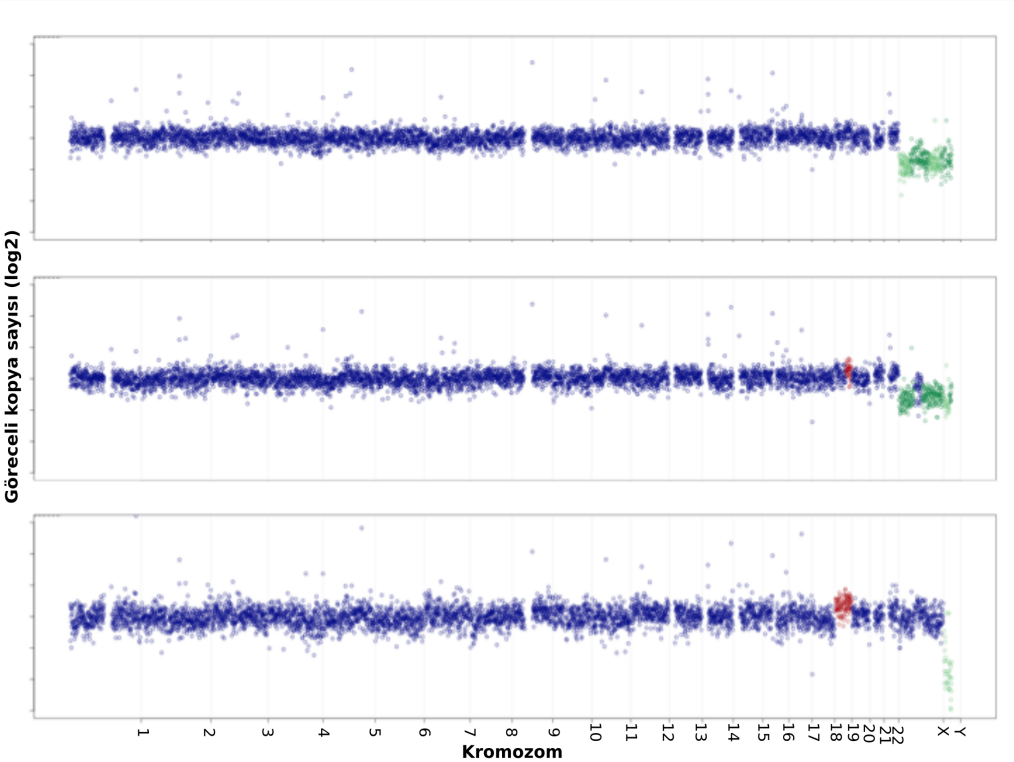
3.2.1 Tüm hücreler arasında KSV ve anöploidi oranı

Toplamda DD filtresini geçen 472 hücreye ait veri *HMMcopy* ve *DNAcopy* yazılımlarınca işlenmiş, iki yazılımın da tutarlı biçimde KSV diye tahmin ettiği segmentler belirlenmiştir (bkz. 2.3.2.6) (**Şekil 9**). İncelenen hücrelerin **~%21**'inde en az bir **KSV** sinyali (**Tablo 4**) bulunmuştur. Hücrelerin **~%6.6**'sında ise en az bir **tüm kromozom** (kromozomun %90'ını kapsayan ölçüde) anöploidi sinyaline rastlanmıştır (**Tablo 5**). Burada AH bireylerde vakaların sıklığının kontrol bireylere kıyasla biraz daha düşük olduğu görülmektedir (bu konu daha sonra incelenecektir) Tartışma bölümünde daha ayrıntılı ele alınacağı üzere, KSV sinyali oranı literatürdekinde rapor edilenlere yakın, tüm kromozom anöploidi sıklığı ise literatürdekinden yüksektir.

3.2.2 Kromozom başına KSV ve anöploidi sıklıkları

472 hücreye ait veriler daha sonra kromozomlar düzeyinde incelenmiş, her bir otozom için anöploidilerin ve KSV olaylarının toplam sayılarının sıklıkları AH ve sağlıklı kontrol bireylerde belirlenmiştir. Sonuçlar **Şekil 10 ve Şekil 11**'de gösterilmiştir. Sinyallerin hangi

kromozomdan geldiğinin KSV veya anöploidi olayı sayılarının dağılımına olan etkisi konusunda istatistiksel testler ve sonuçlarına dair ayrıntılar, diğer biyolojik faktörler için gerekli testlere dair bilgiler ve sonuçlarıyla birlikte **3.2.4.1** numaralı bölümde ele alınacaktır.



Şekil 9: Üç farklı hücrede KSV ve anöploidi tahmini.

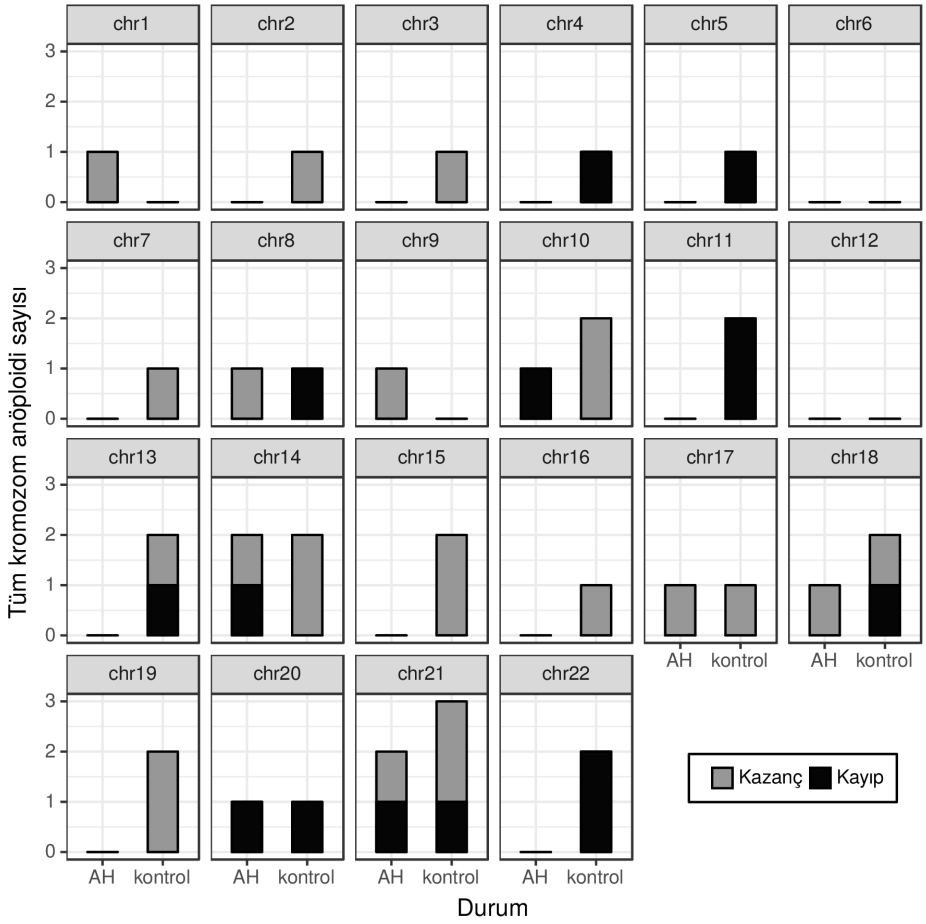
En üstten en alta: KSV tahmin edilmeyen, yalnızca bir kromozomda KSV tahmin edilen, tüm bir kromozomda KSV (anöploidi) tahmin edilen örnek hücreler. Y-ekseni 500 kb'lik pencere başına normalize edilmiş okuma değerlerini göstermektedir. Mavi noktalar dizomi, kırmızı noktalar 3 vey 4 kopya sayısı, yeşil noktalar ise 0 ve 1 kopya sayısı tahmin edilen segmentleri göstermektedir. İlk iki hücre erkek bireylere ait olduğu için X kromozomunda doğru biçimde monozomi tahmin edilmiştir.

Tablo 4: En az bir KSV sinyali gösteren hücrelerin sıklığı.

	Sinyal veren hücre sayısı	**Toplam hücre sayısı**	**Yüzde**
AH	45	200	22.5
Kontrol	54	272	19.9
AH + Kontrol	99	472	21.0

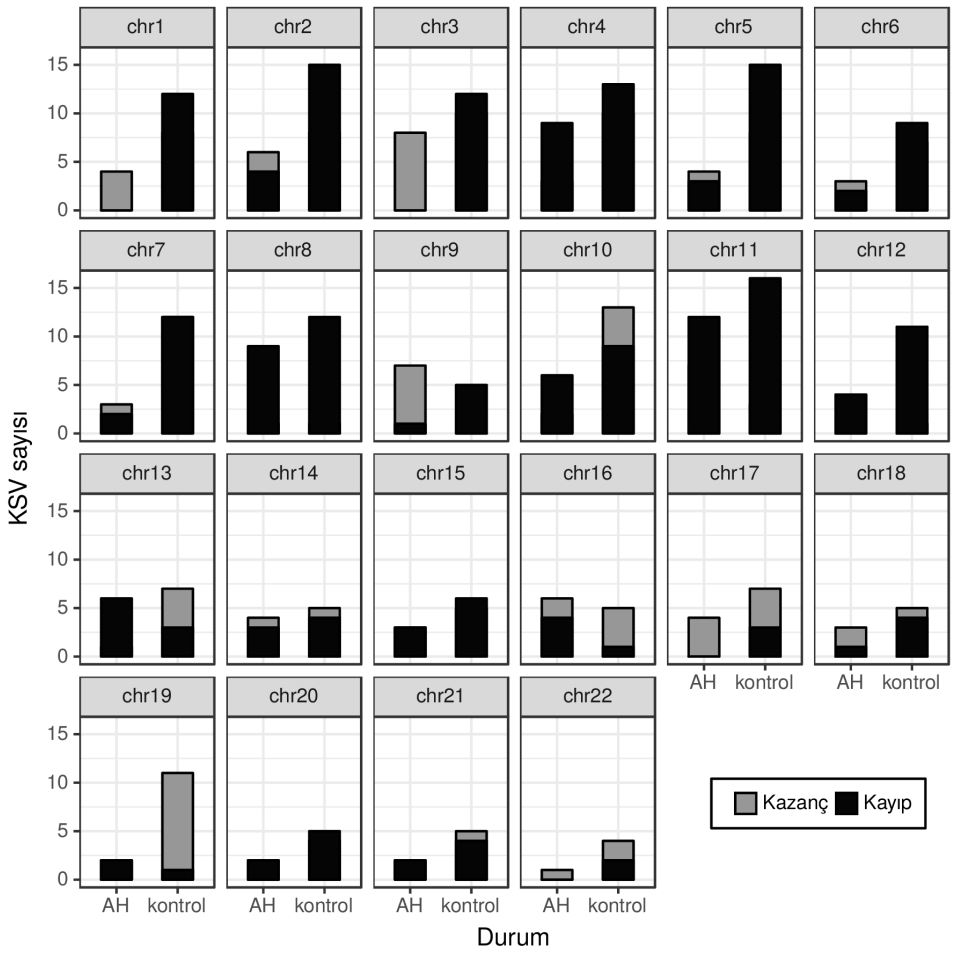
Tablo 5: En az bir anöploidi sinyali gösteren hücrelerin sıklığı.

	Sinyal veren hücre sayısı	Toplam hücre sayısı	Yüzde
AH	10	200	5.0
Kontrol	21	272	7.7
AH + Kontrol	31	472	6.6



Şekil 10: AH ve kontrol örneklerinde kromozom başına tüm kromozom anöploidi olaylarının sayısı.

Toplam 200 AH ve 272 kontrol hücresine ait verilerde kazanç ve kayıpların sayısı sütun grafikleriyle gösterilmiştir.

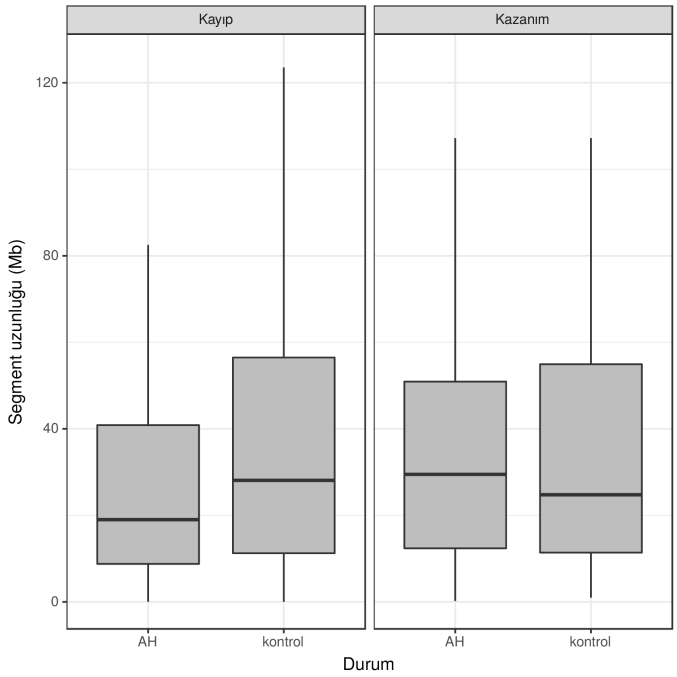


Şekil 11: AH ve kontrol örneklerinde kromozom başına KSV olaylarının sayısı.

Toplam 200 AH ve 272 kontrol hücresine ait verilerde kazanç ve kayıpların sayısı sütun grafikleriyle gösterilmiştir.

3.2.3 KSV'lerin uzunluk dağılımları

472 hücrede KSV olarak tahmin edilen segmentlerin uzunluklarının dağılımı incelenmiştir (Şekil 12). Bu değişken tüm kromozomu kapsayan anöploidi olaylarını diğer bütün KSV olaylarından ayırmak için kullanılmıştır.



Şekil 12: KSV uzunluklarının hastalık durumu ve kayıp / kazanım sinyalleri açısından incelenmesi

3.2.4 Biyolojik faktörlerin KSV ve anöploidi sıklığı üzerinde etkisi

KSV sayısındaki değişkenliğe etki edebileceği düşünülen biyolojik faktörlerin test edilmesi için iki farklı yöntem kullanılmıştır. **Sıfır yığılmalı negatif binom regresyonu** modeli ile 1) hastalık durumu (AH/kontrol), 2) beyin bölgesi, 3) sinyalin alındığı kromozom, 4) DD değişkenlerinin KSV ve anöploidi sıklığı üzerinde etkisi incelenmiştir. Rastgele permütasyon ve bootstrap gibi **parametrik olmayan** istatistiki testler kullanılarak da 1) hastalık durumu ve 2) kayıp ve kazanım sinyali farkı incelenmiştir.

3.2.4.1 Sıfır yığılmalı negatif binom regresyon modeli

Yöntem bölümünde açıklanan bağımlı ve bağımsız değişkenler ile kurulan regresyon modeli, KSV verisini içeren matrisin farklı şekillerde alt gruplara bölünmesiyle birkaç farklı şekilde kullanılmıştır. Verilerin ayrıldığı alt gruplar şunlardır: a) Bulunduğu kromozomda kapsadığı bölgenin uzunluğundan bağımsız olarak bütün KSV sinyalleri, b) Yalnızca bütün kromozom kapsamındaki sinyaller, yani anöploidi sinyalleri. Her iki kategori de daha ileri alt gruplara ayrılarak regresyon modeliyle test edilmiştir: 1) Kayıp veya kazanım olmasına bakılmaksızın bütün KSV sinyalleri, 2) Yalnızca kayıp sinyalleri, 3) Yalnızca kazanım sinyalleri. Sonuç

olarak ilgili model, 6 farklı sinyal türünü seçilen bağımlı değişkenlerin açıklaması açısından test etmekte kullanılmıştır.

3.2.4.1.1 Bütün KSV'leri açıklayan faktörler

3.2.4.1.1.1 Yalnızca kayıp sinyallerinin değerlendirilmesi

Bu KSV kategorisi için, AH-kontrol farkı, beyin bölgesi veya KSV sinyalinin alındığı kromozom faktörleri için farklar anlamlı değildir (AH-kontrol dahil üç faktörün bütün grupları için $p = 1.0$). Verideki varyasyonu anlamlı olarak açıklayan tek bağımsız değişken DD olmuştur ($p < 1e\text{-}14$). DD'nin etkisi pozitiftir, yani gürültü arttıkça KSV'lerin arttığı görülmektedir.

3.2.4.1.1.2 Yalnızca kazanım sinyallerinin değerlendirilmesi

Bu KSV kategorisi için, AH-kontrol farkı, beyin bölgesi veya KSV sinyalinin alındığı kromozom faktörleri için farklar anlamlı değildir (AH-kontrol dahil üç faktörün bütün grupları için $p = 0.999$). Verideki varyasyonu anlamlı olarak açıklayan tek bağımsız değişken DD olmuştur ($p < 0.001$). DD'nin etkisi pozitiftir, yani gürültü arttıkça KSV'lerin arttığı görülmektedir.

3.2.4.1.1.3 Kayıp ve kazanım sinyallerinin birlikte değerlendirilmesi

Kayıp ve kazanım sinyalleri birlikte değerlendirilip her hücrenin bütün kromozomlarında bağımlı değişken olarak bu iki sinyalin sayılarının toplamı kullanıldığında, yukarıdaki testlerle tutarlı olarak yalnızca DD'nin KSV sayısındaki değişkenliği anlamlı olarak açıklayabildiği görülmüştür ($p < 0.001$). DD'nin etkisi yine pozitiftir, yani gürültü arttıkça KSV'lerin arttığı görülmektedir. DD dışındaki diğer faktörlerin (AH-kontrol dahil) alt gruplarına ait p-değerleri yine istatistiki olarak anlamlı değildir ($0.1 < p < 0.99$).

3.2.4.1.2 Tüm kromozom anöploidi sinyallerini açıklayan faktörler

3.2.4.1.2.1. Yalnızca kayıp ve yalnızca kazanım sinyallerinin değerlendirilmesi

Belirtilen bağımsız değişkenler ile mevcut regresyon modeli kullanılamamıştır. Vaka sayısının çok az ve tahmin edilmeye çalışılan parametre değerlerinin yüksek olduğu bu iki durumda, parametre değerlerini tahminde kullanılan döngünün yakınsayamadığı görülmüştür.

3.2.4.1.2.2 Kayıp ve kazanım sinyallerinin birlikte değerlendirilmesi

Tüm kromozom anöploidi sinyalleri içerisinde kayıp veya kazanım olduğuna bakılmaksızın toplam sinyaller bağımlı değişken olarak modeld kullanıldığında, AH-kontrol dahil bağımsız değişkenler arasından yalnızca DD'nin anlamlı bir tahmin edici olduğu görülmüştür ($p < 0.001$). DD'nin etkisi pozitiftir, yani gürültü arttıkça KSV'lerin arttığı görülmektedir.

3.2.4.2 Parametrik olmayan testler

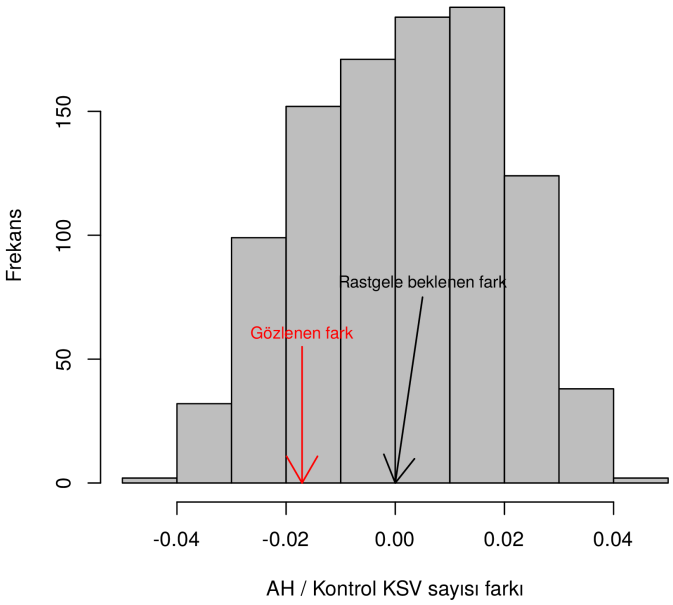
3.2.4.2.1 Bütün KSV'ler ile yapılan testler

3.2.4.2.1.1 Kayıp ve kazanım KSV olaylarının sıklıklarının karşılaştırılması

Literatürde rapor edilene benzer biçimde (bkz. 1. Bölüm) çalışmamızda incelenen 472 hücre arasında da **kayıpların oranı kazanımlara kıyasla daha yüksektir** (**Tablo 4**). Kayıp ve kazanım olaylarının sıklıklarının göreli karşılaştırılması için bootstrap yöntemi kullanılmış, 1000 iterasyonda bireylerin ve bunlara ait eşit sayıda hücrenin rastgele seçilmesiyle kayıplar ve kazanımlar arasındaki fark için hesaplanarak bir güven aralığı belirlenmiştir. Bu şekilde hesaplanan %95 güven aralığı kayıp olaylarının görece sıklığının bu veri setinde anlamlı olmadığına işaret etmektedir (alt limit: -0.0101, üst limit: 0.0176).

3.2.4.2.1.2 AH ve kontrol bireyler arasında KSV sıklıklığının karşılaştırılması

472 hücreye ait veriler incelendiğinde, sağlıklı kontrol bireylerle karşılaştırıldığında **AH bireylerde sınırlı miktarda da olsa daha az KSV olayı rapor edildiği** gözlemlenmiştir (**Tablo 4**). Yukarıda rapor edilen regresyon testleri, bu farkın istatistiksel olarak anlamlı olmadığına işaret etmiştir. Çalışmamızın odak konusu olduğu için bu hipotez parametrik olmayan bir yöntemle de yeniden test edilmiştir. Rastgele beklenen KSV sıklığı farkı miktarını, yani sıfır dağılımını temsil etmesi için, bireylerin hastalık durumunu belirten etiketleri tutan değişken 1000 defa rastgele karılmış, bu iki grup arasındaki rastgele beklenen ortalama farkı temsil eden bir dağılım elde edilmiştir (bkz. 2. Bölüm). İki grup arasındaki gerçek farkın bu dağılım ile karşılaştırılmasıyla ampirik p-değeri $p=0.391$ olarak hesaplanmıştır (**Şekil 13**).



Şekil 13: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen KSV olayı sıklığı farkları.

Gri histogram, iki grup arasında rastgele beklenen farkların rastgele permütasyon yaklaşımıyla hesaplanmış dağılımını, yani ampirik sıfır dağılımını göstermektedir. Siyah ok bu dağılımın ortalama değerini göstermektedir. Kırmızı ok ise gerçek farkı göstermektedir.

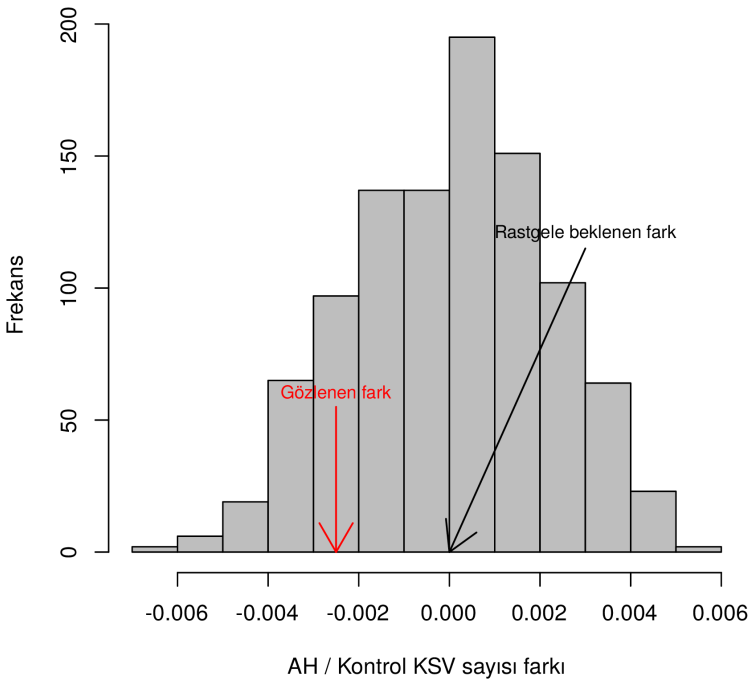
3.2.4.2.2 Tüm kromozom anöploidiler ile yapılan testler

3.2.4.2.2.1 Kayıp ve kazanım anöploidi olaylarının sıklıklarının karşılaştırılması

Kayıp ve kazanım olaylarının sıklıkları bütün KSV olayları ile yapılan testlerde açıklanan şekilde test edildiğinde, iki farklı tür anöploidi olayının göreli sıklıkları arasında anlamlı bir farka rastlanmamıştır (%95 güven aralığı =-0.0047 , 0.0017).

3.2.4.2.2.2 AH ve kontrol bireyler arasında anöploidi sıklıklarının karşılaştırılması

Tablo 5'te görüldüğü gibi incelenen 472 hücre arasında tüm kromozom anöploidi sıklıkları da AH'de kontrol grubuna kıyasla daha düşüktür. Bu farkın anlamlı olup olmadığı yukarıda anlatıldığı şekilde rastgele permütasyon yöntemiyle test edildiğinde, anlamlı bir sonuç bulunmamıştır (p=0.279). Beklenen rastgele sıklık farkı dağılımı ve gözlenen fark arasındaki ilişki **Şekil 14** de görülebilir.



Şekil 14: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen anöploidi olayı sıklığı farkları.

Gri histogram, iki grup arasında rastgele beklenen farkların rastgele permütasyon yaklaşımıyla hesaplanmış dağılımını, yani ampirik sıfır dağılımını göstermektedir. Siyah ok bu dağılımın ortalama değerini göstermektedir. Kırmızı ok ise gerçek farkı göstermektedir.

3.2.5 Teknik gürültünün anöploidi tahminleri üzerinde etkisi

Regresyon modeli ile yapılan testler sonucunda tutarlı olarak görüldüğü üzere KSV sayısındaki değişkenlik üzerinde etkisi test edilen faktörler arasında, dizilenen hücrelerin **değişkenlik değeri** çok anlamlı bir p-değeriyle yegane başarılı tahmin edici olmuştur. Eldeki veri setinde AH-kontrol farkı, birey veya beyin bölgesinin bir etkisi gözlemlenmemiştir.

Bu sonuç üzerine DD ile hücrelerden alınan a) toplam KSV sinyali, b) yalnızca tüm kromozom anöploidi olayları incelenmiştir. Bu inceleme, alınan KSV sinyallerinin büyük çoğunluğunun değişkenlik değeri yüksek hücrelerden geldiğini göstermektedir. Önceki aşamalarda örnek sayısını dramatik olarak düşürmemek adına aşırı sıkı halde olmayan DD üst eşiği / gürültü filtresi **0.5** değerinden **0.45** değerine indirildiğinde, hücreler arasındaki toplam KSV ve tüm kromozom anöploidi olaylarının dağılım ve miktarının (**Tablo 6** ve **Tablo 7**) literatürde şimdiye kadar rapor edilen sıklık değerleriye daha tutarlı olduğu görülmektedir. Bu verilerin, yukarıda rapor edilen verilere kıyasla daha az sayıda sahte pozitif içerdikleri

tahmin edilebilir (ama potansiyel olarak daha yüksek sayıda sahte negatif de içeriyor olabilirler).

Öte yandan gürültü filtresinin sıkılaştırılması sonrasında veriye (sinyalin aşırı seyrelmesi ve algoritmanın yakınsayamaması nedeniyle) regresyon modeli uygulanamamıştır. Aşağıda detaylandırılacağı üzere, verinin bu alt grubu için yalnızca parametrik olmayan test sonuçları rapor edilmiştir.

Tablo 6: Değişkenlik değeri üst eşiği 0.45 olacak şekilde verinin yeniden filtrelenmesi sonrası en az bir KSV gösteren hücrelerin dağılımı

	Sinyal veren hücre sayısı	Toplam Hücre sayısı	Yüzde
AH	10	102	9.8
Kontrol	22	173	12.7
AH + Kontrol	32	275	11.6

Tablo 7: Değişkenlik değeri üst eşiği 0.45 olacak şekilde verinin yeniden filtrelenmesi sonrası en az bir tüm kromozom anöploidi gösteren hücrelerin dağılımı

	Sinyal veren hücre sayısı	Toplam Hücre sayısı	Yüzde
AH	0	102	0
Kontrol	4	173	2.3
AH + Kontrol	4	275	1.5

3.2.5.1 Gürültü filtresi sonrası bağımlı değişkenlerin test edilmesi

Gürültü filtresinin daha sıkı hale getirilmesini takiben, kısım 3.2.4.2 de anlatılan parametrik olmayan testler aynı parametreler ve sırayla yeni filtrelenmiş bu veriye de uygulanmıştır.

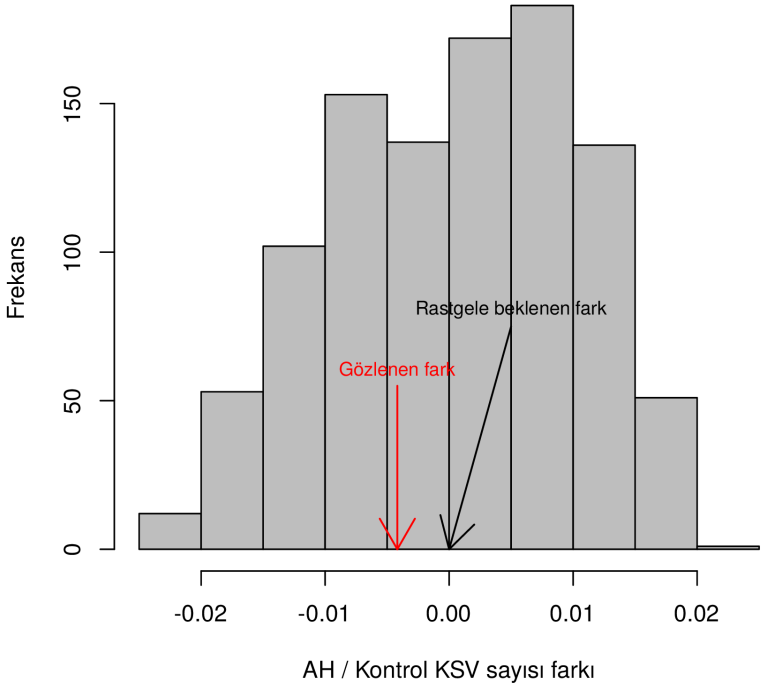
3.2.5.1.1 Bütün KSV'ler ile yapılan testler

3.2.5.1.1.1 Kayıp ve kazanım KSV olaylarının sıklıklarının karşılaştırılması

Uzunluğuna bakılmaksızın bütün KSV olayları ele alındığında, kayıp ve kazanım olaylarının sıklığı arasında anlamlı bir farka rastlanmamıştır.

3.2.5.1.1.2 AH ve kontrol bireyler arasında KSV sıklıklığının karşılaştırılması

Bütün KSV olaylarının sayılarında AH ve kontrol bireylere ait hücrelerden gelmeleri açısından anlamlı bir fark görülmemiştir (p = 0.752). Beklenen rastgele sıklık farkı dağılımı ve gözlenen fark arasındaki ilişki **Şekil 15**'te görülebilir.



Şekil 15: AH ve kontrol grupları arasında gözlenen ve rastgele beklenen KSV olayı sıklığı farkları (DD<0.45)

Gri histogram, iki grup arasında rastgele beklenen farkların rastgele permütasyon yaklaşımıyla hesaplanmış dağılımını, yani ampirik sıfır dağılımını göstermektedir. Siyah ok bu dağılımın ortalama değerini göstermektedir. Kırmızı ok ise gerçek farkı göstermektedir.

4. TARTIŞMA VE SONUÇ

4.1 Bulguların özeti ve değerlendirmesi

Alzheimer Hastalığı hem yaşlı nüfus içinde sıklığı, hem de karmaşık ve şimdiye dek az anlaşılabilmiş biyolojik arkaplanı nedeniyle son onyıllarda ciddi bir araştırma odağı haline gelmiştir. Bu çalışmanın amacı da AH vakalarında somatik KR-KSV'lerin rolü olabileceği hipotezini test etmektir. AH'de bir somatik mutasyon etkisi olup olmadığı, hastalığın erken teşhisi ve/veya hastalığa karşı etkili tedavi yöntemleri aranması açısından önemli bir sorudur.

Mevcut çalışmamıza kadar bu konuda pozitif sonuç rapor eden çalışmalar geleneksel sitometri ve FISH gibi deneysel gürültüye karşı hassas olduğu bilinen yöntemlerle yapılmıştır (Mosch vd., 2007; Iourov vd., 2009; Arendt vd., 2010; Yurov vd., 2014). Şimdiye dek tekil hücre genom dizileme (scWGS) kullanan yegane çalışma ise (van den Bos vd., 2016) sınırlı sayıda AH ve kontrol bireyin (sırasıyla n=10 ve n=6) nöronlarında anöploidi farkı bulamamış (hatta AH grubunda anlamlı olmayan ölçüde daha düşük sinyal tespit etmiştir). Ayrıca söz konusu çalışma kromozom altı KSV'leri incelememiştir. Dolayısıyla AH'de KR-SKV'nin rolü konusu şu ana kadar belirsiz kalmıştır.

Çalışmamız kapsamında 10 AH ve 7 sağlıklı kontrol bireyden, iki farklı yöntemle toplam 1190 kütüphane oluşturulmuş ve dizilenmiş, bunlar arasından 472 hücreye ait veriler kalite ve gürültü filtrelerini geçebilmiştir. Analizimizin sonuçları birkaç başlıkta özetlenebilir.

4.1.1 Düşük dizileme etkisi

Yayımlanmış veriyle yürütülen simülasyon deneyi, anöploidi tahmi için 0.1X yerine 0.01X derinlikte dizileme verisinin yeterli olduğunu göstermektedir.

4.1.2 LCM ve FACS yöntemlerinin karşılaştırması

Çalışmamızın en özgün yanı nöronlara ait çekirdeklerin iki farklı yöntemle toplanmış olmasıdır. Yayımlanmış çalışmalarda kullanılan standart yöntem FACS iken, bu çalışmada bilgimiz dahilinde ilk defa bu işlem için LCM kullanılmıştır. Bu yöntem, özgün morfolojiye sahip hücrelerin boyanmış doku kesitlerinde tanımlanması ve özel olarak seçilebilmelerine olanak tanımaktadır. Böylece örneğin kortikal dokuda (uzun bağlantılarıyla tanınan) piramidal nöronlar ile (kısa bağlantılar yapan) internöronların ayrıştırılması mümkündür. Farklı nöron tiplerinin AH gelişiminde farklı roller üstlenebileceğinden hareketle, beraber çalıştığımız Leipzig ekibi LCM kullanarak piramidal nöronlara odaklanmaya karar vermiştir.

Ayrıca bir tür iç kontrol olarak incelenen bireylerden birine ait doku örneklerinden nöronlar FACS yöntemiyle de ayrıştırılmıştır.

Elde edilen verilerin analizi ve değişkenlik değeri (DD) adlı gürültü ölçüsünün iki yöntemle toplanan hücreler arasında karşılaştırılması, LCM ile toplanan hücrelerin FACS ile toplananlardan anlamlı ölçüde daha gürültülü olduğunu göstermiştir. Bu durum, çalışmamızdaki hücrelerde genel olarak literatürde yayımlanan rakamlara göre DD'nin daha yüksek olmasını da açıklamaktadır.

İki yöntem arasındaki farkın sebebi kesin olarak bilinmese de, LCM'de çevre hücrelere ait parçaların kesilen doku parçasına karışması ihtimali mantıklı gözükmektedir. LCM ile toplanan çekirdeklerde mitokondriyal DNA oranının FACS ile toplananlara kıyasla yüksek olduğu yönündeki gözlemimiz (bu bulgu sunulmamıştır) bu tahmini desteklemektedir, zira çekirdek içinde mitokondriyal DNA beklenmemektedir.

Bu sonuç aslında deney dizaynında bir soruna da işaret etmektedir. Eğer iki yöntemin karşılaştırıldığı dizileme deneyleri, pilot deneylerin parçası olarak ilk aşamada gerçekleştirilmiş olsaydı, deneylerde ağırlıkla FACS kullanılabilir ve toplamda daha temiz veri üretilebilirdi. Ancak işbirliği yaptığımız Alman grup deneyde LCM kullanımı konusunda ısrarcı olmuş ve prokolü gözden geçirme doğrultusundaki önerilerimizi kabul etmemiştir. LCM'in kendisi dışında deneysel çalışmalarda kullanılan DNA izolasyonu veya PCR ile DNA yükseltgeme protokolleri literatürdekilerle benzerdir. Bu işlemlerin verilerde ek gürültü yaratmış olmasını beklemiyoruz.

Yine de özellikle PCR'nin gürültü yaratma potansiyelinin yüksek olduğu not edilmelidir. DNA izolasyon ve PCR koşullarında ufak farklar, veride gürültüyü artırabilir.

4.1.3 KSV ve anöploidi sıklıkları

Analizimizin sonraki aşamalarında, dizilenen kütüphanelerin literatürdekinden daha yüksek gürültülü olduğu hesaba katılarak, gürültü değeri bir teknik faktör olarak analize katılmıştır.

İlk olarak en geniş DD eşiğini geçen, 10 AH ve 7 kontrol bireye ait toplam 472 hücre içinde KSV oranı %21, anöploidi oranı ise %7 civarındadır. İlk rakam için literatürde geniş bir aralık mevcuttur, ancak farklı çalışmalar farklı uzunluklarda KSV'lere odaklanabildikleri için, rakamları karşılaştırmak kolay değildir. Yine de Knouse ve arkadaşlarının (2016), burada kullanılana benzer bir hesaplamalı iş hattı kullanarak farklı veri setlerinde %9 ila %17 arasında >3 megabaz boyunda KSV rapor etmişlerdir. Mevcut tahminimiz bu aralığın dışındadır ama çok uzağında değildir. Ayrıca DD<0.45 ek filtresi verilere uygulandığında, kalan 275 hücre içinde KSV oranının **%12'ye** düştüğü görülmektedir. Bu oranın **daha**

gerçekçi olduğu tahmin edilebilir. Söz konusu 32 vaka, tek başına istatistiksel analiz için fazla dardır.

Çalışmamızda tahmin edilen toplam anöploidi oranı (yaklaşık %7) literatürdeki oranların (genelde %1 veya daha düşük) çok üzerindendir. Ancak DD<0.45 ek filtresi ile bu oran **%1.5'a** düşmektedir. Tüm kromozom anöploidi vakalarının DD korele olması için bir sebep olmadığına göre (çünkü DD kromozom içinde hesaplanmaktadır), %7 oranının büyük ölçüde sahte pozitif sonuç içerdiği tahmin edilebilir. Gerçek oran muhtemelen %1.5'a yakın, hatta belki daha düşüktür.

4.1.4 KSV ve anöploidi sıklığı dağılımını etkileyen faktörler

KSV ve anöploidi sıklıklarının dağılımları incelendiğinde, kullanılan her iki DD değeri eşiğinde kontrol bireylerde görülen vaka oranının, AH vakalarına göre >%10 daha fazla olduğu görülmüştür. Öte yandan bu fark ne parametrik ne de parametrik olmayan analizlerde anlamlı bulunmamıştır. Bu sonuç, van den Bos ve arkadaşlarının (2016) rapor ettikleri bulgularla uyumludur.

Kromozom ve beyin bölgesi de KSV ve anöploidi dağılımı üzerinde anlamlı etkiye sahip gözükmemektedir. Yegane anlamlı etki DD olarak bulunmuştur.

Veride hiçbir biyolojik faktörün anlamlı sonuç göstermemesi verilerimizde **teknik gürültünün** yüksek olmasına ve sahte pozitif ve negatiflerin yoğunluğuna bağlanabilir. Ancak van den Bos ve arkadaşlarının (2016) sonuçları ile beraber değerlendirildiğinde şu çıkarımı yapabiliriz: AH'de sağlıklı kontrollere göre KSV / anöploidi sıklığı muhtemelen sitometri ve FISH deneylerinde rapor edilen kadar yüksek değildir; bir ihtimalle **AH'ye özgü bir fark dahi mevcut değildir**.

4.2 Gelecek çalışmalar

Çalışmamız AH'de hücreler arası somatik varyasyonu inceleyen, yeni bir deneysel yöntem ve buna uygun hesaplamalı bir iş hattı kullanan özgün bir analiz olmuştur. Kendi ekibimiz çalışmayı şu şekilde sürdürecektir: Burada sunulan analizlere ek olarak, yayımlanmış tüm nöronal scWGS verilerinin (ör. McConnell vd., 2013; van den Bos vd., 2016; Knouse vd., 2016) de aynı iş hattından geçirilmesi suretiyle bir meta-analiz yürütülecektir. Söz konusu analiz kapsamında (çok daha fazla hücrenin analiz edilmesi mümkün olacağı için), somatik KSV ve anöploidinin kökenlerine dair (ör. kromozomal lokasyon veya eşey hattı mutasyonlarıyla örtüşme) daha fazla biyolojik analiz yapılması

mümkün olacaktır. Bu çalışma en yakın zamanda makaleleştirilecektir. Sonuçlarımız, alanda yürütülecek başka çalışmalar için iki yönden yol açıcı olacaktır. Birincisi, tekil hücre dizileme çalışmaları için, tüm cazibesine karşın LCM yerine FACS daha kullanışlı bir araç olarak gözükmektedir - en azından şu an için. Bu, alan için önemli bir bilgidir. Genel olarak bundan sonraki çalışmalarda **teknik gürültüyü azaltacak** yöntemlerin tercihi önemli görünmektedir. PacBio ve Oxford Nanopore gibi üçüncü nesil dizileme yöntemleri, PCR aşamasını atlamaları sayesinde, bu açıdan faydalı olabilir. İkincisi, bulgularımız ışığında AH'de KSV'nin rolü yerine, epigenetik değişimlere veya **mikro DNA mutasyonlarına odaklanmanın** daha etkili bir yol olacağı görülmektedir. Bu tip çalışmalarda tekil hücre RNA-dizileme deneyleri ve üçüncü nesil dizileme uygulamaları kullanılabilir.

KAYNAKÇA

Andriani, G. A., Vijg, J., Montagna, C. 2017. "Mechanisms and consequences of aneuploidy and chromosome instability in the aging brain". Mechanisms of Ageing and Development, *161*(Pt A), 19-36.

Arendt, T., Brückner, M. K., Mosch, B., Lösche, A. 2010. "Selective Cell Death of Hyperploid Neurons in Alzheimer's Disease". The American Journal of Pathology, *177*(1), 15-20.

Arendt, T., Holzer, M., Stöbe, A., Gärtner, U., Lüth, H.-J., Brückner, M. K., Ueberham, U. 2000. "Activated Mitogenic Signaling Induces a Process of Dedifferentiation in Alzheimer's Disease That Eventually Results in Cell Death". Annals of the New York Academy of Sciences, *920*(1), 249-255.

Arslantaş, D., Ozbabalik, D., Metintaş, S., Ozkan, S., Kalyoncu, C., Ozdemir, G., Arslantas, A. 2009. "Prevalence of dementia and associated risk factors in Middle Anatolia, Turkey". Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia, *16*(11), 1455-1459.

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Erişim 1 Ocak 2017.

Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., Jones, E. 2011. "Alzheimer's disease". Lancet (London, England), *377*(9770), 1019-1031.

Barber, R. C. 2012. "The Genetics of Alzheimer's Disease" Scientifica (Cairo). 2012;2012:246210.

Baslan, T., Kendall, J., Rodgers, L., Cox, H., Riggs, M., Stepansky, A., Hicks, J. 2012. "Genome wide copy number analysis of single cells". Nature Protocols, *7*(6), 1024-1041.

Bishop, N. A., Lu, T., Yankner, B. A. 2010. "Neural mechanisms of ageing and cognitive decline". Nature, *464*(7288), 529-535.

Bolger, A. M., Lohse, M., Usadel, B. 2014. "Trimmomatic: a flexible trimmer for Illumina sequence data". Bioinformatics, *30*(15), 2114-2120.

Bushman, D. M., Kaeser, G. E., Siddoway, B., Westra, J. W., Rivera, R. R., Rehen, S. K., Chun, J. 2015. "Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains". eLife, *4*, e05116.

Cai, X., Evrony, G. D., Lehmann, H. S., Elhosary, P. C., Mehta, B. K., Poduri, A., Walsh, C. A. 2014. "Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain". Cell Reports, *8*(5), 1280-1289.

Faggioli, F., Wang, T., Vijg, J., Montagna, C. 2012. "Chromosome-specific accumulation of aneuploidy in the aging mouse brain". Human Molecular Genetics, *21*(24), 5246-5253.

Fischer, H.-G., Morawski, M., Brückner, M. K., Mittag, A., Tarnok, A., Arendt, T. 2012. "Changes in neuronal DNA content variation in the human brain during aging". Aging Cell, *11*(4), 628-633.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Sibert, J. 2012. "AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models". Optimization Methods and Software, *27*(2), 233-249.

Garvin, T., Aboukhalil, R., Kendall, J., Baslan, T., Atwal, G. S., Hicks, J., Schatz, M. C. 2015. "Interactive analysis and assessment of single-cell copy-number variations". Nature methods, *12*(11), 1058-1060. Gurvit, H., Emre, M., Tinaz, S., Bilgic, B., Hanagasi, H., Sahin, H., Harmanci, H. 2008. "The prevalence of dementia in an urban Turkish population". American Journal of Alzheimer's Disease and Other Dementias, *23*(1), 67-76.

Iourov, I. Y., Vorsanova, S. G., Liehr, T., Yurov, Y. B. 2009. "Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: Differential expression and pathological meaning". Neurobiology of Disease, *34*(2), 212-220.
Jonsson, T., Atwal, J. K., Steinberg, S., Snaedal, J., Jonsson, P. V., Bjornsson, S., Stefansson, K. 2012. "A mutation in *APP* protects against Alzheimer's disease and age-related cognitive decline". Nature, *488*(7409), 96-99.
Knouse, K. A., Wu, J., Amon, A. 2016. "Assessment of megabase-scale somatic copy number variation using single-cell sequencing". Genome Research, *26*(3), 376-384.
Knouse, K. A., Wu, J., Hendricks, A. 2017. "Detection of Copy Number Alterations Using Single Cell Sequencing". Journal of Visualized Experiments: JoVE, (120).
Knouse, K. A., Wu, J., Whittaker, C. A., Amon, A. 2014. "Single cell sequencing reveals low levels of aneuploidy across mammalian tissues". Proceedings of the National Academy of Sciences, *111*(37), 13409-13414.
Lai, D., Ha, G. t.y. "HMMcopy: A package for bias-free copy number estimation and robust CNA detection in tumour samples from WGS HTS data", 14.
Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Amouyel, P. 2013. "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease". Nature Genetics, *45*(12), 1452-1458.
Lee, M.-H., Siddoway, B., Kaeser, G. E., Segota, I., Rivera, R., Romanow, W. J., Chun, J. 2018. "Somatic APP gene recombination in Alzheimer's disease and normal neurons". Nature, *563*(7733), 639.
Li, H. 2011. "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data". Bioinformatics (Oxford, England), *27*(21), 2987-2993.
Li, H., Durbin, R. 2009. "Fast and accurate short read alignment with Burrows-Wheeler transform". Bioinformatics (Oxford, England), *25*(14), 1754-1760.
López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., Kroemer, G. 2013. "The Hallmarks of Aging". Cell, *153*(6), 1194-1217.
Masters, C. L., Bateman, R., Blennow, K., Rowe, C. C., Sperling, R. A., Cummings, J. L. 2015. "Alzheimer's disease". Nature Reviews Disease Primers, *1*, 15056.
Matevossian, A., Akbarian, S. 2008. "Neuronal Nuclei Isolation from Human Postmortem Brain Tissue". Journal of Visualized Experiments : JoVE, (20).
McConnell, M. J., Lindberg, M. R., Brennand, K. J., Piper, J. C., Voet, T., Cowing-Zitron, C., Gage, F. H. 2013. "Mosaic Copy Number Variation in Human Neurons". Science, *342*(6158), 632-637.
Mosch, B., Morawski, M., Mittag, A., Lenz, D., Tarnok, A., Arendt, T. 2007. "Aneuploidy and DNA replication in the normal human brain and Alzheimer's disease". The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, *27*(26), 6859-6867.
Ning, L., Li, Z., Wang, G., Hu, W., Hou, Q., Tong, Y., … He, J. 2015. "Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons". Scientific Reports, *5*, 11415.
Olshen, A. B., Venkatraman, E. S., Lucito, R., Wigler, M. 2004. "Circular binary segmentation for the analysis of array-based DNA copy number data". Biostatistics (Oxford, England), *5*(4), 557-572.
Qiu, C., Kivipelto, M., von Strauss, E. 2009. "Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention". Dialogues in Clinical Neuroscience, *11*(2), 111 128.

Rehen, S. K., Yung, Y. C., McCreight, M. P., Kaushal, D., Yang, A. H., Almeida, B. S. V., Chun, J. 2005. "Constitutional aneuploidy in the normal human brain". The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, *25*(9), 2176-2180.
Rohrback, S., April, C., Kaper, F., Rivera, R. R., Liu, C. S., Siddoway, B., Chun, J. 2018. "Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing". Proceedings of the National Academy of Sciences, *115*(42),10804-10809.
Rosenkrantz, J. L., Carbone, L. 2017. "Investigating somatic aneuploidy in the brain: why we need a new model". Chromosoma, *126*(3), 337-350.
van den Bos, H., Spierings, D. C. J., Taudt, A., Bakker, B., Porubský, D., Falconer, E., Lansdorp, P. M. 2016. "Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons". Genome Biology, *17*(1), 116.
van Dyck, C. H. 2018. "Anti-Amyloid-β Monoclonal Antibodies for Alzheimer's Disease: Pitfalls and Promise". Biological Psychiatry, *83*(4), 311-319.
Vitak, S. A., Torkenczy, K. A., Rosenkrantz, J. L., Fields, A. J., Christiansen, L., Wong, M. H., Adey, A. 2017. "Sequencing thousands of single-cell genomes with combinatorial indexing". Nature Methods, *14*(3), 302-308.
Yurov, Y. B., Vorsanova, S. G., Liehr, T., Kolotii, A. D., Iourov, I. Y. 2014. "X chromosome aneuploidy in the Alzheimer's disease brain". Molecular Cytogenetics, *7*(1), 20.