ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

Sezer İSLAMBEY

PYTHON İLE CRISPR/CAS NÜKLEAZ RNA KILAVUZLU GENOM DÜZENLEME KÜTÜPHANESİ GELİŞTİRİLMESİ

BİYOTEKNOLOJİ ANABİLİM DALI

ADANA-2022

ÖZ

YÜKSEK LİSANS TEZİ

PYTHON İLE CRISPR/CAS NÜKLEAZ RNA KILAVUZLU GENOM DÜZENLEME KÜTÜPHANESİ GELİŞTİRİLMESİ

Sezer İSLAMBEY

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİYOTEKNOLOJİ ANABİLİM DALI

Danışman : Prof. Dr. Zeynel CEBECİ
Yıl : 2022, Sayfa : 103
Jüri : Prof. Dr. Zeynel CEBECİ
: Doç. Dr. B. Devrim ÖZCAN
: Dr. Öğr. Üyesi Erkut TEKELİ

Bu çalışmada Python programlama dili kullanılarak CRISPR/Cas sistemlerinde gRNA'ları tanımlamak için pycrisprcas9 adında bir açık kaynaklı Python kütüphanesi geliştirilmiştir. Geliştirilen yazılımda gRNA verimliliklerini hesaplamak için Doench ve ark (2014) tarafından oluşturulan kural seti 1, Hsu ve ark (2013) tarafından oluşturulan ceza matrisi kulanılmıştır. Geliştirilen kütüphanenin doğruluğu R paketlerinden CRISPRseek paketi (Zhu ve ark., 2014) kullanılarak test edilmiştir. Geliştirilen paket için kullanım rehberleri hazırlanarak GitHub üzerinde *https://github.com/sezerislambey* adresindeki pycrisprcas9 adlı depodan dağıtıma açılmıştır.

Anahtar Kelimeler: Biyoinformatik, Python, CRISPR/Cas9 sistemi, gRNA analizi, Hedef-dışı analiz

ABSTRACT

MSc THESIS

DEVELOPMENT OF CRISPR/CAS NUCLEASE RNA GUIDED GENOME EDITING LIBRARY WITH PYTHON

Sezer İSLAMBEY

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF BIOTECHNOLOGY

Supervisor : Prof. Dr. Zeynel CEBECİ
Year : 2022, Pages: 103
Jury : Prof. Dr. Zeynel CEBECİ
: Assoc. Prof. Dr. B. Devrim ÖZCAN
: Assist. Prof. Dr. Erkut TEKELİ

In this study, an open-source Python library named pycrisprcas9 was developed to identify gRNAs in CRISPR/Cas systems using the Python programming language. In the developed software, rule set 1 created by Doench et al (2014) and penalty matrix created by Hsu et al (2013) were used to calculate gRNA efficiencies. The datasets used to test the accuracy of the developed library were tested using the package in the CRISPRseek package (Zhu et al., 2014), one of the R packages previously developed in R. results were compared. The accuracy of the results has been confirmed in the analysis. The user guides for the library are available on the Github repository that can be accessed at *https://github.com/sezerislambey*

Keywords: Bioinformatics, Python, CRISPR/Cas9 system, gRNA analysis, gRNA identification, Off-target analysis

GENİŞLETİLMİŞ ÖZET

CRISPR-Cas sistemleri, kesim için yabancı DNA dizilerini hedefleyen bakterilerdeki çeşitli RNA-protein kompleksleri ailesidir. Bu komplekslerin türevleri, CRISPR'dan türetilmiş bir kılavuz RNA (gRNA) sekansına ve Cas9 proteininin kaynağına bağlı olarak spesifik hedef sekansları ayırmak üzere tasarlanmıştır. gRNA'ların tasarımı için önemli hususlar, hedef dışı kesimi en aza indirirken istenen hedef bölgede amaçlanan aktiviteyi maksimize etmektir. Mevcut CRISPR-Cas9 türevli RNA güdümlü nükleazların anlaşılmasındaki hızlı ilerlemeler ve yeni RNA güdümlü nükleaz sistemlerinin geliştirilmesi nedeniyle hedefe özgü RNA güdümlü nükleaz sistemlerinin tasarımı için çok sayıda farklı parametreyi barındırabilecek hesaplama araçlarının mevcut olması önemlidir. Belirli bir giriş dizisini hedefleyen gRNA'ları tanımlamak ve seçilen herhangi bir genom içindeki diğer bölgelerde hedef dışı bölünmeyi en aza indirmek için Python programlama dilinde oldukça esnek, açık kaynaklı bir kütüphane olan pycrisprcas9 geliştirilmiştir. Geliştirilen pycrisprcas9, farklı bakteri türlerinden CRISPR-Cas9 sistemleri için ilgi çekici bir diziyi hedefleyen potansiyel gRNA'ları belirleyebilmekte; yayınlanmış veya kullanıcı tarafından sağlanan ağırlık matrislerini ve pozisyona özgü uyumsuzluk ceza puanlarını kullanarak potansiyel hedef dışı diziler için bir bölünme puanı oluşturabilmektedir. Tanımlanmış gRNA'lar, yalnızca artan özgüllük için eşleştirilmiş yönlerde meydana gelenleri ve/veya restriksiyon enzimi bölgeleriyle örtüşenleri içerecek şekilde ayrıca filtrelenebilmektedir. pycrisprcas9 kütüphanesindeki ***Non-TargetAnalysis*** fonksiyonu temel olarak ihtiyaç duyulan tüm CRISPR işlemlerini gerçekleştirip, hedef dışı analiz sonuçlarını

Summary.xls (hedef dizi özetleri), RecutDetails.xls (nükleaz kesim sitesinin üst üste gelen RE kesim siteleri) ve pairedSpacers.xls (potansiyel eşleştirilmiş aralayıcılar) ve Off-targetAnalysis.xls (hedef dışı diziler ve puanlar) olmak üzere dört farklı dosya çıktısı olarak raporlamaktadır.

TEŞEKKÜR

Yüksek lisans tezi konusunu veren, yapıcı ve yönlendirici fikirleri ile bana daima yol gösteren danışman hocam Sayın Prof. Dr. Zeynel CEBECİ' ye sonsuz teşekkürlerimi sunarım.

Yüksek Lisans eğitimim boyunca yardımlarını benden esirgemeyen hocalarım Doç. Dr. B. Devrim ÖZCAN ve Doç. Dr. Uğur SERBESTER'e teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZ I
ABSTRACT II
GENİŞLETİLMİŞ ÖZET III
TEŞEKKÜR V
İÇİNDEKİLER VI
ÇİZELGELER DİZİNİ IX
ŞEKİLLER DİZİNİ X
1. GİRİŞ 1
1.1. CRISPR-Cas Sistemi 3
2. ÖNCEKİ ÇALIŞMALAR 5
2.1. CRISPR-Cas9 Genom Düzenleme Teknolojisi 5
2.2. CRISPR için Biyoinformatik Araçlar 6
2.2.1.CRISPR-Cas İle İlgili Veritabanları 7
2.2.1.1. EENdb: Endonükleaz Mühendisliği İçin Veritabanı 7
2.3.1.2. CrisprGe: CRISPR/Cas Tabanlı Genom Düzenleme Merkezi 8
2.2.1.3. WGE: Genom Mühendisliği İçin Bir CRISPR Veritabanı 10
2.2.1.4. CRISPRdb ve CRISPRCasdb 12
2.2.1.5. CRISPRone 13
2.2.1.6. Anti-CRISPRdb 13
2.2.2. CRISPR-Cas Sistemlerinin Tahmini 13
2.2.2.1. Cas-OFFinder 13
2.2.2.2. SSFinder 15
2.2.2.3. CRISPRseek 17
2.2.2.4. sgRNAcas9 18
2.2.2.5. pgRNAFinder 19
2.2.2.6. CasOT 20

2.2.2.7. CRISPRdirect 20
2.2.2.8. CRISPRFinder 21
2.2.2.9. CRISPRCasFinder 21
2.2.2.10. PILER-CR 22
2.2.2.11. CRISPRDetect 22
2.2.2.12. CRISPy-web 23
2.3. gRNA Tasarımı 24
2.4. Hedef Tanımlama 25
2.5. Hedef Bölgede Verimlilik Tahmin Modelleri 25
2.6. Hedef Dışı Tahmin Algoritmaları 26
3. MATERYAL VE METOT 27
3.1. Materyal 27
3.2. Metot 30
3.2.1. gRNA Tanımlanması 30
3.2.2. Potansiyel Hedef Dışı Dizilerin Belirlenmesi 32
3.2.3. pycrisprcas9 Kütüphanesi 34
3.2.3.1. footnoteOffTargets Fonksiyonu 34
3.2.3.2. CharacteristicVectorForScoring 34
3.2.3.3. computegRNAEfficiency 35
3.2.3.4. strainergRNAs 36
3.2.3.5. strainerOffTarget 37
3.2.3.6. determinegRNAs 38
3.2.3.7. calculateOfftargetScore 41
3.2.3.8. UniquePatterns 41
3.2.3.9. Non-TargetAnalysis 42
3.2.3.10. searchingHits2 47
3.2.3.11. convertPattern 48

3.2.3.12. uniqueRestrictionEnzymes 48
3.2.3.13. writingHits2 49
4. BULGULAR VE TARTIŞMA 51
4.1. pycrisprcas9 Kullanımı 51
4.2. Tartışma 88
5. SONUÇLAR VE ÖNERİLER 91
KAYNAKLAR 93
ÖZGEÇMİŞ……...……103

ÇİZELGELER DİZİNİ

Çizelge 2.1. CRISPR/Cas sistemi tarafından hedeflenen en önemli genlerin listesi.10
Çizelge 3.1. IUPAC nükleotid belirsizlik kodlarının baz karşılığı……28
Çizelge 3.2. Kullanılan kaynak dosyalar……..33
Çizelge 4.1. pycrisprcas9 ve CRISPRseek paketinin karşılaştırılması…….....89

ŞEKİLLER DİZİNİ

Şekil 2.1. EENdb'nin menü başlığı ve veritabanının beş ana bölümü arasındaki ilişki..…….……8
Şekil 2.2. CrisprGE asarımı……...9
Şekil 2.3. Genoverse genom tarayıcısında (üst panel) önceden hesaplanmış CRISPR sitelerinin görüntülenmesi…….……..…11
Şekil 2.4. (A) Cas-OFFinder şeması (B) CasOFFinder'ın iş akışı. (C) CPU (siyah kareler) ve GPU (kırmızı daireler) aracılığıyla giriş hedef sitelerinin sayısının bir fonksiyonu olarak hedef site başına çalışma süresi ...…..15
Şekil 2.5. Windows işletim sisteminde SSFinder kullanımı …….…….….16
Şekil 2.6. SSFinder tarafından verilen çıktı dosyasına genel bakış …….....…..16
Şekil 2.7. CRISPRseek sonuç çıktısı gRNA'ların özeti……..…....17
Şekil 2.8. CRISPRseek sonuç çıktısı hedef dışı bölgelerin detaylı bilgileri……18
Şekil 3.1. Non-TargetAnalysis iş akışı ……..……...……..29
Şekil 4.1. pycrisprcas9 kütüphanesinin indirilmesi ……..……..51

1. GİRİŞ

Kümelenmiş Düzenli Aralıklarla Bölünmüş Kısa Palindromik Tekrar Kümeleri (Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR), tek hücreli canlılarda (mikroorganizma) bulunan kazanılmış bağışıklık sistemidir. Arkeal genomların yaklaşık %90'ı ve bakteri genomlarının %48'i, CRISPR-Cas adı verilen istilacı genetik elementlere karşı adaptif bir savunma sistemi barındırır (Barrangou ve van der Oost, 2013). Bakterilerde ve arkelerde, CRISPR-Cas sistemleri istilacı faj ve plazmit DNA parçalarını (20 baz uzunluğunda) CRISPR lokuslarına dahil ederek homolog dizilerin bozulmasına rehberlik etmek için karşılık gelen CRISPR RNA'ları (crRNA'lar) kullanarak bağışıklık sağlar (Terns ve Terns, 2011). Bakteriler, entegre olan bu yabancı genetik materyali kopyalayıp CRISPR sisteminde kullanılacak kılavuz RNA'yı (gRNA: Guided RNA) sentezler. CRISPR sisteminin diğer parçası olan Cas enzimi gRNA ile birleşip hücre içine nüfuz eden bakteriyofaj DNA'sını keserek etkisiz hale getirir (Barrangou ve van der Oost, 2013).

CRISPR ile ilişkili proteinlerden biri olan Cas9, bir RNA bölümü tarafından hedefe yönlendirilen ve her iki DNA zincirini de kesen bir endonükleazdır. Cas9, sentetik tek kılavuz RNA (sgRNA: Single Guided RNA) adı verilen tek bir zincir halinde sentezlenebilir. RNA'nın genomik DNA'ya bağlanan bölümü 18-20 nükleotitten oluşur. Kesmek için, 2 ile 5 nükleotid arasında özel bir DNA dizisi (tam dizilim Cas9'u üreten bakterilere bağlı) gRNA'nın 3' ucunda yer almalıdır. Buna protospacer bitişik motif (PAM: protospacer adjacent motif) denir. DNA kesilmesinden sonra onarım, homolog olmayan uç birleştirme (NHEJ: non-homolog end joining) veya homolog rekombinasyon (HR) olmak üzere iki yolla gerçekleşebilir (Hsu, Lander ve Zhang, 2014).

Biyoenformatik, CRISPR-Cas sisteminin keşfinde ve yeni fonksiyonların tanımlanmasında kilit rol oynar. CRISPR dizilerinin kendine özgü tekrarlayıcı yapısı, ilk olarak 1987'de Yoshizumi Ishino tarafından tanımlanmıştır (Ishino ve

ark., 1987). Daha sonra, biyoenformatik analizlerin yardımıyla Francisco Mojica, CRISPR dizilerinin sadece *Escherichia coli*'de değil, aynı zamanda çeşitli arkeal ve bakteri genomlarında da bulunduğunu göstermiştir (Mojica ve ark., 2000). Sonraki biyoinformatik analizler, PAM bölgelerinin bakteriyofajlarla eşleşmesini ortaya çıkarmıştır. Bu da CRISPR-Cas sistemlerinin kazanılmış bir bağışıklık sistemi olarak hareket ettiğini doğrulamıştır (Mojica ve ark., 2005; Pourcel ve ark., 2005; Bolotin ve ark., 2005). Shah ve ark (2009), aralayıcı eşleşmeleri üzerine yapmış oldukları bir biyoinformatik çalışmasında, CRISPR-Cas sistemlerinin RNA'dan ziyade öncelikle DNA'yı hedef alacağını başarılı bir şekilde öngörmüştür. Bu nedenle, biyoenformatik, CRISPR sistemlerinin keşfinde ve ayrıca ilk işlevsel hipotezlerin oluşturulmasında çok önemli bir rol oynamıştır.

CRISPR-Cas9 sistemleri için tek veya eşleştirilmiş hedef bölgeleri ayıran ve hedef dışı analizi gerçekleştiren gRNA'ları tanımlayan biyoinformatik araçlara ihtiyaç vardır. Bu tez çalışmasında, deneysel olarak türetilmiş hedef dışı puanlama şemaları ve herhangi bir genom dizisi kullanılarak karakterize bir PAM dizisi ve hedef dışı analiz ile herhangi bir CRISPR-Cas9 sistemi için hedef-spesifik gRNA'ların tasarımını kolaylaştırmak için açık kaynak bir Python kütüphanesinin geliştirilmesi amaçlanmıştır. Ek olarak, gRNA'ları tanımlamak için ilgili iki giriş dizisinde hedef bölgeleri puanlama işlevselliği sağlanmıştır. Geliştirilen yazılım isteğe uygun uyumsuzluk ceza matrisi girişi, PAM sekanslarında varyasyonlar tanımlama, kılavuz sekanslar için spesifik modeller belirleme ve diğer arama kriterlerini değiştirme esnekliği ile zenginleştirilmiştir. Bu nedenle, bu özellikler CRISPR-Cas türevli sistemlerin ek sınıfları karakterize edildikçe ve hedef dışı aktivite hakkında yeni deneysel bilgiler elde edildikçe, hedef bölge arama parametrelerinin değiştirilmesine izin verebilmektedir.

1.1. CRISPR-Cas Sistemi

Birçok prokaryotta bulunan CRISPR, yabancı DNA dizilerini tespit etmek ve ayırmak için uyarlanabilir bir bağışıklık sistemi olarak işlev görecek şekilde Cas ve bir tracrRNA (trans-aktive edici crRNA (crisprRNA)) ile birlikte hareket eden **crRNA**'ları kodlar (Sampson ve ark., 2013; Koonin ve Makarova, 2009; Horvath ve Barrangou, 2010). Cas9 proteinlerinin nükleaz aktivitesini genom içindeki spesifik hedeflere yönlendirmek için CRISPR'den türetilmiş bir gRNA dizisi kullanan CRISPR-Cas sistemlerinin modifiye edilmiş versiyonları geliştirilmiştir (Mali ve ark., 2013). gRNA-yönlendirmeli Cas9 sistemleri, insan kök hücreleri ile model ve model olmayan organizmalarda hedeflenmiş genetik değişiklikler yaratabilir (Hwang ve ark., 2013; Cong ve ark., 2013, Li ve ark., 2013). Bu nükleazlar, homolog olmayan uç birleştirme yoluyla ekleme / silmeler (indeller) veya bir DNA donör molekülü ile homolojiye yönelik onarım yoluyla eklenen spesifik dizi değişiklikleri de dahil olmak üzere çeşitli genom modifikasyonlarına yol açabilen çift zincirli DNA kopmaları oluşturur (Ran ve ark., 2013a). gRNA-Cas9 komplekslerinin diğer yararlı türevleri arasında nikazlar, sadece bir DNA zincirini ayırarak eşleştirilmiş konfigürasyonlardaki spesifiteyi arttırmak için kullanılabilir ve herhangi bir DNA bölünme aktivitesine sahip olmayan, ancak ek efektör etki alanlarını toplayarak gen transkripsiyonunu artırabilir veya azaltabilir (Mali ve ark., 2013a; Ran ve ark., 2013a).

gRNA-Cas9 kompleksleri, iki bileşenden oluşan spesifik hedef sekansları tanır. Kılavuz sekans tipik olarak gRNA'daki değişken bir bölge ile aynıdır. Kılavuz dizisinin doğrudan bitişiğinde, Cas9 proteini tarafından tanınan kısa bir dizi olan PAM dizisi bulunur (Nishimasu ve ark., 2014). *S. pyogenes*'ten türetilmiş en yaygın olarak kullanılan CRISPR-Cas9 sisteminde (Mali ve ark., 2013a; Ran ve ark., 2013a), kılavuz sekans tipik olarak 20 bazlık değişken bir bölge olup tercih edilen PAM sekansı ise NGG'dir (PAM sekansı = NAG olduğunda düşük aktiviteye sahiptir). CRISPR-Cas9 sistemleri için potansiyel bir sınırlama, gRNA tarafından

hedeflenen diziyle tam olarak eşleşmeyen dizilerde bölünebilmeleridir (Cho ve ark., 2014; Fu ve ark., 2013; Hsu ve ark., 2013; Mali ve ark., 2013b). Bu nedenle, CRISPR-Cas9 sistemlerinin tasarımı için önemli bir husus, düşük hedef dışı bölünme oranlarına sahip gRNA'ların seçilmesidir. İlk çalışmalar sadece PAM sekansına en yakın kılavuz sekansındaki 12 bazın özgüllüğe katkıda bulunduğunu düşündürmekteydi (Cong ve ark., 2013). Hsu ve ark (2013) tarafından gerçekleştirilen bir analizde, hedef dışı bölgelerin bulundukları yerlere, uyuşmazlıkların sayısı ve düzenlerine göre hedef bölünme olasılığını kapsamlı deneysel veriler kullanarak modellemiştir. Hedef dışı bölünme ile gRNA'ları seçmek adına http://crispr.mit.edu adresinde bir web uygulaması olarak kullanılan bir uyumsuzluk ceza matrisi ve puanlama şeması türetilmiştir (Ran ve ark., 2013a). Hedef dışı oranlar, eşleştirilmiş Cas9 nikazları (Cho ve ark., 2014; Mali ve ark., 2013b; Ran ve ark., 2013b) veya FokI nükleazına kaynaşmış ölü bir Cas9 nükleazının dimerleri (Tsai ve ark., 2014) gibi iki gRNA-Cas9 kompleksinin bağlanmasını gerektiren stratejiler kullanılarak daha da azaltılabilir.

2. ÖNCEKİ ÇALIŞMALAR

2.1. CRISPR-Cas9 Genom Düzenleme Teknolojisi

En önde gelen üç genom düzenleme teknolojisi, Transkripsiyon Aktivatör Benzeri Efektör Nükleazlar (TALEN: Transcription Activator-like Effector Nucleases), Çinko Parmak Nükleazları (ZFN: Zinc Finger Nucleases) ve CRISPR-Cas9, genomun belirli bir bölgesini kesin ve öngörülebilir bir şekilde değiştirmek için programlanabilir nükleazları kullanır. TALEN'ler ve ZFN'ler protein bazlı sisteme bağlıyken CRISPR-Cas9 teknolojisi RNA güdümlü nükleazları kullanır.

Genetik mühendisliği açısından bakıldığında, *Streptococcus pyogenes* bakterisinden uyarlanan Tip II CRISPR-Cas9 sistemi kimerik sgRNA ve Cas9 endonükleaz olmak üzere iki bölümden oluşur. sgRNA, Cas9 ile fonksiyonel bir kompleks oluşturur ve kullanılan Cas9 proteininin tipine göre değişen PAM dizisini bularak hedef DNA dizisini parçalanması için nükleaza rehberlik eder. En yaygın olarak kullanılan Cas9, PAM 5'-NGG-3' özellikli SpCas9'dur. Cas9 proteini, diziyi yalnızca PAM varlığında parçalar. *In vitro* veya *in vivo* yapılan sgRNA, trans-aktive edici crRNA (tracrRNA) dizisine kaynaşmış özel olarak tasarlanmış CRISPR RNA (crRNA) dizisinden oluşur. crRNA dizisi, gRNA ayırıcı olarak bilinen 5'-uçlu 20 nükleotitten (nt) oluşan değişken bir kısımdır ve farklı DNA bölgelerini hedeflemek üzere programlanabilen hedef DNA dizisinin tamamlayıcısıdır. tracrRNA, sgRNA'nın sabit bir parçasıdır ve birkaç kök döngü oluşturur. Cas9 bağlama, crRNA işleme ve Cas9 aracılı hedef bölünme için gereklidir (Jiang ve Doudna 2017).

CRISPR-Cas9 gen düzenleme deneyinin tasarımı, gRNA ayırıcı dizisinin seçilmesiyle başlar. CRISPR-Cas9-sgRNA sisteminin kritik bir adımı, sgRNA'nın optimum tasarımıdır. PAM sekans bölgelerini tarayarak genomda hedef bölgeleri bulmak basit olsa da, etkinlik ve özgüllük bağlamında gen düzenlemenin kesinliğini belirleyen çeşitli zorluklar mevcuttur. Burada etkinlik, belirli sgRNA'nın hassas kesmeyi kolaylaştırma olasılığı olarak tanımlanır ve özgüllük, sgRNA'nın hedef dışı

bölgelere bağlanma olasılığı olarak tanımlanır. Cas9, farklı hedef bölgelerde değişken verimlilik sergiler. Birkaç baz uyuşmazlığını ve DNA/RNA çıkıntılarını kabul eder (Lee ve ark., 2016). Bu nedenle hedeflenen etkinliği en üst düzeye çıkarmak ve potansiyel hedef dışı etkileri en aza indirmek için tahmin modelleri/algoritmaları ve hesaplama araçları geliştirmeye ihtiyaç vardır (Doench ve ark., 2016).

En iyi hedef bölgelerin seçiminde yardımcı olan sgRNA tasarımı için hesaplama araçlarının geliştirilmesiyle sonuçlanan modeller oluşturulmuştur (Parkhi ve ark., 2018; Jacob ve Gregory, 2016). Gen regülasyonu, genom çapında özel sgRNA kitaplıkları yapımı ve özel gen setleri için sgRNA tasarımını tek bir adımda kolaylaştırmak adına hesaplama araçları geliştirilmiştir. Ayrıca CRISPR kaynaklı mutasyonları ve mutasyon modellerini tespit etmek ve genom düzenleme sonuçlarını değerlendirmek için NGS platformları kullanılmaktadır. CRISPR tarafından tanıtılan mutasyonları tanımlamak, NGS veri analizleri ve sonuçların yorumlanabilmesi için biyoinformatik araçlar geliştirilmiştir.

2.2. CRISPR için Biyoinformatik Araçlar

Biyoinformatik araçlar gRNA'nın tasarımı ve deney sonrası kalite kontrolünün yapılmasını sağlayan yazılımlardır. Ayrıca gRNA tasarımında daha önce hangi şartlarda çalıştığını görme ve izleme konusunda veritabanlarına da ihtiyaç vardır. Bunlar arasında, 2013'te yayınlanan EENdb (Engineered Endonuclease Database), 2015'te yayınlanan CrisprGE (CRISPR Genom Editing) ve WGE (Wellcome Trust Sanger Institute Genome Editing Database) yaygın kullanılmakla birlikte birçok biyoenformatik araç geliştirilmiştir.

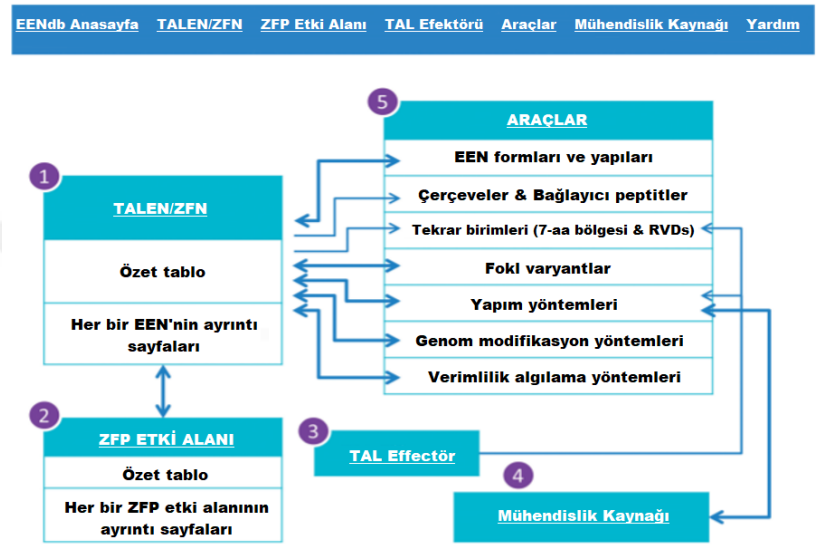
2.2.1. CRISPR-Cas İle İlgili Veritabanları

2.2.1.1. EENdb: Endonükleaz Mühendisliği İçin Veritabanı

Tasarlanmış endonükleazlar (EEN: Engineered endonuclease), belirli DNA dizilerini *in vitro* veya *in vivo* olarak bağlamak ve parçalamak için tasarlanmıştır. ZFN'ler ve transkripsiyon TALEN'ler, ortak birçok özellik sergileyen en yaygın kullanılan iki özelleştirilebilir EEN türüdür (Baker, 2012). Hedef dizileri, işlenebilir DNA bağlama alanları aracılığıyla tanırlar ve DNA'yı spesifik olmayan nükleaz alanları ile keserler (Kim ve ark., 1996; Christian ve ark., 2010). Bu yapay endonükleazlarla, sahaya özgü DNA çift sarmallı kırılmalar herhangi bir belirli genomda teorik olarak indüklenebilir ve bu sayede hedeflenen genom modifikasyonu kolayca elde edilebilir. Çoğu ZFN ve TALEN, dimer olarak işlev görür. Bu EEN'lerin her monomeri, ZFN'ler için çinko parmak proteinlerinden (ZFP'ler) veya TALEN'ler için transkripsiyon aktivatörü benzeri (TAL) efektörlerinden (TALE'ler) türetilen spesifik bir DNA bağlama alanı içeren yapay olarak oluşturulmuş bir hibrit proteinden oluşur (Berg, 1990; Scholze ve Boch., 2010).

EENdb birbiriyle ilişkili beş bölümden oluşur (Şekil 2.1). Ana bölüm EEN'ler hakkında bilgi toplanması ve özetlenmesidir. "ZFP Etki Alanı" ve "TAL Etkileyici", ana bölümü tamamlamak üzere oluşturulmuş iki bölümdür. "ZFP Etki Alanı", ZFN'lerden ve diğer ZFP'lerden 7-aa bölgelerinin toplanmasını sağlar. 'TAL Etkileyici', TALEN'ler dışındaki TALE proteinlerinden TALE DNA bağlama alanlarının bilgilerini toplar. EEN'ler hakkında temel bilgilerin toplanması, EENdb'nin ana bileşenidir. Kullanıcılar, özel bir türü hedeflemek için tasarlanmış tüm TALEN'leri veya ZFN'leri listeleyebilir veya 'TALEN/ZFN' sayfasından gen sembolleri, gen kimlikleri, tam veya kısmi hedef dizileri veya referans PMID'leri veya ilk yazarların soyadı ile kolayca arama yapabilir (Şekil 2.1). Karşılık gelen ZFN'lerin, TALEN'lerin veya iki EEN türünün entegrasyonunun sonuçları, kullanıcılar tarafından belirlenen arama seçeneğine göre döndürülür.

EENdb, açık kaynak teknolojileri ile uygulanmaktadır. Veriler bir MySQL ilişkisel veritabanında saklanır. Web sitesinde kodlama PHP dilinde yaplmıştır ve hizmet bir Apache sunucu üzerinden sağlanmaktadır (Xiao ve ark., 2012).

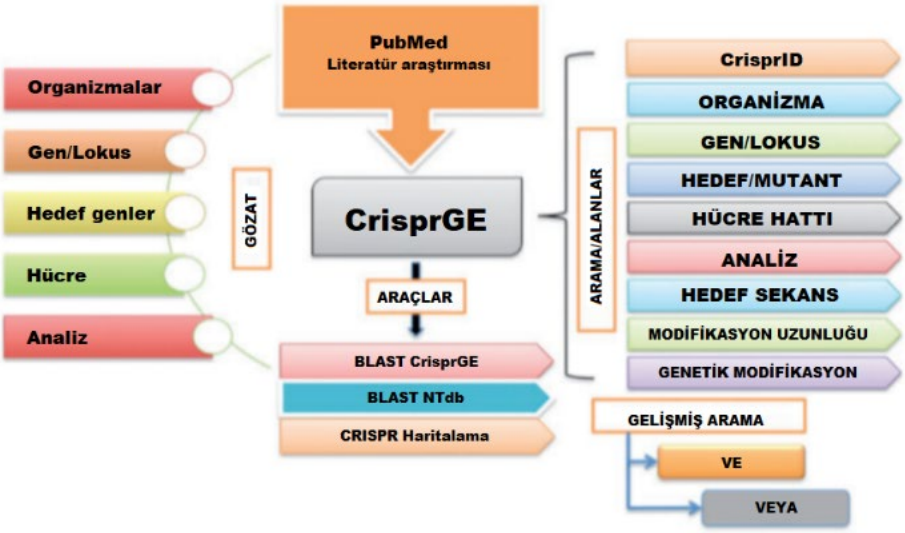


Şekil 2.1. EENdb genel mimarisi (Xiao ve ark., 2012'ten değiştirilmiştir)

CrisprGE, CRISPR/Cas sistemi için düzenlenmiş toplam 4680 gene sahip özel bir veritabanıdır. Toplam 32 modelde ve diğer organizmalarda hedeflenen 223 benzersiz (unique) genin yanı sıra onarım mekanizmaları tarafından indüklenen farklı modifikasyonlardan oluşur. Veritabanında rapor edilen deneyler, farklı hücre hatları üzerinde gerçekleştirilmiştir. Bu veritabanında, CRISPR/cas yöntemiyle en az 70 kez hedeflenen ilk 20 gen Çizelge 2.1'de gösterilmektedir.

CrisprGE veritabanı, kolay arama ve tarama seçenekleriyle donatılmıştır. BLAST, CrisprGE, BLAST NTdb ve CRISPR mapper gibi analiz araçlarımnı da içerir (Şekil 2.2). Bireysel girişler, UniProt, KEGG ve PubMed, vb. gibi diğer kaynaklara köprüyle bağlanır. Web arayüzünün uygulanması CrisprGE, Red Hat

Enterprise Linux 5'te MySQL ve Apache arka uçta açık kaynaklı LAMP sunucusu kullanılarak oluşturulur. Ön uç PHP ile uygulanır (Kaur ve ark., 2015).



Şekil 2.2. CrisprGE mimarisi (Kaur ve ark., 2015'ten değiştirilmiştir)

Çizelge 2.1. CRISPR/Cas sistemi tarafından hedeflenen en önemli genlerin listesi (Kaur ve ark., 2015)

Organizma	***Gen***	***Girdi sayısı***
Mus musculus, Rattus norvegicus, Xenopus tropicalis, Danio rerio	*Tyr*	*252*
Arabidopsis thaliana, Nicotiana benthamiana	*ADH1*	*238*
Nicotiana tabacum, Nicotiana benthamiana, Oryza sativa, citrus sinensis	*PDS*	*155*
Rattus norvegicus, Mus musculus	*Prkdc*	*125*
Rattus norvegicus, Mus musculus	*Tet1*	*118*
Arabidopsis thaliana	*TT4*	*108*
Rattus norvegicus, Mus musculus	*B2m*	*95*
Oryza sativa	*YSA*	*92*
Rattus norvegicus, Mus musculus	*Tet2*	*88*
Glycine max	*DDM1*	*87*
Homo sapiens	*CCR5*	*86*
Mus musculus	*PCSK9*	*81*
Homo sapiens	*DMD*	*80*
Danio rerio	*fh*	*72*
Homo sapiens	*Pcdh*	*72*
Homo sapiens	*HBB*	*70*
Rattus norvegicus, Danio rerio	*ApoE*	*69*
Rattus norvegicus, Mus musculus	*Tet3*	*68*
Rattus norvegicus, Mus musculus	*Prf1*	*67*
Arabidopsis thaliana, Nicotiana benthamiana	*PDS3*	*66*

2.2.1.3. WGE: Genom Mühendisliği İçin Bir CRISPR Veritabanı

WGE veritabanı, CRISPR bölgelerini hesaplamak, görselleştirmek ve seçmek için Wellcome Trust Sanger Enstitüsü tarafından genom düzenleme veritabanını olarak oluşturulmuştur (Hodgkins ve ark., 2015). WGE veritabanı şu anda tek ve eşleştirilmiş CRISPR bölgelerini ve fare ve insan ekzomlarında bulunan CRISPR'ler için önceden hesaplanmış hedef dışı bilgileri depolamaktadır. Hedef dışı bölgelerin puanlanması ve görüntülenmesi basit ve sezgisel çalışmaktadır. Yüksek kaliteli CRISPR bölgelerini hızlı bir şekilde belirlemek için filtreler uygulanabilmektedir. WGE veritabanı ayrıca gen modelleri, protein translasyonu ve varyasyon izleri ile birlikte aynı genom tarayıcısında gen hedefleme vektörlerinin tasarımı ve gösterimi için bir araç sağlamaktadır. WGE açık, genişletilebilir ve

herhangi bir genom için CRISPR bölgelerini hesaplamak ve sunmak üzere ayarlanabilmektedir.

Tüm CRISPR bölgeleri, önceden hesaplanmış ve kullanıcı tarafından talep edilen hedef dışı puanların yanı sıra kullanıcı tarafından oluşturulan vektör tasarımları WGE veritabanında depolanarak CRISPR konumlarının ve hedef dışı bilgilerin toplu olarak indirilmesi, WGE web sitesi aracılığıyla veya bir REST API ile doğrudan WGE veritabanını sorgulayarak elde edilebilmektedir.

WGE'de CRISPR-bulma bölümüne, bir türü (şu anda insan veya fare) ve incelenecek genin işaretçi sembolü seçilerek bir hedef Ensembl ekson seçilmesi istenir. Tüm olası CRISPR siteleri ve eşleştirilmiş siteler, Ensembl'den (Flicek ve ark., 2014) mevcut gen modellerini, protein çevirisini, mevcut varyasyonu ve kullanıcı tarafından oluşturulan herhangi bir hedefleme vektör tasarımını içeren kaydırılabilir bir Genoverse genom tarayıcı görünümünde gösterilir. Eşleştirilmiş CRISPR bölgeleri (Shen ve ark., 2014), karşıt dizilerdeki CRISPR siteleri 30 bç'ne kadar bir ayrılmaya veya 10 bç'ne kadar bir örtüşmeye sahip olduğunda gösterilir (Şekil 2.3) (Hodgkins ve ark., 2015).

Şekil 2.3. Genoverse genom tarayıcısında (üst panel) önceden hesaplanmış CRISPR bölgelerinin görüntülenmesi (Hodgkins ve ark., 2015)

Şekil 2.3'te insan APOE geninin bir bölgesi, açıklamalı gen modelinin altında CRISPR siteleri (mavi renkte PAM bölgesi olan yeşil çubuklar) ile gösterilmektedir. Tek tek CRISPR sitelerine tıklamak, hedef dışı bilgileri gösteren bir açılır pencere döndürür (alt panel). Orijinal CRISPR sitesi için genomik bilgiler mavi renktedir. Hedef dışı dizideki uyumsuzluklar kırmızı renktedir (Hodgkins ve ark., 2015).

2.2.1.4. CRISPRdb ve CRISPRCasdb

CRISPRdb arke ve bakteri genomlarından CRISPR dizilerini içeren açık erişimli bir veritabanıdır. Toplam 232 arkeden 870 CRISPR dizisini ve 6782 bakteri genomundan 8069 CRISPR dizisini kapsar. CRISPRdb, varsayılan CRISPR'leri tanımlamak için CRISPRFinder adı verilen bir arama aracını kullanarak NCBI Referans Dizi Veritabanı'ndan yeni yayınlanan genom dizilerini yansıtmak için aylık olarak otomatik olarak güncellenir (Grissa ve ark., 2007b). CRISPRdb, (i) tür bilgilerini içeren taksonomi; (ii) genomik dizi (kromozom veya plazmit); (iii) girişin kimlik, başlangıç ve bitiş konumları, oluşturma tarihi ve değişiklik tarihini sağlayan CRISPR yeri; (iv) doğrudan tekrarların ID, uzunluk ve konsensüs dizisi; (v) aralayıcıların kimlikleri, uzunlukları ve dizileri hakkında bilgileri içerir. Kullanıcılar, CRISPR dizilerinden oluşan kendi özel veritabanlarını oluşturabilir ve ardından bunları BLAST tarafından genel verilerle karşılaştırabilir (Grissa ve ark., 2007b).

Genişletilmiş CRISPR-Cas tahmin aracı kullanılarak CRISPRCasFinder'ın CRISPRCasdb adlı yeni bir sürümü kullanıma sunulmuştur. CRISPRCasdb, 240 arkeal ve 9242 bakteri genomu için CRISPR dizileri ve Cas açıklamaları hakkındaki bilgileri birleştirir (Alkhnbashi, ve ark., 2020).

2.2.1.5. CRISPRone

CRISPRone 11102 tam ve 21186 taslak genomda tanımlanan CRISPR-Cas sistemleri için bir web kaynağıdır. Ayrıca, CRISPRone, sahte CRISPR'ler olarak adlandırılan yanlış-CRISPR dizilerine özel bir odaklanmaya sahiptir. Ek olarak, web sitesi CRISPR dizilerinin tahmini, Cas proteinlerinin tanımlanması ve konsensüs tekrarını tamamlayıcı olan ve tracrRNA'nın bir parçası olan anti-tekrarın tahmini hizmetlerini sunar (Alkhnbashi, ve ark., 2020).

2.2.1.6. Anti-CRISPRdb

Anti-CRISPRdb, antiCRISPR proteinleri (Acr) için tek halka açık veri tabanıdır. Yayınlanmış anti-CRISPR proteinlerinin ve bunların kamuya açık veri tabanlarından homologlarının manuel bir koleksiyonundan oluşur. Kullanıcıya bilinen ve potansiyel yeni CRISPR karşıtı proteinlere göz atma, arama, tarama, indirme ve veri paylaşma olanağı sağlar (Dong ve ark., 20178).

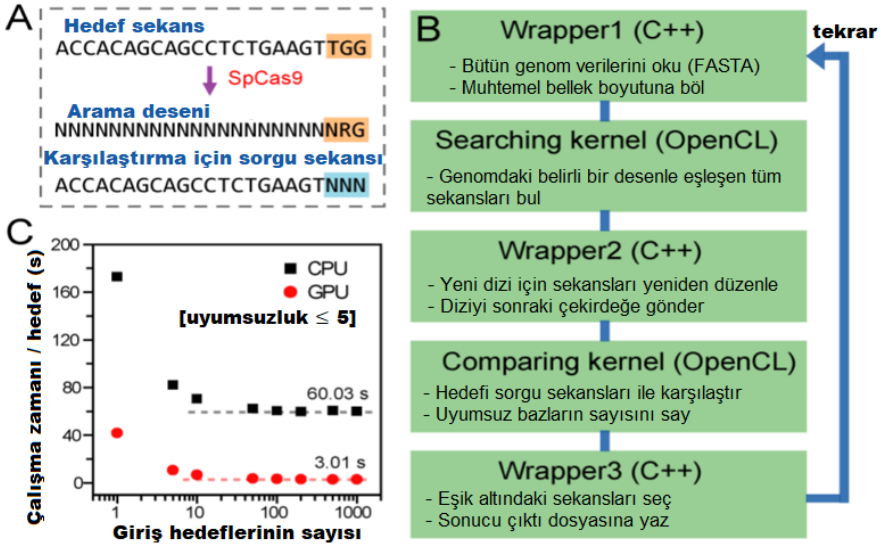
2.2.2. CRISPR-Cas Sistemlerinin Tahmini

2.2.2.1. Cas-OFFinder

Cas9, insan hücrelerindeki genleri düzenlemek için kullanılmıştır. Cas9 proteinleri, farklı PAM dizilerini tanır. *Streptococcus pyogenes* (SpCas9) kaynaklı Cas9, 5'-NGG-3' PAM dizilerini ve nadiren 5'-NAG-3' PAM dizilerini tanır. *Streptococcus thermophilus*'tan (StCas9) (Cong ve ark., 2013) ve *Neisseria meningitidis*'ten (NmCas9) (Hou ve ark., 2013) kaynaklı Cas9 sırasıyla 5'-NNAGAAW-3' (W = A veya T) ve 5'-NNNNGMTT-3' (M = A veya C) tanır. Potansiyel hedef dışı bölgeler aranırken Cas9 tarafından PAM tanımasındaki bozulma hesaba katılmalıdır. SpCas9 durumunda, CasOFFinder ilk olarak ilgili sgRNA dizisine ve 5'-NRG-3' PAM dizilerine karşılık gelen 20-bç dizilerinden oluşan tüm 23-bç DNA dizilerini derler (Şekil 2.4A). Cas-OFFinder daha sonra tüm

derlenmiş dizileri sorgu dizisiyle karşılaştırır ve 20-bç sgRNA dizisindeki uyumsuz bazların sayısını sayar.

Bae ve ark. (2014) tarafından geliştirilen Cas-OFFinder iki farklı OpenCL çekirdeği (arama çekirdeği ve karşılaştırma çekirdeği) ve C++ bölümlerinden oluşur (Şekil 2.4B). İlk olarak, Cas-OFFinder genom dizisi veri dosyalarını tek veya çok dizili FASTA formatlarında okur. FASTA dosyalarını okumak ve ayrıştırmak için açık kaynaklı bir FASTA/FASTQ ayrıştırıcı kitaplığı kullanılır. OpenCL çeşitli işlemcileri desteklese de cihazların hafızası her zaman büyük veri analizi için yeterince büyük değildir. OpenCL cihazlarının bellek sınırlamasının üstesinden gelmek için genom verilerini cihaz belleği tarafından izin verilen mümkün olan en büyük boyuttaki birimlere böler. Bu bölünmüş parçalar daha sonra tüm genomda bir PAM dizisi içeren tüm bölgeleri derleyen arama çekirdeğine yüklenir. Bu belirli bölgeleri aramak ve seçmek için, arama çekirdeği bir işlemcinin her hesaplama biriminde bağımsız olarak çalışır, yani hesaplama birimlerindeki tüm arama işlemleri aynı anda gerçekleştirilir. Bu adımdan sonra PAM dizilerini içeren belirli bölgeler hakkındaki bilgileri toplar ve bu dizileri, eşleşmeyen bazların sayısını sayan karşılaştırmalı çekirdeğe iletir. Arama çekirdeğine benzer şekilde, hesaplama birimlerindeki tüm karşılaştırma işlemleri aynı anda gerçekleştirilir. Son olarak belirli bir eşikten daha az eşleşmeyen baza sahip potansiyel hedef dışı siteleri seçer ve sonuçları, kromozom sayısı, konum, yön, eşleşmeyen bazların sayısı ve küçük harflerle belirtilen uyumsuz bazlara sahip potansiyel hedef dışı DNA dizileri ile birlikte bir çıktı dosyasına yazar. Bu işlemler, tüm bölünmüş parçalar yüklenene kadar tekrarlanır (Bae ve ark. 2014). Cas-OFFinder IUPAC kodlanmış deseni kullanarak genomdaki bütün dizilerin verilerini sağlar ve düzenleme aralığı dışında hiçbirşey hesaplamaz.

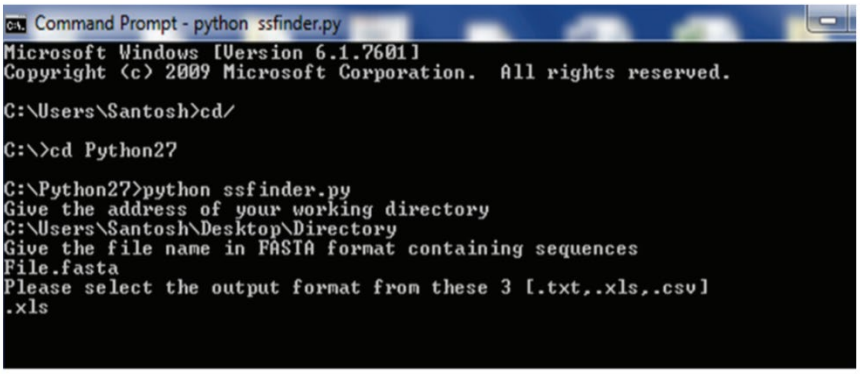


Şekil 2.4. (A) Cas-OFFinder şeması. (B) CasOFFinder'ın iş akışı. (C) CPU (siyah kareler) ve GPU (kırmızı daireler) aracılığıyla giriş hedef sitelerinin sayısının bir fonksiyonu olarak hedef site başına çalışma süresi (Bae ve ark., 2014'ten değiştirilmiştir)

2.2.2.2. SSFinder

Upadhyay ve Sharma, (2014) tarafından geliştirilen SSFinder bir Python betiğidir. Yalnızca sistemde yüklü olan uyumlu bir Python sürümüne (2.2 veya daha yüksek sürüm) ihtiyaç duyar. SSFinder'ı kullanmak için, kullanıcıların komut dosyasını indirmesi veya komut dosyasını bir not defterine kopyalayıp yapıştırması gereklidir. Windows kullanıcılarının Python yürütülebilir dosyasını (C:\Python) barındıran dizine "ssfinder.py" olarak kaydetmesi gerekir (Şekil 2.5). FASTA formatlı nükleotid dizilerinin girdi dosyasını içeren bir çalışma dizini de gereklidir. Linux kullanacıları terminalde "$python SSFinder.py" komutunu çalıştırarak kullanabilir. Kullanıcı, SSFinder tarafından istendiğinde, çalışma dizininin adresini, giriş dizilerinin dosya adını ve istenen çıktı biçimini (.xls, .txt veya .csv gibi) sağlar. SSFinder, bir nükleotidin adım boyutunda 23 nükleotitlik bir pencereyi hareket ettirerek bir DNA dizisini tarar. 3' "NGG" PAM dizisine sahip dilimler seçilir ve 12

nükleotid tohum dizisinin varlığı için analiz edilir. Tohum dizisi, CRISPR-Cas sisteminin özgüllüğünü belirlediğinden (Mali ve ark., 2013c; Upadhyay ve ark., 2013), araç, bu dizilerin tüm girdi genom verilerinde tekrarlanmamasını sağlar (Jinek ve ark., 2013). Tüm giriş dizisi dosyasında farklı tohum dizisine sahip diziler, çıktı dosyasında belirli hedef bölgeler olarak görüntülenir. Çıktı dosyası otomatik olarak "SSFinder-output" dosya adıyla aynı dizine kaydedilir. Çıktı dosyası dizilerin tanımlayıcısı, başlangıç ve bitiş nükleotidlerine göre dört grupta sınıflandırma, "NGG" PAM dizisi dahil potansiyel hedef bölgeler, hedef bölgelerin başlangıç ve bitiş konumları, spesifik 12 nükleotid tohum dizilimi ve spesifik CRISPR-Cas hedef bölgeleri içerir (Şekil 2.6) (Upadhyay ve Sharma, 2014).



Şekil 2.5. Windows işletim sisteminde SSFinder kullanımı (Upadhyay ve Sharma, 2014)

Identifier	Classification	CRISPR-target site with "NGG" PAM	Position		Condition with seed sequences	Specific CRISPR-target site
			Start	End		
Sequence_1	A or TN18A or T	ACTTCTTCGTCCAACTTCTTCGG	6	28	A or TN7S1IA or T	ACTTCTTCGTCCAACTTCTT
Sequence_1	A or TN18G or C	TTTGCAAGCCTCATCCATTGTGG	386	408	A or TN7S1IG or C	TTTGCAAGCCTCATCCATTG
Sequence_1	G or CN18A or T	CCAAAGTTCTATTTGAGCTAAGG	68	90	G or CN7S1IA or T	CCAAAGTTCTATTTGAGCTA
Sequence_1	G or CN18G or C	CTAACCGACCTTCAGCTAACAGG	154	176	G or CN7S1IG or C	CTAACCGACCTTCAGCTAAC

Şekil 2.6. SSFinder tarafından verilen çıktı dosyasına genel bakış (Upadhyay ve Sharma, 2014)

2.2.2.3. CRISPRseek

Zhu ve ark., (2014) tarafından geliştirilen ve Bioconductor'da dağıtılan bir R paketi olan CRISPRseek dizileri temsil etmek ve eşleşmeleri bulmak için Biostrings paketini, hedef dışı dizilerin herhangi bir ekzonla örtüşüp örtüşmediğini belirlemek ve genomik koordinatları temsil etmek için GenomicRanges paketini, hedef dışı olan kuşatma sekanslarını almak için Bsgenome paketini kullanır. CRISPRseek, Bsgenome nesnesi ve Bioconductor'daki TransciptDb nesnesi gibi ek açıklamalar için çeşitli genom kaynaklarından yararlanır.

CRISPRseek, çeşitli kısıtlamalar kullanarak belirli bir giriş dizisini hedeflemek adına CRISPR-Cas9'dan türetilen nükleazlar için aday gRNA'ları tanımlar ve değerlendirir (Şekil 2.7). Ayrıca her gRNA için potansiyel hedef dışı dizileri raporlar ve sıralar (Şekil 2.8) (Zhu ve ark., 2014).

gRNAsPlusPAM	toViewInUCSC	target efficacy	Top10 offtarget Score	RE site
CTACTGTGTGCACTTCATCCTGG	chr4:3215834-3215856	0.0768	20	
TGAAGTGCACACAGTAGATGAGG	chr4:3215828-3215850	0.2668	21.9	
GAAGTGCACACAGTAGATGAGGG	chr4:3215827-3215849	0.3687	11.9	
CACACAGTAGATGAGGGAGCAGG	chr4:3215821-3215843	0.2049	31.5	BsaXI Cac8I
AGTAGATGAGGGAGCAGGCGTGG	chr4:3215816-3215838	0.6173	12.3	BsaXI Cac8I
GTAGATGAGGGAGCAGGCGTGGG	chr4:3215815-3215837	0.6881	15.8	

Şekil 2.7. CRISPRseek sonuç çıktısı gRNA'ların özeti (Zhu, 2020)

OffTargetSequence	inExon	inIntron	entrez_id	gene	score	n.mismatch	mismatch.distance2PAM	alignment	NGG	strand	chrom	chromStart
CCAGTTTGTGGATCCTGCTCTGG	TRUE		2623	GATA1	100	0		.....	1	+	chrX	48649564
ACAATTTCTGGATCCTGCTCCAG		TRUE			2.6	3	20,17,13	A..A...C.....	0	-	chrX	70453482
GCAGTTTATGGATCCTGCTGGAG					1.5	3	20,13,1	G.....A.....G	0	+	chr11	128187317
TGAGTTTGTGGTTCCTGCTCTGG		TRUE			1.4	3	20,19,9	TG.....T.....	1	-	chr4	115807036
ACAGATTGTGGATCCTCCTCTGG		TRUE			1.3	3	20,16,4	A...A.....C...	1	+	chr5	58007599
TCACTTTGTGGACCCTGCTCTGG					1.2	3	20,17,8	T..C.....C.....	1	+	chr11	127110216
CCAGGTTGTTGAGCCTGCTCAAG		TRUE			1	3	16,11,8	....G....T..G.....	0	+	chr6	124177937
ACAGTATGTGAATCCTGCTCCGG					0.9	3	20,15,10	A....A....A.....	1	-	chr8	54555407
ACAGTGTGTGAATCCTGCTCTGG					0.9	3	20,15,10	A....G....A.....	1	-	chr4	136302357
TCAGTGTGTGGTTCCTGCTCCAG					0.9	3	20,15,9	T....G.....T.....	0	-	chr8	58413423
CCACTTTGGGGTTCCTGCTCCGG		TRUE			0.8	3	17,12,9	...C....G..T.....	1	-	chr15	89911561
GTAGTTTGTGGATCCTGTTCTAG		TRUE			0.7	3	20,19,3	GT.....T..	0	+	chrX	10161855
CCAGTTTTTGGACCCTGCTGGAG		TRUE			0.5	3	13,8,1	.....T....C.....G	0	+	chr3	76589098
GTGTCCTCCACACCAGAATCAGG	TRUE		2623	GATA1	100	0		.....	1	+	chrX	48649586
GTGTGCTACACACCAGAATCTGG					5.7	2	16,13	....G..A.....	1	-	chr1	244432157
CTGCCCTCCTCACCAGAATCAGG		TRUE			2.5	3	20,17,11	C.C.....T.....	1	-	chr2	165419882
CTGTTGTCCACACCAGAATCTGG					1.5	3	20,16,15	C...TG.....	1	-	chr18	73966192
CAGTCCTCCACATCAGAATCAGG		TRUE			1.2	3	20,19,8	CA.....T.....	1	+	chr9	19596750
GTGTTTTCCACACCAGAATGCAG					0.8	3	16,15,1	....TT.....G	0	+	chrX	81194638
TTGTCCTTCACACCAGATTCCAG					0.7	3	20,13,3	T.....T.....T..	0	-	chr22	16556222
GTGTCGTCCGCACCAGCATCAGG		TRUE			0.6	3	15,11,4	.....G...G.....C...	1	+	chr2	239777630
GTGTCCCCCAGGCCAGAATCAAG					0.5	3	14,10,9	.....C...GG.....	0	-	chr8	43427226
TGTCCTCCACACCAGAATCAGGG	TRUE		2623	GATA1	100	0		.....	1	+	chrX	48649587
TGTCCTCAACACCAGAATGATAG		TRUE			2.9	2	13,2	.....A.....G.	0	-	chrX	83738197
TTTCTTCTACACCAGAATCACAG					2.5	3	19,16,13	.T..T..T.....	0	-	chr5	119290894
TGTCCTGCACACCAGAATGATAG					2.2	2	14,2	.....G.....G.	0	-	chr17	51150403
TGTCCTCCTCCCCAGAATCAAAG		TRUE			1.9	2	12,10	.....T.C.....	0	+	chr19	51673303
TGGGCTGCACACCAGAATCACAG		TRUE			1.6	3	18,17,14	..GG..G.....	0	-	chr16	29695131
TGTGCTACAAACCAGAATCACAG					1.6	3	17,14,11	...G..A.A.....	0	-	chr10	70009486
TGACCCCCAAACCAGAATCATGG					1.4	3	18,15,11	..A..C...A.....	1	+	chr9	101650683
TGTGCTACACACCAGAATCTGGG					1	3	17,14,1	...G..A.....T	1	-	chr1	244432156
TTTCCTACACAACAGAATCACAG					0.9	3	19,14,9	.T....A....A.....	0	+	chr1	81398698
TGTCCCCCAGGCCAGAATCAAGG					0.8	3	15,11,10	.....C...GG.....	1	-	chr8	43427225
TCTTCTCCACACCAGAAGCAGAG		TRUE			0.7	3	19,17,3	.C.T.....G..	0	-	chr1	156082357

Şekil 2.8. CRISPRseek sonuç çıktısı hedef dışı bölgelerin detaylı bilgileri (Zhu, 2020)

2.2.2.4. sgRNAcas9

Xie ve ark., (2014) tarafından Perl dilinde kodlanmış olan sgRNAcas9 (sürüm 2.0.6), Windows, Linux ve Mac OS X sistemlerinde yerel olarak çalıştırlabilen bir yazılım paketidir. Pakette sgRNAcas9.pl (ana komut), format_genome.pl, ot2gtf.pl, pot2gtf.pl, check_sgRNA_seq.pl, sgRPrimer.pl, ve extract_targetSeq.pl olmak üzere her biri farklı görevler yürüten 7 Perl betiği vardır.

sgRNAcas9 iş akışının ana adımları; (1) CRISPR hedef bölgelerini arar. (2) Hedef dışı etkileri değerlendirir. (3) sgRNA ekspresyon vektörünü seçilir ve oligonükleotitler tasarlanır. (4) Cas9 endonükleaz bölünme aktivitesini doğrulamak adına PCR primer çiftlerini tasarlamak için hedef üzerindeki veya dışındaki bölünme

bölgelerini çevreleyen istenen uzunluktaki nükleotit dizilerini çıkarır. Adım 1 ve 2, format_genome.pl, sgRNAcas9.pl, ot2gtf.pl ve pot2gtf.pl komut dosyası tarafından gerçekleştirilir. Adım 3, check_sgRNA_seq.pl ve sgRPrimer.pl tarafından gerçekleştirilir; Adım 4, extract_targetSeq.pl tarafından gerçekleştirilir.

Giriş dizileri, 5' ila 3' yönünde FASTA formatında sağlanır. Girdi dosyası olarak verilen herhangi bir dizi, bu programda "sense strand" olarak adlandırılır. "Her iki dizi arama modu" veya "eşleştirilmiş-gRNA arama modu" seçildiğinde, dizi, sgRNAcas9 programı tarafından ters tamamlayıcı karşılığına dönüştürülür. Genom ve GTF dosyaları, NCBI web sitesinden (ftp://ftp.ncbi.nlm.nih.gov/genomes) veya Ensembl ftp sitesinden (http://www.ensembl.org/info/data/ftp/index.html) yüklenebilir. FASTA başlıklarına sahip genom DNA dizisi, ana program sgRNAcas9.pl'yi çalıştırmadan önce format_genome.pl komut dosyası tarafından ön işleme tabi tutulmalıdır.

2.2.2.5. pgRNAFinder

pgRNAFinder'ın web arayüzü tek gRNA tasarım modülü, kısa mesafe eşleştirilmiş gRNA tasarım modülü ve uzun mesafeli eşleştirilmiş gRNA tasarım modülü olmak üzere üç modülden oluşur (Xiong ve ark., 2017). Kullanıcılar bir hedef dizi veya geni seçer veya yükler. Kullanıcı girişi veya yüklemesinde, bir gen sembolü, genom bölgesi (gen yapısı sorgusu ile) veya doğrudan dizi girme gibi üç seçenek vardır. Daha sonra kullanıcıdan bir tür ve bir gRNA tasarım modülü seçmeleri istenir.

gRNAFinder'ın web arayüzü hem gRNA arama hem de puanlama sistemini içerir. Kullanıcılar sorgu dizilerini girdikten sonra, 3' protospacer-bitişik motif (PAM) ile gRNA'yı ve olası hedef dışı hedefleri arar ve silinen dizilerin korunması puanlanır. CRISPR bölgelerini belirlemek için filtreler uygulanabilir (Xiong ve ark., 2017).

2.2.2.6. CasOT

CasOT bir Perl betiğidir (Xiao ve ark,, 2014). Tekli-gRNA, eşleştirilmiş-gRNA ve hedef ve hedef dışı olmak üzere üç arama modunu içerir. Tek-gRNA arama modunda, kullanıcıların hedef dışı bölgeleri veya herhangi bir kullanıcı tarafından belirlenmiş diziyi aramak için 21-33-nükleotit hedef bölgeleri (18-30-nükleotit protospacer artı 3'-uçta PAM -NGG) içeren bir giriş FASTA dosyası ve potansiyel hedef dışı bölgelerin eksonlarda yer alıp almadığını belirlemek için bir genom açıklama dosyası (Ensembl GTF formatı) sağlamaları gerekir.

Program hedef dizilerini okur ve bunları PAM'ler, tohum bölgeleri ve tohum olmayan bölgelere ayırır. Çekirdek bölgelerine benzer tüm olası 12-nükleotit dizileri ve bunların ters tamamlayıcı dizilerini içerecek şekilde bir havuz oluşturulur. Uyumsuzluklar, kullanıcı tarafından belirtilen seçeneğin izin verdiğinden daha fazla değildir. Genom dizileri veya kullanıcı tarafından sağlanan diğer diziler, önceden oluşturulmuş tohum havuzunda herhangi bir 12-nükleotit fragmanının görünüp görünmediğini belirlemek için nükleoti- nükleotit taranır. Bulunursa, parçanın bitişik genomik dizileri çıkarılır ve tohum olmayan bölgedeki PAM tipi ve uyumsuzluk numarasının kullanıcı tarafından belirtilen seçeneklerle eşleşip eşleşmediğini kontrol etmek için ek test adımları yürütülür. Tüm kriterler karşılanırsa, fragmanın genom veya dizideki konumu ve dizisi kaydedilir. Ardından, konumlar, diziler, PAM tipi ve sayılar ve uyumsuzlukların ayrıntılı açıklaması dahil olmak üzere potansiyel hedef dışı bölgelerin ayrıntılı bilgileri, her bir hedef bölgeye karşılık gelen ayrı dosyalar olarak çıktılanır. Açıklama dosyası sağlanırsa, bu site bir eksonda bulunduğunda gen kimliği ve gen sembolü, hedef dışı bilgilere eklenir (Xiao ve ark., 2014).

2.2.2.7. CRISPRdirect

CRISPRdirect, bir giriş dizisinden rasyonel CRISPR/Cas hedefleri seçmek için basit ve işlevsel bir web sunucusudur (Naito ve ark., 2015). Hedef dizi gereksinimleri iki yönlüdür. İlk olarak, 5′-NGG PAM dizisi, hedef diziye bitişik

olarak yerleştirilmelidir. İkinci olarak, hedef dışı düzenlemeden kaçınmak için hedef dizi tüm genom içinde spesifik olmalıdır. CRISPRdirect, genomik dizilere karşı kapsamlı aramalar gerçekleştirerek, kullanıcıların minimum hedef dışı bölgelere sahip rasyonel hedef dizileri kolayca seçmelerini sağlar. Sunucu belirli sayıda genomik dizileri içermektedir (Naito ve ark., 2015).

2.2.2.8. CRISPRFinder

CRISPRFinder, Debian Linux altında Perl'de geliştirilmiştir (Grissa ve ark., 2007a). Web aracının girişi, 'FASTA' formatında 67 Mb uzunluğa kadar genomik bir sorgu dizisidir. CRISPR'lerin (en az bir motiften oluşan) olası konumları, maksimum tekrarlar bulunarak tespit edilir. Bir maksimum tekrar bir uyumsuzluk meydana gelmeden her iki yönde de uzatılamayan bir tekrardır. N boyutundaki bir dizideki maksimum tekrarların toplam sayısı doğrusaldır (n'den küçüktür) ve hesaplama doğrusal zamanda bir sonek-ağaç tabanlı algoritma kullanılarak yapılabilir.

Programın çalışması 4 ana adımdan oluşur. İlk adımda 25-60 bç'lik dizilerle aralıklı 23-55 bç uzunluğundaki maksimum tekrarlar taranır. İkinci adımda korunmuş bölge konsensüsünün tüm genomda aday korunmuş bölgelerin oluşum sayısını dikkate alarak tanımlanmış bir puana göre seçilmesi ve ilk veya son nükleotidlerdeki uyumsuzluklar yerine korunmuş bölgeler arasındaki dahili uyumsuzluklara öncelik verilir. Üçüncü adımda CRISPR tanımına uyup uymadıklarını kontrol ettikten sonra aday CRISPR'leri tanımlanır. Dördüncü adımda ise artık tandem tekrarlarını ortadan kaldırır (Grissa ve ark., 2007a).

2.2.2.9. CRISPRCasFinder

CRISPRCasFinder web sunucusu, bağımsız ön ve arka uçlara dayanmaktadır (Couvin ve ark., 2018). Ön uç, .NET Core (dotNet core) geliştirme platformu ve C# programlama dili kullanılarak oluşturulan bir web uygulamasdır.

CRISRFinder'ın bir güncellemesi olan CRISPRCasFinder, taşınabilir bir sürüm, gelişmiş performans içerir ve Cas proteinleri için aramayı entegre eder. CRISPRCasFinder, Linux (Linux için Windows Alt Sistemi dahil) ve MacOS sistemleriyle uyumlu bağımsız bir program olarak da mevcuttur. Program Perl'de yazılmıştır (Couvin ve ark., 2018).

2.2.2.10. PILER-CR

PILER-CR, tekrarların analizi için PILER algoritma ailesine dayanan CRISPR tekrarlarının tanımlanması için tasarlanmış bir araçtır (Edgar, 2007). PILER-CR algoritması, sağlanan girdi genomunun kendisiyle yerel hizalamalarını bulmakla başlar. Bu kendi kendine hizalamalar, kısa bir mesafeyle ayrılmış bir CRISPR dizisinin iki potansiyel tekrar dizisini temsil eder. Algoritma daha sonra bir yığın kavramını tanıtır. Yığın, her biri tekrarlayan bir elemanı temsil eden en az bir yerel hizalama ile kaplanmış sürekli bir taban seti olarak tanımlanır. Yığınlar yerel hizalamalarla birbirine bağlanır ve bunlar yığınların bir bağlantı grafiği oluşturmaya yarar. Burada yığınlar grafiğin düğümleridir ve iki yığın arasındaki her yerel hizalama bir kenardır. Her bağlı alt grafik, kısmi eşleşmelerin filtrelenmesi ve tekrar adaylarının genişletilmesi gibi birkaç adımda rafine edilen bir CRISPR dizisi adayını temsil eder. Tekrar sınırları kontrol edilir ve koruma seviyesi %90'ın altında ise kolon tekrarlardan silinir. Son olarak, oldukça benzer (>%95) konsensüs tekrar dizilerine sahip veya birbirine yakın (<tekrar +aralayıcı uzunluğu) CRISPR dizi adayları birleştirilir (Alkhnbashi ve ark., 2020).

2.2.2.11. CRISPRDetect

Biswas ve ark., (2016) tarafından geliştirilmiş olan CRISPRDetect ile CRISPR dizilerinin saptanması, beş adımlı bir prosedürdür. İlk olarak, ilk adaylar, en az üç tekrar ile 11 bç tekrarlanan alt diziler için bir düzenli ifade araması yürütülerek tanımlanır. Ardından, ikinci adımda, aday aralayıcılar hizalanır ve

benzerlikleri çok yüksekse, aday dizi tandem tekrarı olarak işaretlenir ve atılır. Kalan CRISPR adaylarının her biri, tekrar dizilerinin her iki yönde uzatılmasıyla başlayan bir iyileştirme prosedüründen geçer. Bunun için aday dizinin tüm tekrarlarındaki bitişik bazlar incelenir ve %75'i aynı ise tekrarlar uzatılarak sonraki bazlar kontrol edilir. Tekrar uzantısından sonra, algoritma tüm CRISPR dizisini genişletmeye çalışır. Bunun için, temsili bir tekrar dizisi seçilir ve yerel bir hizalama yoluyla yan bölgelerdeki eksik tekrarları tespit etmek için kullanılır. CRISPRDetect algoritması Perl'de uygulanır ve Linux işletim sistemine sahip makinelerde bir komut satırı aracı olarak kullanılabilir.

2.2.2.12. CRISPy-web

Blin ve ark., (2016) tarafından geliştirlmiş olan CRISPy-web analiz için bir dizi hazırlama ve ardından Cas9 hedef sitelerini tarama ve sıralama olmak üzere iki adımdan oluşur. Hazırlık adımında, girdi dosyaları ayrıştırılır ve dosyalar antiSMASH'dan aktarıldıysa Biopython kullanılarak ikincil metabolit gen kümesi konumları ile açıklama yapılır. Tarama adımında ise PAM'ın akış yukarısında aynı 13 bç çekirdek dizisine sahip diziler için doğrudan hedef dışı isabetlerin sayısını sayarak, ilgilenilen seçili bölgenin hem ileri hem de geri dizilerinde 'NGG' PAM dizisini arar. Ardından, tüm olası kombinasyonları göz önünde bulundurarak 13 bç dizisi için tüm genom üzerinde 0, 1, 2 veya 3 bp yanlış eşleşme içeren hedef dışı isabetlerin sayısını hesaplar. Tüm olası Cas9 hedeflerini topladıktan sonra, doğrudan hedef dışı eşleşmeleri olan bölgeler, tek bir baz çifti mutasyonu ile hedef dışı eşleşmeleri olan bölgeler veya iki baz çifti mutasyonu ile ondan fazla hedef dışı eşleşmesi olan bölgeler atılır. Bunlar bir Genbank dosyasında saklanır. Kullanıcı arayüzünde görüntülenen sitelerin sayısını daha da azaltmak için, hedefler artık 2 bç uyumsuzluk hedef dışı, 3 bç uyumsuzluk hedef dışı ve genomdaki koordinatlara göre sıralanır ve kullanıcıya yalnızca ilk 1000 hedef gösterilir.

2.3. gRNA Tasarımı

Genom düzenlemede, ilgili bir DNA dizisi genom içine doğrudan eklenir, değiştirilir veya silinir (Wang ve ark., 2016). CRISPR-Cas9 ve diğer Cas9 benzeri proteinler, hedef tasarımın basitliği, verimlilik, minimum maliyet, çoğullanmış düzenleme ve diğerleri gibi diğer gen düzenleme yöntemleriyle karşılaştırıldığında çeşitli avantajlara sahiptir. Bugüne kadar birçok çalışma, CRISPR-Cas9 ve ilgili sistemlerin insan, maymun, zebra balığı, fare, meyve sineği ve domuz dahil olmak üzere çeşitli türler ve hücre türleri için en etkili gen düzenleme araçları olduğunu göstermiştir (Wang ve ark., 2016; Sander ve Joung, 2014; Hsu ve ark., 2014; Baliou ve ark., 2018). 2014'ten bu yana CRISPR-Cas9, genom düzenleme için baskın araç haline geldi ve bu da son derece spesifik ve verimli gRNA'lar tasarlamak için hesaplama araçlarının hızla artmasına yol açtı. Bu gRNA'ların tasarımı, aralarında hedef özgüllüğü, yani hedef dışı etkilerin olmaması gibi karşılanması gereken birkaç kritere sahiptir.

gRNA tasarımı için geliştirilen araçların hepsi aynı temel ilkeleri takip eder. İlk adım, ilgilenilen gene, uygulamaya, örneğin nakavt, bastırma veya aktivasyona ve kullanılan CRISPR-Cas sisteminin PAM'sine bağlı aday hedef bölgeleri belirlemektir. Aday hedef bölgeler yüksek hedefte verimlilik ve düşük hedef dışı aktivite olmak üzere değerlendirilir. İkinci adım puanlar temel olarak uyumsuzluk pozisyonlarına dayanır ve oldukça iyi tahmin gücü elde eder (%98'e kadar güvenilir). Hedef verimlilik puanları daha az ayırt edicidir. Ancak bu puanlar yalnızca deneysel olarak oluşturulmuş hedefleme dışı kurallara dayanmaktadır.

Hedef dışı hedeflerin güvenilir tahmini, yalnızca puanlamaya değil, aynı zamanda tüm genomlarda hedef dışı aday bölgeleri bulmak için kullanılan algoritmaya da bağlıdır. Bu tür bölgeler, gRNA ile 1-3 uyumsuzluğa sahiptir ve okuma haritalama araçlarıyla verimli bir şekilde tanımlanabilir (Alkhnbashi ve ark., 2020).

2.4. Hedef Tanımlama

CRISPR-Cas sistemi tarafından spesifik hedefleme, iki mekanizma ile sağlanır. Birincisi, aralayıcılar, hedeflenen nükleik asit için neredeyse mükemmel bir şekilde tamamlayıcı olmalıdır. İkinci olarak, hedefe, ilgili Cas genlerine özgü olan doğru PAM eşlik etmelidir. Bir uygulama ortamında iyi çalışılmış CRISPR-Cas9 sistemi (PAM dahil) için, hedefleme verimliliği ve potansiyel hedef dışı hedefler büyük ilgi görmektedir. Bu yönleri incelemek için çeşitli araçlar geliştirilmiştir. Tanımlanmamış PAM'lere sahip yeni tanımlanmış CRISPR-Cas sistemleri durumunda, hedef tanımlama, aralayıcı dizilerine dayalı benzerlik aramalarına başvurur (Alkhnbashi ve ark., 2020).

2.5. Hedef Bölgede Verimlilik Tahmin Modelleri

sgRNA'ların Cas9 hedeflemedeki etkinliklerinin farklılık gösterdiğini için CRISPR-Cas9 sistemi, bölünme aktivitelerinde yüksek derecede değişkenlik gösterir (Doench ve ark., 2014; Lee ve ark., 2016; Wilson ve ark., 2018). Yapılan çalışmalarda, gRNA'nın nükleotid bileşiminin, PAM'ın hemen yukarı akışındaki bir G tercihi gibi konuma özgü nükleotidlerin, GC içeriğinin, gRNA erime sıcaklığının, poli-T dizilerinin, sgRNA'nın termodinamik stabilitesinin, transkripsiyon başlangıç bölgesine göre hedef bölge konumu ve hedef bölge tanımı (yalnızca 20 bç hedef dizi + PAM + yan dizilim dahil), CRISPR-Cas9 etkinliği üzerinde önemli bir etkiye sahip olduğunu göstermiştir (Doench ve ark., 2014; Graham ve Root, 2015; Wilson ve ark., 2018). Deneysel verileri farklı algoritmalar/makine öğrenme yöntemleri ile birlikte bir eğitim seti olarak kullanarak, sgRNA hedef etkinliğini tahmin etmek için bu özellikleri içeren etkinlik tahmin modelleri geliştirilmiştir (Cui ve ark., 2018). Kural Seti 1 ve 2 (Doench ve ark., 2014; Doench ve ark., 2016), SVM (Destek Vektör Makinesi) (Labaj ve ark., 2017), SVM Sınıflandırıcı modeli (Chari ve ark., 2015) ve Elastic Net (Zou ve Hastie, 2005), sgRNA tasarım araçlarında kullanılan en yaygın modellerdir.

Model tahmin verimliliği ile ilgili olarak, sgRNA aktivitesini tek bir özelliğin yönetmediği gözlemlenmiştir. Model tahmin verimliliğinin, eğitim seti deneysel koşulları ve kullanılan veri seti, gRNA transkripsiyon yöntemi ve CRISPR-Cas9 aktivitesinin ölçülme şeklinden güçlü bir şekilde etkilendiği gözlemlenmiştir (Wilson ve ark., 2018).

2.6. Hedef Dışı Tahmin Algoritmaları

Mevcut sgRNA tasarım araçları, hedef diziyi referans genoma göre hizalamak için CRISPOR'da (Haeussler ve ark., 2016) BWA (Li ve Durbin, 2009), CCTop'ta (Stemmer ve ark., 2017) Papyon (Langmead ve ark., 2009) ve CHOPCHOP'ta (Montague ve ark., 2014) Ceza Matrisi (Hsu ve ark., 2013) gibi hizalama araçlarını kullanmıştır. Potansiyel hedef dışı bölgeleri belirlemek için, sgRNA tasarım aracı, tohum bölgelerinde düşen uyumsuzluk sayıları ve konumları arar ve buna göre hedef dışı bölgelere bir puan atar. Araçlar, farklı temel uyumsuzluk toleransı seçme seçeneğine sahip olduğundan dolayı hedef dışı tahmin sonuçlarında farklılık gösterir. Hizalama araçları tarafından tahmin edilen, deneysel olarak doğrulanmış hedef dışı konumları karşılaştıran çalışmalar, yanlış-pozitif (hedef dışı değerler aslında CRISPR/Cas9 veya fazla tahmin ile bölünür) ve yanlış-negatif (hedefler dışında kalanlar, hedef bölgeden veya eksik tahminden önemli ölçüde farklıdır) hedef site tahminleriyle karşılaştırılmıştır (Wilson ve ark., 2018). Yanlış-pozitif ve yanlış-negatif tahminleri dengelemek ve yanlış-pozitifleri filtrelemek için puanlama algoritmaları geliştirilmektedir. Geliştirilen iki puanlama yöntemi MIT-Broad (Hsu ve ark., 2013) ve CFD (cutting frequency determination) puanıdır (Doench ve ark., 2016).

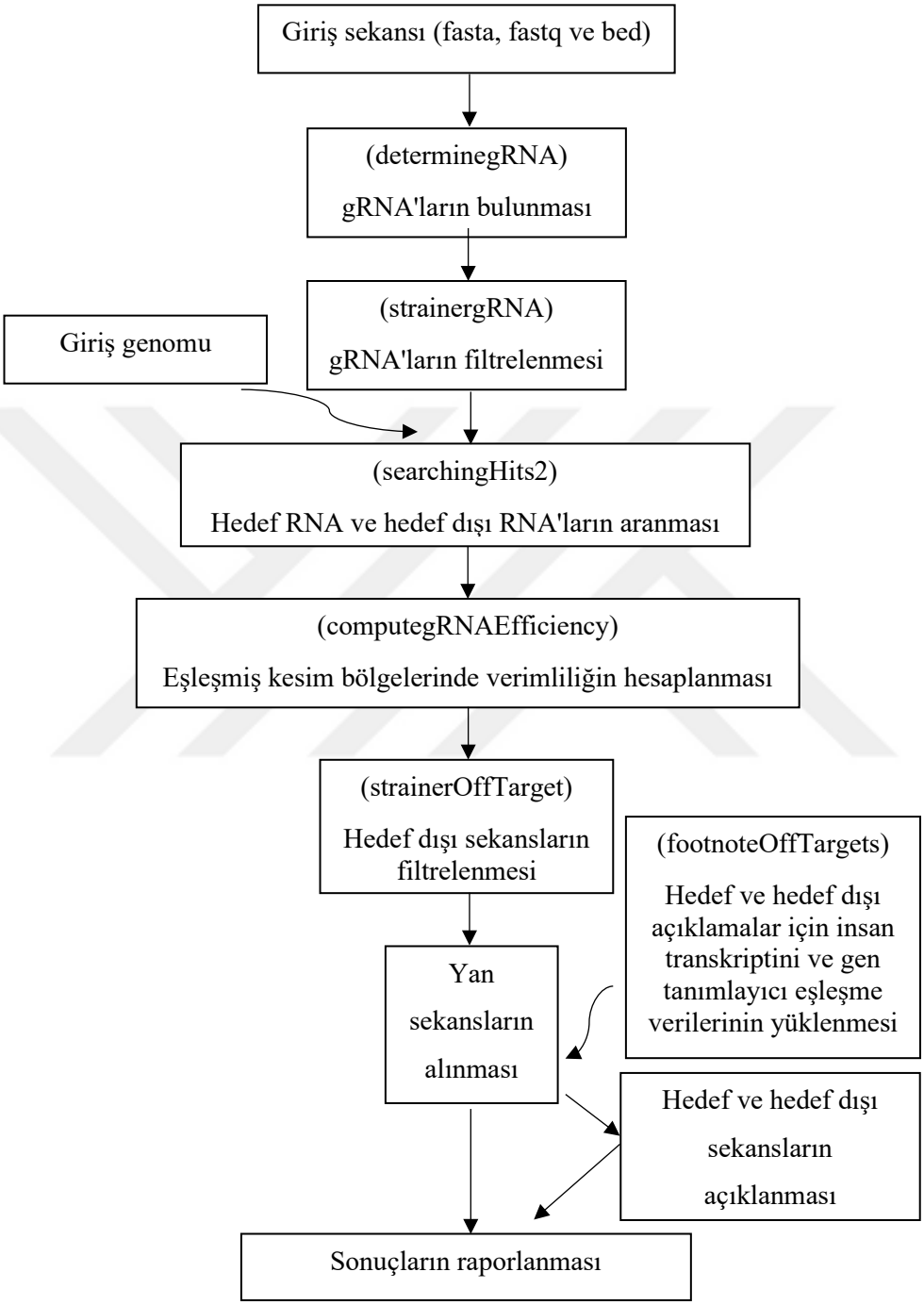
3. MATERYAL VE METOT

3.1. Materyal

İstatistiksel hesaplamalara ve grafiklere olanak tanıyan bir programlama dili olan Python'da CRISPR-Cas9 türevli sistemler için ortak bir gRNA tasarımı ve hedef dışı analiz için Şekil 3.1'de görülen genel mimari uygulanmıştır. Bileşenlerin yeniden kullanılması, eşleşmeleri bulması, hedef dışı dizilerin herhangi bir eksonla örtüşüp örtüşmediğini verimli bir şekilde belirlenmesi, genomik koordinatların temsili, hedef dışı olan yan sekansların alınması ve ek açıklamalar gibi çeşitli genom kaynaklarından yararlanılmıştır. Potansiyel ötelemeler için bölünme skorunu tahmin etmek amacıyla deneysel olarak belirlenmiş bir pozisyona özgü uyumsuzluk ceza matrisi ve skorlama şeması kullanılmıştır (Hsu ve ark., 2013). gRNA tanımlama, filtreleme, hedef dışı arama, puanlama ve bilgi notu, tek bir iş akışı fonksiyonu ***Non-TargetAnalysis*** içine entegre edilmiştir. Kütüphane yüklendikten ve parametreler ayarlandıktan sonra, bir kod satırı (***Non-TargetAnalysis***) çeşitli yardımcı fonksiyonları çağırarak tüm analizi gerçekleştirir. Kodlama ve testler için Python 3.8.5 sürümü kullanılmıştır.

Çizelge 3.1. Kullanılan kaynak dosyalar

Yararlanılan dosyalar	***Kaynak***
Ceza matrisi ve skorlama şeması	*RNA kılavuzlu Cas9 nükleazlarının DNA hedefleme özgüllüğü (Hsu ve ark., 2013)*
CFD	*Etkinliği en üst düzeye çıkarmak ve CRISPR-Cas9'un hedef dışı etkilerini en aza indirmek için optimize edilmiş sgRNA tasarımı (Doench ve ark., 2016)*
Kural seti 1	*CRISPR-Cas9 aracılı gen inaktivasyonu için son derece aktif sgRNA'ların rasyonel tasarımı (Doench ve ark., 2014)*
CRISPRscan kural seti	*CRISPRscan: in vivo olarak CRISPR-Cas9 hedeflemesi için yüksek verimli sgRNA'lar tasarlama (Moreno-Mateos ve ark., 2015)*
UCSC tarafından sağlanan ve genom dizileri (2bit dosya formatında)	*https://genome.ucsc.edu/*
UCSC tarafından oluşturulan bir açıklama veritabanları (sqlite formatında)	*https://genome.ucsc.edu/*



Şekil 3.1. Non-TargetAnalysis fonksiyonu mimarisi

3.2. Metot

Çift nikaz olarak kullanılabilen veya örtüşen restriksiyon enzimi kesim bölgesine sahip olan/olmayan çiftler olarak mevcut olan/olmayan potansiyel hedef dizileri otomatik olarak bulur. Daha sonra, kullanıcı tanımlı bir maksimum uyumsuzluk sayısı ile genom çapında hedefler dışında arama yaparak, hedef dışı ve ceza ağırlık matrisindeki uyuşmazlıklara dayalı olarak her hedef dışı bölgenin puanını hesaplar. Hedef dışı hedefleri kullanıcı tanımlı filtreyle filtrelenebilir yan dizilerle birlikte hedeflere ek açıklama eklenebilir. Toplam hedef dışı puana göre sıralanan gRNA'ları, restriksiyon enzimi kesim bölgeleri ve olası eşleştirilmiş gRNA'ları içeren bir özet raporu da dahil olmak üzere çeşitli raporlar oluşturulur. Ayrıntılı eşleştirilmiş gRNA bilgileri, restriksiyon enzimi kesim bölgeleri ve hedef dışı diziler ve puanlar, kullanıcı tarafından belirtilen çıktı dizininde ayrı dosyalarda saklanır. Toplamda, çıktı dizininde dört sekmeyle ayrılmış dosya oluşturulur: Non-TargetAnalysis.xls (hedef dışı ayrıntılar), Summary.xls (gRNA özeti), RecutDetails.xls (her gRNA'nın restriksiyon enzimi kesme bölgeleri) ve pairedgRNAs.xls (potansiyel eşleştirilmiş gRNA'lar). Bu raporlar, *Streptococcus pyogenes* CRISPR-Cas9 nükleazlarını ve türevlerini tanımlamak, seçmek ve kullanmak için kapsamlı bir bilgi seti sağlar. Paket, alternatif genomik dizileri hedeflemek için kullanılabilen diğer bakteri türlerinden CRISPR-Cas9 kompleksleri için farklı gRNA ve PAM dizisi gereksinimlerini kabul edecek şekilde kolayca değiştirilebilir. Ayrıca, *Streptococcus pyogenes* ve diğer türler için CRISPR-Cas9 nükleazları için yeni deneysel ve hesaplamalı sonuçlar kullanıma sunuldukça, hedef dışı dizileri puanlamak için geliştirilmiş ağırlık matrislerini veya puanlama yöntemini içerecek şekilde değiştirilebilir.

3.2.1. gRNA Tanımlanması

Çeşitli kısıtlamalar kullanarak belirli bir girdi dizisini hedeflemek için CRISPR-Cas9 türevi nükleazlar için aday gRNA'ları tanımlar ve değerlendirir.

Ayrıca her gRNA için potansiyel hedef dışı dizileri raporlar ve sıralar. ***Non-TargetAnalysis*** fonksiyonu belirli bir sekansı hedeflemek için gRNA'lar ve belirli bir genomdaki hedef dışı bölgeler için özelleştirilmiş aramalar yapmak için kullanılabilir. İlgilenilen bir genomik DNA dizisi, fasta formatlı bir dosya olarak girilir. Bu aramada tanımlanan hedef ve hedef dışı bölgeleri belirlemek için çeşitli parametreler ayarlanabilir. Tanımlanan gRNA sekansları, birçok sekans analiz programında görselleştirmeye izin vermek için bir genbank dosya formatı dahil olmak üzere çeşitli formatlarda dışa aktarılabilir. ***Non-TargetAnalysis*** fonksiyonu için varsayılan parametre ayarları şunlardır:

Hedef bölgeler, gRNA'nın değişken bölgesi tarafından belirlenen kılavuz sekans ve Cas9 proteini tarafından tanınan PAM sekansından oluşan iki bileşenden oluşur. Belirli bir hedef bölge kümesi zaten tanımlanmışsa, ***determinegRNAs*** komutu False olarak ayarlanmış olarak çok sıralı bir fasta dosyası olarak girilebilirler. Hedef bölge araması için varsayılan kısıtlamalar, 20 baz kılavuz sekans ve 3 baz tercih edilen PAM sekansından (NGG) oluşan, *S. pyogenes'den* yaygın olarak kullanılan Cas9 nükleazı tarafından tanınan bölgeler için ayarlanır. PAM uzunluğu ve sekansı, PAM ve PAM_size parametreleri kullanılarak diğer türlerden Cas9 sistemleri için tercih edilen sekansları kullanmak üzere değiştirilebilir. gRNA_size parametresi, Cas9 sistemleri için farklı gRNA uzunluklarını kullanan diğer türlerden veya daha fazla özgüllük sağlayabilen kesilmiş gRNA'lar için tanımlanan kılavuz sekans uzunluğunu ayarlamak üzere değiştirilebilir. gRNA dizisi, gRNA_pattern hedef bölgedeki belirli özellikleri gerektirecek veya hariç tutacak şekilde ayarlayarak CRISPR-Cas9 sistemlerinin verimliliğini optimize etmek için de kısıtlanabilir. Örneğin, ilk baz guanin ise, gRNA'ların konakçı U6 promotörlerinden in vivo sentezi daha etkilidir ve gRNA_pattern ="^G" olarak ayarlanarak belirlenebilir.

Diğer hedeflere göre aday hedef bölgeleri filtrelenebilir. Bölünme özgüllüğünü arttırmak için CRISPR-Cas9 sistemlerinin nikaz türevleri ile uygun

şekilde yönlendirilmiş ve aralıklı hedef bölge çiftleri kullanılabilir. Bildirilen hedef bölgeler, yalnızca eşlenmiş bölgeler kurtarılacak şekilde filtrelenebilir. Bu alanlar için aralık *S. pyogenes*'den türetilen CRISPR-Cas9 sistemleri için tarif edilmiş olmakla birlikte, diğer türlerden CRISPR-Cas9 sistemleri veya CRISPR-RNA güdümlü FokI çekirdeklendirmeleri gibi yeni dimerik nükleazlar için başka boşluk kısıtlamalarının gerekli olması mümkündür. ***findPairedgRNAOnly=True*** komutu çıktıyı eşleştirilmiş sitelerle kısıtlarken, ***min_gap*** ve ***max_gap*** komutları eşleştirilmiş hedef bölgeler için alternatif boşlukların ayarlanmasına izin verir. Nükleaz kesim olaylarını izlemek için yararlı olabilecek ilave bir özellik, gRNA-Cas9 kesim sahası ile örtüşen restriksiyon enzim (RE) kesim sahalarının varlığıdır. ***findgRNAsWithREcutOnly, RepatternFile, minREpatternSize, overlap_gRNA_positions***, Cas9 bölünme bölgesinin belirli bir uzunluk tanıma alanına sahip RE bölgeleri ile çakışmasını gerektirecek şekilde ayarlanabilir.

3.2.2. Potansiyel Hedef Dışı Dizilerin Belirlenmesi

gRNA'ları seçmek için konakçı genomundaki diğer bölgelerin nükleaz tarafından parçalanma olasılığını en aza indirmektir. ***Non-TargetAnalysis***, varsayılan genomun tüm kromozomlarını varsayılan ***chromToSearch*** ayarında ("all") arar. Hedef dışı arama adımını atlamak için ("") veya örneğin belirli bir gen kümesinde aramak için ("chr1") ayarlanabilir. Potansiyel hedef dışı bölgeler için varsayılan genomik sekanslar araması, max_mismatch parametresi ile kullanıcılar tarafından değiştirilebilen maksimum 4 uyumsuzluğa izin verir. İstenirse, PAM_pattern parametresi tarafından ayarlanan hedef dışı aramalar için PAM dizisi, PAM dizilerinin varyantlarının potansiyel hedef dışı bölgeler olarak taranmasına izin vermek üzere hedef tanımlama için kullanılan PAM dizisinden farklı olabilir. Örneğin, *S. pyogenes*'ten Cas9, NGG PAM sekanslarını ve nadiren NAG'yi tanır. Belirlenen gRNA'ların özgüllüğünü arttırmak için, aday gRNA tanımlaması için

PAM = "NGG" seçeneğini belirlerken, potansiyel hedef dışı bulma için PAM_pattern = "NRG"(R: A veya G) ayarlanabilir.

Göreceli hedef dışı bölünme oranlarının tahmini, hedef dışı dizideki tekli ve çoklu uyumsuzlukların göreceli etkisini puanlamak için deneysel olarak belirlenmiş bir ceza matrisinin kullanılmıştır. Bununla birlikte, ağırlıklar parametresi, *S. pyogenes* Cas9, alternatif matrisler, diğer türlerden Cas9'lar veya kesilmiş kılavuz dizileri olan gRNA'lar için belirlenen farklı ceza matrislerini dahil etmek üzere kolayca ayarlanabilir. Öngörülen hedef dışı bölünme puanı, bölünmesi öngörülen bölgeler için maksimum 100 ile hedeflenen hedef, hedef sekansa benzemeyen bölgeler için 0'a kadar değişir.

Hedef dışı analiz sonuçları Summary.xls (hedef dizi özetleri), RecutDetails.xls (nükleaz kesim sitesinin üst üste gelen RE kesim siteleri) ve pairedSpacers.xls (potansiyel eşleştirilmiş aralayıcılar) ve Off-targetAnalysis.xls (hedef dışı diziler ve puanlar) olmak üzere dört farklı dosya çıktısı olarak kaydedilir. gRNA'lar, diğer yazılımlarla kolayca daha fazla analiz veya görselleştirmeye izin vermek için genbank ve fasta formatlarında da rapor edilir.

Çizelge 3.2**.** IUPAC nükleotid belirsizlik kodlarının baz karşılığı

IUPAC nükleotit kod	***Baz***
'Y'	*['C', 'T']*
'R'	*['A', 'G']*
'S'	*['C', 'G']*
'W'	*['A', 'T']*
'K'	*['G', 'T']*
'M'	*['A', 'C']*
'B'	*['C', 'G', 'T']*
'D'	*['A', 'G', 'T']*
'H'	*['A', 'C', 'T']*
'V'	*['A', 'C', 'G']*
'N'	*['A', 'C', 'G', 'T']*

3.2.3. pycrisprcas9 Kütüphanesi

3.2.3.1. footnoteOffTargets Fonksiyonu

Açıklama

Ekson veya intronun içinde olup olmadığını ve genin içindeyse gen kimliğini belirtmek için hedef dışı açıklamalar ekleyin.

Kullanım

```
footnoteOffTargets(
 scores,
 txdb,
 orgAnn,
 ignore_strand = True
)
```

Argümanlar

Girdi argümanı	Açıklama
scores	calculateOfftargetScore veya strainerOffTarget'tan bir veri çerçevesi çıktısı. strand (hedef dışı dizi, artı için + ve eksi için -), chrom (hedef dışı kromozom), chromStart (hedef dışı başlangıç konumu), chromEnd (hedef dışı son konumu), name (gRNA adı), gRNAPlusPAM (PAM dizisinin birleştirilmiş olduğu gRNA dizisi), OffTargetSequence (hedef dışının genomik dizisi), n.mismatch (hedef dışı ve gRNA arasındaki uyumsuzluk sayısı), forViewInUCSC (UCSC genom tarayıcısında görüntülemek için dize, ör., chr14:31665685-31665707), score (hedef dışı skor), mismatch.distance2PAM (PAM ile tüm uyumsuzlukların virgülle ayrılmış mesafeleri, ör. 14,11, bir uyumsuzluğun PAM'den 14 bp uzakta olduğu ve diğer uyumsuzluğun PAM'den 11 bp uzakta olduğu anlamına gelir), alignment (gRNA ve hedef dışı arasındaki hizalama, ör., ...G..C....., G ve C'nin uyumsuzluk olması dışında bu hedef dışı hedefin gRNA ile hizalandığı anlamına gelir), NGG (bu hedef dışı standart PAM içerir veya içermez, evet için 1 ve hayır için 0), mean.neighbor.distance.mismatch (komşu uyumsuzluklar arasındaki ortalama mesafe)
Txdb	Açıklama verileri, örneğin sıçan için TxDb.Rnorvegicus.UCSC.rn5.refGene, fare için TxDb.Mmusculus.UCSC.mm10.knownGene, insan için TxDb.Hsapiens.UCSC.hg19.knownGene, Drosophila (meyve sineği) için TxDb.Dmelanogaster.UCSC.dm3.ensGene ve C.elegans için TxDb.Celegans.UCSC.ce6.ensGene
orgAnn	İnsan için org.Hs.eg.db paketindeki org.Hs.egSYMBOL gibi organizma açıklama haritalaması

3.2.3.2. CharacteristicVectorForScoring

Açıklama

Hedef dışı puanları hesaplamak ve özellik vektörleri oluşturmak için kullanılan bir fonksiyondur.

Kullanım

```
CharacteristicVectorForScoring(
 hits,
 gRNA_size = 20,
 canonical_PAM = "NGG",
 subPAM_position = [22,23]),
 PAM_size = 3,
 PAM_location = "3prime"
)
```

Argümanlar

Girdi argümanı	Açıklama
hits	searchHits'ten oluşturulan bir veri çerçevesi içerir. IsMismatch.posX (Bu X pozisyonunun uyumsuz olup olmadığını gösteren gösterge değişkeni, 1 evet anlamına gelir ve 0 hayır anlamına gelir, X = 1-gRNA_size) kılavuz RNA'daki tüm pozisyonları temsil eder, gRNA olarak kısaltılır), strand (hedef dışı dizi, artı için + ve eksi için -), chrom (hedef dışı kromozom), chromStart (hedef dışı başlangıç konumu), chromEnd (hedef dışı son konumu), name (gRNA adı), gRNAPlusPAM (PAM dizisinin birleştirilmiş olduğu gRNA dizisi), OffTargetSequence (hedef dışının genomik dizisi), n.mismatch (hedef dışı ve gRNA arasındaki uyumsuzluk sayısı), forViewInUCSC (UCSC genom tarayıcısında görüntülemek için dize, ör., chr14:31665685-31665707), score (100 olarak ayarlanmıştır ve calculateOfftargetScore'da hesaplanacaktır)
gRNA_size	gRNA boyutu, varsayılan uzunluk 20'dir
canonical_PAM	Standart PAM, spCas9 için varsayılan NGG, Cpf1 için TTTN
subPAM_position	Alınacak alt PAM'nin başlangıç ve bitiş konumları. Tercih edilen PAM olarak 20bp gRNA ve NGG ile SP için varsayılan olarak 22 ve 23'tür
PAM_size	PAM boyutu, varsayılan olarak spCas9 için 3, Cpf1 için 4
PAM_location	gRNA'ya göre PAM konumu. Örneğin, spCas9 PAM için varsayılan olarak 3prime.cpf1 PAM için 5prime

3.2.3.3. computegRNAEfficiency

Açıklama

Belirli bir dizi dizisi ve özellik ağırlık matrisi için gRNA verimliliğini hesaplamak için kullanılan bir fonksiyondur.

Kullanım

```
computegRNAEfficiency(
 extendedSequence,
 baseBeforegRNA,
 featureWeightMatrix,
 gRNA_size = 20,
 enable_multicore = False,
 n_cores_max = 6
)
```

Argümanlar

Girdi argümanı	Açıklama
extendedSequence	gRNA artı PAM artı yan diziler içeren dizilerdir. Her dizi, featureWeightMatrix'te belirtilen özellikleri oluşturmak için yeterince uzun olmalıdır.
baseBeforegRNA	gRNA verimliliğini hesaplamak için kullanılan gRNA'dan önceki baz sayısı, varsayılan 4
featureWeightMatrix	Önemli özellikleri içeren ilk sütunu ve karşılık gelen özelliklerin ağırlığını içeren ikinci sütunu içeren bir veri çerçevesi. Aşağıdaki örnekte DoenchNBT2014 ağırlık matrisi kullanılmıştır. Kısaca, özellikler arasında INTERCEPT,GC_LOW (gRNA dizisindeki düşük GC içeriği için ceza puanı), GC_HIGH (gRNA dizisindeki yüksek GC içeriği için ceza puanı), G02 (genişletilmiş Dizinin ikinci konumunda G anlamına gelir), GT02 (GT dizisi anlamına gelir) bulunur. Genişletilmiş Dizinin 2. konumunda başlayan nükleotidler).
gRNA_size	gRNA'nın boyutu, varsayılan 20 baz uzunluğu

3.2.3.4. strainergRNAs

Açıklama

Restriksiyon enzimi kesim bölgesini içeren gRNA'ları filtreleme işlemi yapan bir fonksiyondur.

Kullanımı

```
strainergRNAs(
 all_gRNAs,
 pairOutputFile = "",
 findgRNAsWithREcutOnly = False,
 REpatternFile = os.getcwd() + os.sep + os.sep.join(["extdata", "NEBenzymes.fa"]),
 format = "fasta",
 minREpatternSize = 4,
 overlap_gRNA_positions = [17, 18],
 overlap_allpos = True
)
```

Argümanlar

Girdi argümanı	Açıklama
all_gRNAs	gRNAs, determinegRNAs çıktısı
pairOutputFile	Eşleştirilmiş gRNA'ların dosya yolu
findgRNAsWithREcutOnly	Restriksiyon enzimi tanıma modeliyle örtüşen gRNA'ların bulunup bulunmadığını belirtin
REpatternFile	Restriksiyon enzimi kesim desenlerini içeren dosya yolu
format	REpatternFile formatı, varsayılan olarak fasta
minREpatternSize	Aranacak enzim modeli için gereken minimum restriksiyon enzimi tanıma modeli uzunluğu, varsayılan 4
overlap_gRNA_positions	gRNA ve restriksiyon enzimi kesim bölgesinin gerekli örtüşme konumları, varsayılan 17 ve 18
overlap_allpos	Varsayılan True, yani yalnızca gRNA'ların tüm konumlarla örtüşmesi False olarak korunur, yani gRNA'ların konumlardan biri veya her ikisi ile örtüşmesi korunur

3.2.3.5. strainerOffTarget

Açıklama

Bu fonksiyonda minimum puan, topN gibi kullanıcılar tarafından belirlenen kriterleri karşılayan hedef dışı filtreler. Ek olarak, hedef dışı, yan sekans, gRNA bölünme verimliliği ve fetchSequence'ın True'ya ve annotateExon'un True'ya ayarlı olup olmadığı bir ekzon içinde olup olmadığı ile açıklanmıştır.

Kullanım

```
strainerOffTarget(
 scores,
 min_score = 0.01,
 topN = 200,
 topN_OfftargetTotalScore = 20,
 annotateExon = True,
 txdb,
 orgAnn,
 ignore_strand = True,
 outputDir,
 oneFilePergRNA = False,
 fetchSequence = True,
 upstream = 200,
 downstream = 200,
 BSgenomeName,
 baseBeforegRNA = 4,
 baseAfterPAM = 3,
 featureWeightMatrixFile = pd.read_csv(os.getcwd() + os.sep + os.sep.join(["extdata",
"DoenchNBT2014.csv"])),
 rule_set = "Root_RuleSet1_2014",
 calculategRNAefficacyForOfftargets = True
)
```

Argümanlar

Girdi argümanı	Açıklama
scores	calculateOfftargetScore'dan bir veri çerçevesi çıktısı. strand (hedef dışı dizi, artı için + ve eksi için -), chrom (hedef dışı kromozom), chromStart (hedef dışı başlangıç konumu), chromEnd (hedef dışı son konumu), name (gRNA adı), gRNAPlusPAM (PAM dizisinin birleştirilmiş olduğu gRNA dizisi), OffTargetSequence (hedef dışının genomik dizisi), n.mismatch (hedef dışı ve gRNA arasındaki uyumsuzluk sayısı), forViewInUCSC (UCSC genom tarayıcısında görüntülemek için dize, ör., chr14:31665685-31665707), score (hedef dışı skor), mismatch.distance2PAM (PAM ile tüm uyumsuzlukların virgülle ayrılmış mesafeleri, ör. 14,11, bir uyumsuzluğun PAM'den 14 bp uzakta olduğu ve diğer uyumsuzluğun PAM'den 11 bp uzakta olduğu anlamına gelir), alignment (gRNA ve hedef dışı arasındaki hizalama, ör., ...G..C....., G ve C'nin uyumsuzluk olması dışında bu hedef dışı hedefin gRNA ile hizalandığı anlamına gelir), NGG (bu hedef dışı standart PAM içerir veya içermez, evet için 1 ve hayır için 0), mean.neighbor.distance.mismatch (komşu uyumsuzluklar arasındaki ortalama mesafe)

min_score	son çıktıya dahil olacak hedef dışı miimumi puan, varsayılan 0,5
topN	Nihai çıktıya dahil edilecek ilk N hedef dışı, varsayılan 100
topN_OfftargetTotalScore	Toplam hedef dışı puanı hesaplamak için kullanılan ilk N hedef dışı, varsayılan 10
annotateExon	Hedefin bir ekzon içinde olup olmadığının gösterilip gösterilmeyeceğini belirlenmesi, varsayılan True
txdb	Açıklama verileri, örneğin sıçan için TxDb.Rnorvegicus.UCSC.rn5.refGene, fare için TxDb.Mmusculus.UCSC.mm10.knownGene, insan için TxDb.Hsapiens.UCSC.hg19.knownGene, Drosophila (meyve sineği) için TxDb.Dmelanogaster.UCSC.dm3.ensGene ve C.elegans için TxDb.Celegans.UCSC.ce6.ensGene
orgAnn	insan için org.Hs.eg.db paketindeki org.Hs.egSYMBOL gibi organizma açıklama haritalaması
outputDir	hedef dışı analiz ve raporların yazılacağı dizin
oneFilePergRNA	her gRNA için bir dosyaya yaz veya yazma, varsayılan olarak False
fetchSequence	Hedef dışı yan sıra dizisini getir veya getirme, varsayılan True
upstream	hedef dışı başlangıçtan yukarı akış ofseti, varsayılan 200
downstream	hedef dışı uçtan aşağı akış ofseti, varsayılan 200
BSgenomeName	Örneğin, hg19 için BSgenome.Hsapiens.UCSC.hg19, mm10 için BSgenome.Mmusculus.UCSC.mm10, ce6 için BSgenome.Celegans.UCSC.ce6, rn5 için BSgenome.Rnorvegicus.UCSC.rn5 ve BSgenome.UCDmelanogaster. dm3 için dm3
baseBeforegRNA	gRNA verimliliğini hesaplamak için kullanılan gRNA'dan önceki baz sayısı, varsayılan 4
baseAfterPAM	gRNA verimliliğini hesaplamak için kullanılan PAM'den sonraki baz sayısı, varsayılan 3
featureWeightMatrixFile	gRNA verimliliğini hesaplamak için kullanılan özellik ağırlık matrisi dosyası. Varsayılan olarak DoenchNBT2014 ağırlık matrisi kullanılır. Alternatif ağırlık matrisi dosyasını kullanmak için, ilk sütunda önemli özellikleri ve ikinci sütunda özelliklere karşılık gelen ağırlıkları içeren bir csv dosyası girin.
rule_set	gRNA etkinliğini hesaplamak için bir kural seti skorlama sistemi.
calculategRNAefficacyForOfftargets	Hedef dışı ve aynı zamanda hedefler için gRNA etkinliğinin çıktısını almak için varsayılan olarak True'ya ayarlayın. Yalnızca analizi hızlandırmak için hedefler için hesaplanan gRNA etkinliğine ihtiyaç duyuyorsanız, bunu False olarak ayarlayın.

3.2.3.6. determinegRNAs

Açıklama

Fasta formatında dizileri içeren bir giriş dosyası için potansiyel gRNA'ları bulur.

Kullanım

```
determinegRNAs(
 inputFilePath,
 baseEditing = False,
 targetBase = "C",
 editingWindow = [3,4,5,6,7,8],
 format = "fasta",
 PAM = "NGG",
 PAM_size = 3,
 findPairedgRNAOnly = False,
 annotatePaired = True,
 paired_orientation = "PAMout",
```

```
 enable_multicore = False,
 gRNA_pattern = "",
 gRNA_size = 20,
 overlap_gRNA_positions = [17,18],
 primeEditing = False,
 PBS_length = 13,
 RT_template_length =
[7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28],
 RT_template_pattern = "D$",
 corrected_seq,
 targeted_seq_length_change,
 bp_after_target_end = 15,
 target_start, target_end,
 primeEditingPaired_output = "pairedgRNAsForPE.xls",
 min_gap = 0,
 max_gap = 20,
 pairOutputFile,
 name_prefix = "",
 featureWeightMatrixFile = pd.read_csv(os.getcwd() + os.sep +
os.sep.join(["extdata", "DoenchNBT2014.csv"])),
 baseBeforegRNA = 4,
 baseAfterPAM = 3,
 calculategRNAEfficacy = False,
 efficacyFile,
 PAM_location = "3prime",
 rule_set = "Root_RuleSet1_2014"
)
```

Argümanlar

Girdi argümanı	Açıklama
inputFilePath	Potansiyel gRNA'lar için aranacak dizileri içeren dizi giriş dosyası yolu veya bir
baseEditing	Temel düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtilmesi. Varsayılan olarak False. Eğer True ise baseEditing = True, targetBase ve editingWidow'u buna göre ayarlayın.
targetBase	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. CBE sisteminde C'yi T'ye dönüştürmek için varsayılan olarak C'dir. ABE sisteminde A olarak ayarlanır.
editingWindow	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Orijinal CBE sistemi için varsayılan olarak 4 ila 8 olan etkin düzenleme penceresini belirtmek için kullanılır. Kullandığınız sistem farklı bir düzenleme penceresine sahipse buna göre değiştirin.
format	Giriş dosyasının formatı, fasta ve fastq desteklenir, varsayılan fasta
PAM	gRNA'dan sonra protospacer-bitişik motif (PAM) dizisi, varsayılan NGG
PAM_size	PAM uzunluğu, varsayılan 3
findPairedgRNAOnly	Eşleştirilmiş gRNA'ları yalnızca, ilki ters gRNA adı verilen eksi dizide ve ikincisi ileri gRNA adı verilen artı dizide olacak şekilde aranıp aranmayacağını seçin. True veya False, varsayılan False
annotatePaired	Eşleştirilmiş bilgi çıkışının yapılıp yapılmayacağını belirtin, varsayılan True
paired_orientation	PAMin yönelimi, duyu ve duyu olmayan ipliklerdeki iki bitişik PAM'ın N21GG ve CCN21 gibi birbirine doğru baktığı anlamına gelirken, PAMout yönelimi, CCN21 ve N21GG gibi birbirlerinden uzağa baktıkları anlamına gelir.
gRNA_size	gRNA'nın boyutu, varsayılan uzunluk 20
overlap_gRNA_positions	gRNA ve restriksiyon enzimi kesim bölgesinin gerekli örtüşme konumları, varsayılan 17 ve 18
primeEditing	Birincil düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtin. Varsayılan olarak False'tır. True ise PBS_length, RT_template_length, RT_template_pattern,

	target_seq_length_change, bp_after_target_end, target_start ve target_end'i uygun şekilde ayarlanır
PBS_length	Yalnızca primeEditing gerçekleştirildiğinde uygulanabilir. Astar bağlanma bölgesi için çıktının sayısını belirlemek için kullanılır.
RT_template_length	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablonu için gereken taban sayısını belirtmek için kullanılır, varsayılan olarak 8'den 18'e kadardır. Düzenleme büyük ekleme ise uzunluğu artırın. Çıktıda yalnızca hesaplanmış RT_template_length belirtilen aralığa düşen gRNA'lar olacaktır. Aşağıdaki gibi hesaplanır. RT_template_length = target_start – cut_start + (target_end – target_start) targeted_seq_length_change + bp_after_target_end
RT_template_pattern	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. https://doi.org/10.1038/s41586-019-1711-4'e göre varsayılan olarak C ile bitmeyen RT şablon dizi modelini belirtmek için kullanılır.
corrected_seq	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. Başarılı düzenlemeden sonra mutasyona uğramış veya eklenen dizileri belirtmek için kullanılır.
targeted_seq_length_change	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. Hedeflenen dizi uzunluğu değişikliğinin sayısını belirtmek için kullanılır. Taban değişiklikleri için 0, ekleme için pozitif sayılar ve silme için negatif sayı olarak ayarlayın. Örneğin, 10, düzeltilmiş dizinin 10bp eklemeye sahip olacağı anlamına gelir, -10, düzeltilmiş dizinin 10bp silmeye sahip olacağı anlamına gelir ve 0, yalnızca bazların değiştirildiği ve dizi uzunluğunun aynı kaldığı anlamına gelir.
bp_after_target_end	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablonunun bir parçası olarak hedef değişikliği bitiş sitesinden sonra eklenecek baz sayısını belirtmek için kullanılır.
target_start	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablon dizisini elde etmek için kullanılacak değişiklikleri yapmak için giriş dizisindeki başlangıç konumunu belirtmek için kullanılır
target_end	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablon dizisini elde etmek için kullanılacak değişiklikleri yapmak için giriş dizisindeki bitiş konumunu belirtmek için kullanılır.
primeEditingPaired_output	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. PBS, RT_template, gRNA dizileri, varsayılan pairedgRNAsForPE.xls
min_gap	Geçerli eşleştirilmiş gRNA'lar olması için zıt yönlü iki gRNA arasındaki minimum mesafe. Varsayılan 0
max_gap	Geçerli eşleştirilmiş gRNA'lar olmak için zıt yönlü iki gRNA arasındaki maksimum mesafe. Varsayılan 20
pairOutputFile	Eşleştirilmiş gRNA bilgilerini yazmak için çıktı dosyası
name_prefix	Bulunan gRNA'lara ad atanırken kullanılan önek, varsayılan gRNA, kılavuzlu RNA'nın kısaltması.
baseBeforegRNA	gRNA verimliliğini hesaplamak için kullanılan gRNA'dan önceki baz sayısı, spCas9 için varsayılan 4'tür. 5 üssünde bulunan PAM için, PAM dizisi artı PAM boyutundan önceki baz sayısını belirtmeniz gerektir.
baseAfterPAM	gRNA verimliliğini hesaplamak için kullanılan PAM'ten sonraki baz sayısı, spCas9 için varsayılan 3,
featureWeightMatrixFile	gRNA verimliliğini hesaplamak için kullanılan özellik ağırlık matrisi dosyası. Varsayılan olarak DoenchNBT2014 ağırlık matrisi kullanılır. Alternatif ağırlık matrisi dosyasını kullanmak için, ilk sütunda önemli özellikleri ve ikinci sütunda özelliklere karşılık gelen ağırlıkları içeren bir csv dosyası girin.
calculategRNAEfficacy	Varsayılan olarak False'tır, gRNA etkinliğini hesaplamaz
efficacyFile	gRNA etkinliklerini yazmak için dosya yolu
PAM_location	gRNA'ya göre PAM konumu. Örneğin, spCas9 PAM 3 üssünde bulunurken cpf1 PAM 5 üssünde bulunur
rule_set	gRNA etkinliğini hesaplamak için bir kural kümesi puanlama sistemi belirtin.

3.2.3.7. calculateOfftargetScore

Açıklama

Verilen özellik vektörleri ve ağırlık vektörü ile her bir hedef dışı hedef için puanı hesaplar.

Kullanım

```
calculateOfftargetScore(
 featureVectors,
 weights = [0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613,
0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583]
)
```

Argümanlar

Girdi argümanı	Açıklama
featureVectors	CharacteristicVectorForScoring'den oluşturulan bir veri çerçevesi. gRNA'daki tüm pozisyonları temsil eden IsMismatch.posX (X pozisyonunun uyumsuz olup olmadığını gösteren gösterge değişkeni, 1 evet anlamına gelir ve 0 hayır anlamına gelir, X = 1-gRNA.size) içerir), strand (hedef dışı dizi, artı için + ve eksi için -), chrom (hedef dışı kromozom), chromStart (hedef dışı başlangıç konumu), chromEnd (hedef dışı son konumu), name (gRNA adı), gRNAPlusPAM (PAM dizisinin birleştirilmiş olduğu gRNA dizisi), OffTargetSequence (hedef dışının genomik dizisi), n.mismatch (hedef dışı ve gRNA arasındaki uyumsuzluk sayısı), forViewInUCSC (UCSC genom tarayıcısında görüntülemek için dize, ör., chr14:31665685-31665707), score (hedef dışı skor), mismatch.distance2PAM (PAM ile tüm uyumsuzlukların virgülle ayrılmış mesafeleri, ör. 14,11, bir uyumsuzluğun PAM'den 14 bp uzakta olduğu ve diğer uyumsuzluğun PAM'den 11 bp uzakta olduğu anlamına gelir), alignment (gRNA ve hedef dışı arasındaki hizalama, ör., ...G..C....., G ve C'nin uyumsuzluk olması dışında bu hedef dışı hedefin gRNA ile hizalandığı anlamına gelir), NGG (bu hedef dışı standart PAM içerir veya içermez, evet için 1 ve hayır için 0), mean.neighbor.distance.mismatch (komşu uyumsuzluklar arasındaki ortalama mesafe)
weights	gRNA uzunluğunun sayısal bir vektör boyutu, varsayılan [0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583]

3.2.3.8. UniquePatterns

Açıklama

Bir dizi ve model listesi girin ve modellerin dizide yalnızca bir kez oluşup oluşmadığını belirleyin. gRNA'daki bir RE bölgesinin yan bölgede de oluşup oluşmadığını belirlemek için kullanılır.

Kullanım

```
UniquePatterns(
 seq,
 patterns
)
```

Argümanlar

Girdi argümanı	Açıklama
seq	bir gRNA'nın yan dizisi
patterns	RE sitelerinin bir listesi gibi DNAStringSet gibi desenler

3.2.3.9. Non-TargetAnalysis

Açıklama

Hedefe özel kılavuz RNA'ların (gRNA'lar) tasarımı ve otomatik olarak determinegRNA'ları, strainergRNA'ları, searchingHits, CharacteristicVectorForScoring, calculateOfftargetScore, strainerOffTarget'ı çağırarak, gRNA bölünme verimliliğini hesaplayarak ve gRNA etkinliğini, CRISPRCas9 sistemi için tahmin etmektedir.

Kullanım

```
Non-TargetAnalysis(
 İnputFilePath,
 outputDir,
 BSgenomeName,
 Txdb,
 orgAnn,
 chrom_acc,
 corrected_seq,
 target_start,
 target_end,
 targeted_seq_length_change,
 gRNAoutputName,
 format = "fasta",
 header = False,
 determinegRNAs = True,
 exportAllgRNAs = "all",
 determinegRNAsWithREcutOnly = False,
 annotatePaired = True,
 REpatternFile = os.getcwd() + os.sep + os.sep.join(["extdata", "NEBenzymes.fa"]),
 minREpatternSize = 4,
 overlap_gRNA_positions = [17, 18],
 findPairedgRNAOnly = False,
 paired_orientation = "PAMout",
 min_gap = 0,
 max_gap = 20,
 gRNA_name_prefix = "",
 PAM_size = 3,
 gRNA_size = 20,
 PAM = "NGG",
 chromToSearch = ["all"],
 chromToExclude    =    ["chr17_ctg5_hap1",    "chr4_ctg9_hap1",    "chr6_apd_hap1",
chr6_cox_hap2", "chr6_dbb_hap3", "chr6_mann_hap4", "chr6_mcf_hap5", "chr6_qbl_hap6",
"chr6_ssto_hap7"],
```

```
 max_mismatch = 3,
 PAM_pattern = "NNG$",
 allowed_mismatch_PAM = 1,
 gRNA_pattern = "",
 baseEditing = False,
 targetBase = "C",
 editingWindow = [4, 8],
 editingWindow_offtargets = [4, 8],
 primeEditing = False,
 PBS_length = 13,
 RT_template_length = [8, 28],
 RT_template_pattern = "D$",
 bp_after_target_end = 15,
 primeEditingPaired_output = "pairedgRNAsForPE.xls",
 min_score = 0,
 topN = 1000,
 topN_OfftargetTotalScore = 10,
 annotateExon = True,
 ignore_strand = True,
 fetchSequence = True,
 upstream = 200,
 downstream = 200,
 upstream_search = 0,
 downstream_search = 0,
 weights = [0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613,
0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583],
 baseBeforegRNA = 4,
 baseAfterPAM = 3,
 featureWeightMatrixFile = pd.read_csv(os.getcwd() + os.sep + os.sep.join(["extdata",
"DoenchNBT2014.csv"])),
 useEfficacyFromInputSeq = False,
 outputUniqueREs = True,
 gRNA_backbone =
"GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU
",
 temperature = 37,
 overwrite = False,
 scoring_method = "Hsu-Zhang",
 subPAM_activity = {"AA" : 0,
                    "AC" : 0,
                    "AG" : 0.259259259,
                    "AT" : 0,
                    "CA" : 0,
                    "CC" : 0,
                    "CG" : 0.107142857,
                    "CT" : 0,
                    "GA" : 0.069444444,
                    "GC" : 0.022222222,
                    "GG" : 1,
                    "GT" : 0.016129032,
                    "TA" : 0,
                    "TC" : 0,
                    "TG" : 0.038961039,
                    "TT" : 0},
 subPAM_position = [22, 23],
 PAM_location = "3prime",
 rule_set = "Root_RuleSet1_2014",
```

```
 calculategRNAefficacyForOfftargets = True,
 mismatch_activity_file = pd.read_csv(os.getcwd() + os.sep + os.sep.join(["extdata",
"NatureBiot2016SuppTable19DoenchRoot.csv"])),
 method_indelFreq = "Lindel",
 baseBeforegRNA_indelFreq = 13,
 baseAfterPAM_indelFreq = 24
)
```

Argünamlar

Girdi argümanı	Açıklama
inputFilePath	Potansiyel gRNA'lar için aranacak dizileri içeren dizi giriş dosyası yolu veya bir
format	Giriş dosyasının formatı, fasta, fastq ve bed desteklenmesi, varsayılan fasta
header	Giriş dosyasının başlık içerip içermediğini belirtin, varsayılan False, yalnızca bed biçimi için geçerlidir
gRNAoutputName	inputFilePath dosya yolu yerine DNAStringSet nesnesi olduğunda gRNA çıktı dosyasının adını belirtin
determinegRNAs	Girdi dosyasındaki dizilerden gRNA'ları bulup bulmayacağınızı belirtin veya gRNA'ları bulma adımını atlayın, varsayılan True. Giriş dosyası, kullanıcı tarafından seçilen gRNA'ları ve zaten PAM'ı içeriyorsa, bunu False olarak ayarlayın.
exportAllgRNAs	Tüm potansiyel gRNA'ların bir dosyaya fasta formatında mı, genbank formatında mı yoksa her ikisinde mi çıktı alınacağını belirtin. Her ikisi için de varsayılan.
findgRNAsWithREcutOnly	Restriksiyon enzimi tanıma modeliyle örtüşen gRNA'ların bulunup bulunmadığını belirtin
REpatternFile	Restriksiyon enzimi kesim kalıplarını içeren dosya yolu
minREpatternSize	Aranacak enzim modeli için gereken minimum kısıtlama enzimi tanıma modeli uzunluğu, varsayılan 4
overlap_gRNA_positions	gRNA ve kısıtlama enzimi kesim bölgesinin gerekli örtüşme konumları, varsayılan 17 ve 18
findPairedgRNAOnly	Eşleştirilmiş gRNA'ları yalnızca, ilki ters gRNA adı verilen eksi dizide ve ikincisi ileri gRNA adı verilen artı dizide olacak şekilde aranıp aranmayacağını seçin. True veya False, varsayılan False
annotatePaired	Eşleştirilmiş bilgi çıkışının yapılıp yapılmayacağını belirtin, varsayılan True
paired_orientation	PAMin yönelimi, duyu ve duyu olmayan ipliklerdeki iki bitişik PAM'ın N21GG ve CCN21 gibi birbirine doğru baktığı anlamına gelirken, PAMout yönelimi, CCN21 ve N21GG gibi birbirlerinden uzağa baktıkları anlamına gelir.
min_gap	Geçerli eşleştirilmiş gRNA'lar olması için zıt yönlü iki gRNA arasındaki minimum mesafe. Varsayılan 0
max_gap	Geçerli eşleştirilmiş gRNA'lar olmak için zıt yönlü iki gRNA arasındaki maksimum mesafe. Varsayılan 20
gRNA_name_prefix	Bulunan gRNA'lara ad atanırken kullanılan önek, varsayılan gRNA, kılavuzlu RNA'nın kısaltması.
PAM_size	PAM uzunluğu, varsayılan 3
gRNA_size	gRNA'nın boyutu, varsayılan uzunluk 20
PAM	gRNA'dan sonra PAM dizisi, varsayılan NGG
BSgenomeName	Örneğin, hg19 için BSgenome.Hsapiens.UCSC.hg19, mm10 için BSgenome.Mmusculus.UCSC.mm10, ce6 için BSgenome.Celegans.UCSC.ce6, rn5 için BSgenome.Rnorvegicus.UCSC.rn5, Zv9 için BSgenome.Drerio.UCSCdan ve dm3 için BSgenome.Dmelanogaster.UCSC.dm3
chromToSearch	Aranacak kromozomu belirtin, varsayılan olarak tümü, yani tüm kromozomları arayın. Örneğin, ["chrX"] yalnızca X kromozomunda eşleşme aramayı belirtir
chromToExclude	Arama yapılmayacak kromozomu belirtin. "" olarak belirtilirse, chromToSearch tarafından belirtilen kromozomları aramak anlamına gelir. Varsayılan olarak, haplotip bloklarını hg19'da hedef dışı aramadan hariç tutmak için, yani chromToExclude = ["chr17_ctg5_hap1","chr4_ctg9_hap1", "chr6_apd_hap1",

	"chr6_cox_hap2", "chr6_dbb_hap3", "hapr4", "_hapch_bl", "chr4", "_hapch_bl", "chr6_ssto_hap7"]
max_mismatch	Hedef dışı aramada izin verilen maksimum uyumsuzluğa, varsayılan 3. Uyarı > 3 olarak ayarlanırsa önemli ölçüde daha yavaş olacaktır.
PAM_pattern	Protospacer-bitişik motifin (PAM) normal ifadesi, spCas9 için varsayılan NNG$\|NGN$. cpf1 için ^TTTN
allowed_mismatch_PAM	Hedef dışı arama için PAM dizisinde izin verilen maksimum yanlış eşleşme sayısı, hedef dışı tanımlama için NGN ve NNG PAM modeline izin vermek için varsayılan olarak 1'dir.
gRNA_pattern	Normal ifade veya gRNA modelini temsil etmek için IUPAC Genişletilmiş Genetik Alfabesi, varsayılan kısıtlama yoktur. Örneğin gRNA'nın GG ile başlaması gerektiğini belirtmek için, onu ^GG olarak ayarlayın.
baseEditing	Temel düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtin. Varsayılan olarak False'a ayarlayın. True ise, baseEditing = True, targetBase ve editingWidow'u buna göre ayarlayın.
targetBase	Yalnızca baseditingTrue olarak ayarlandığında uygulanabilir. Temel düzenleme sistemlerinin hedef tabanını belirtmek için, CBE sisteminde C'yi T'ye dönüştürmek için C olarak ayarlanır. ABE sisteminde A olarak ayarlanır. Varsayılan C'dir.
editingWindow	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Orijinal CBE sistemi için varsayılan olarak 4 ila 8 olan etkin düzenleme penceresini belirtmek için kullanılır. Kullandığınız sistem farklı bir düzenleme penceresine sahipse buna göre değiştirin
editingWindow.offtargets	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Yalnızca hedef dışı arama için dikkate alınacak etkili düzenleme penceresini belirtmek için kullanılır, varsayılan olarak 4 ila 8'dir (1, 3' PAM'den en uzak site, 5' PAM'den en yakın bölge anlamına gelir), orijinal CBE sistemi. Kullandığınız sistemin farklı bir düzenleme penceresi varsa veya hedef tabanlı hedef dışı hedefleri daha büyük bir düzenleme penceresine dahil etmek istiyorsanız, buna göre değiştirin.
primeEditing	Birincil düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtin. Varsayılan olarak False'tır. True ise PBS_length, RT_template_length, RT_template_pattern, target_seq_length_change, bp_after_target_end, target_start ve target_end' uygun şekilde ayarlanır.
PBS_length	Yalnızca primeEditing gerçekleştirildiğinde uygulanabilir. Astar bağlanma bölgesi için çıktının sayısını belirlemek için kullanılır.
RT_template_length	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablonu için gereken taban sayısını belirtmek için kullanılır, varsayılan olarak 8'den 18'e kadardır. Düzenleme büyük ekleme ise uzunluğu artırın. Çıktıda yalnızca hesaplanmış RT.template.length belirtilen aralığa düşen gRNA'lar olacaktır. Aşağıdaki gibi hesaplanır. RT_template_length = target_start – cut_start + (target_end – target_start) target_seq_length_change + bp_after_target_end
RT_template_pattern	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. https://doi.org/10.1038/s41586-019-1711-4'e göre varsayılan olarak C ile bitmeyen RT şablon dizi modelini belirtmek için kullanılır.
Corrected_seq	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. Başarılı düzenlemeden sonra mutasyona uğramış veya eklenen dizileri belirtmek için kullanılır.
targeted_seq_length_change	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. Hedeflenen dizi uzunluğu değişikliğinin sayısını belirtmek için kullanılır. Taban değişiklikleri için 0, ekleme için pozitif sayılar ve silme için negatif sayı olarak ayarlanır. Örneğin, 10, düzeltilmiş dizinin 10bp eklemeye sahip olacağı anlamına gelir, -10, düzeltilmiş dizinin 10bp silmeye sahip olacağı anlamına gelir ve 0, yalnızca bazların değiştirildiği ve dizi uzunluğunun aynı kaldığı anlamına gelir.
bp_after_target_end	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablonunun bir parçası olarak hedef değişikliği bitiş sitesinden sonra eklenecek baz sayısını belirtmek için kullanılır.
target_start	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablon dizisini elde etmek için kullanılacak değişiklikleri yapmak için giriş dizisindeki başlangıç konumunu belirtmek için kullanılır.

target_end	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. RT şablon dizisini elde etmek için kullanılacak değişiklikleri yapmak için giriş dizisindeki bitiş konumunu belirtmek için kullanılır.
primeEditingPaired_output	Yalnızca primeEditing, True olarak ayarlandığında geçerlidir. PBS, RT_template, gRNA dizileri, varsayılan pairedgRNAsForPE.xls
min_score	Nihai çıktıya dahil edilecek hedef dışı minimum puan, varsayılan 0
topN	Nihai çıktıya dahil edilecek ilk N hedef dışı puanı, varsayılan 1000
topN_OfftargetTotalScore	toplam hedef dışı puanı hesaplamak için kullanılan ilk N hedef dışı, varsayılan 10
annotateExon	Hedef dışının bir ekzon içinde olup olmadığının gösterilip gösterilmeyeceğini seçin, varsayılan True
txdb	Açıklama verileri, örneğin sıçan için TxDb.Rnorvegicus.UCSC.rn5.refGene, fare için TxDb.Mmusculus.UCSC.mm10.knownGene, insan için TxDb.Hsapiens.UCSC.hg19.knownGene, Drosophila için TxDb.Dmelanogaster.UCSC.dm3.ensGene ve C.elegans için TxDb.Celegans.UCSC.ce6.ensGene
orgAnn	insan için org.Hs.eg.db paketindeki org.Hs.eg gibi organizma açıklama haritalaması
outputDir	hedef dışı analiz ve raporların yazılacağı dizin
fetchSequence	Hedef dışı veya değil, yan sıra sekansları getir, varsayılan True
upstream	hedef dışı başlangıçtan yukarı akış ofseti, varsayılan 200
downstream	hedef dışı uçtan aşağı akış ofseti, varsayılan 200
upstream_search	bed girişinden yukarı akış ofseti gRNA'ları aramaya başlar, varsayılan 0
downstream_search	bed girişinden aşağı akış ofseti gRNA'ları aramaya başlar, varsayılan 0
weights	Yalnızca puanlama.yöntemi Hsu-Zhang olarak ayarlandığında, gRNA uzunluğunun sayısal bir vektör boyutu, varsayılan [0, 0, 0,014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583] geçerlidir.
baseBeforegRNA	gRNA verimliliğini hesaplamak için kullanılan gRNA'dan önceki baz sayısı, varsayılan 4. 5 üssünde bulunan PAM için, PAM dizisi artı PAM boyutundan önceki baz sayısını belirtmeniz gerekir.
baseAfterPAM	gRNA verimliliğini hesaplamak için kullanılan PAM'ten sonraki üs sayısı, spCas9 için varsayılan 3, 5prime'da bulunan PAM için, gRNA'nın uzunluğunu artı 3 asal üzerindeki genişletilmiş sırayı içermesi gerekir.
featureWeightMatrixFile	gRNA verimliliğini hesaplamak için kullanılan özellik ağırlık matrisi dosyası. Varsayılan olarak DoenchNBT2014 ağırlık matrisi kullanılır. Alternatif ağırlık matrisi dosyasını kullanmak için, ilk sütunda önemli özellikleri ve ikinci sütunda özelliklere karşılık gelen ağırlıkları içeren bir csv dosyası girin.
useEfficacyFromInputSeq	Varsayılan False. True olarak ayarlanırsa özet dosyası, hedef dışı analiz yerine giriş dizilerinden hesaplanan gRNA etkinliğini içerecektir. Giriş dizisi, hedef dışı analiz için kullanılandan farklı bir türden ise True olarak ayarlayın.
outputUniqueREs	Varsayılan True. True olarak ayarlanırsa özet dosyası, gRNA dizisini çevreleyen 100 veya 200 baz içinde bölünme bölgesine özgü RE'ler içerecektir.
gRNA_backbone	gRNA omurga sabit bölge dizisi. Sp gRNA omurgasındaki dizi için varsayılan
temperature	santigrat cinsinden sıcaklık. Varsayılan 37 santigrat.
scoring_method	Hedef dışı bölünme oranı tahmini için hangi yöntemin kullanılacağını gösterir, şu anda iki yöntem desteklenmektedir, Hsu-Zhang ve CFDscore
subPAM_activity	Yalnızca puanlama yöntemi, tercih edilen PAM dizisine göre her bir alternatif alt PAM dizisi için bölünme oranını temsil etmek üzere CFDscore A karma değerine ayarlandığında geçerlidir
subPAM_position	Yalnızca scoring_method CFDscore olarak ayarlandığında geçerlidir Alt PAM'nin başlangıç ve bitiş konumları. Tercih edilen PAM olarak 20bp gRNA ve NGG ile spCas9 için varsayılan olarak 22 ve 23'tür. Cpf1 için [1,2] olabilir.
PAM_location	gRNA'ya göre PAM konumu. Örneğin, spCas9 PAM için varsayılan olarak 3prime. cpf1 PAM için 5prime olarak ayarlanır.
rule_set	gRNA etkinliğini hesaplamak için bir kural kümesi puanlama sistemi
calculategRNAefficacyForOfftargets	Hedef dışı ve aynı zamanda hedefler için gRNA etkinliğinin çıktısını almak için varsayılan olarak True ayarlanır. Yalnızca analizi hızlandırmak için hedefler için hesaplanan gRNA etkinliğine ihtiyaç duyuyorsanız, bunu False olarak ayarlanır.

mismatch_activity_file	Yalnızca scoring_method CFDscore olarak ayarlandığında geçerlidir gRNA'nın her pozisyonundaki olası tüm tek nükleotid uyumsuzluğu türleri için bölünme oranlarını içeren (csv) dosya.

3.2.3.10. searchingHits2

Açıklama

Verilen gRNA'lar, BSgenome ve maksimum uyumsuzluklar için hedef dışı arama yapmaktadır.

Kullanım

```
searchingHits2(
 gRNAs,
 BSgenomeName,
 chromToSearch = "all",
 chromToExclude = "",
 max_mismatch = 0,
 PAM_size = 3,
 gRNA_size = 20,
 PAM = "NGG",
 PAM_pattern = "NGG",
 allowed_mismatch_PAM = 1,
 PAM_location = "3prime",
 BaseEditing = False,
 targetBase = "C",
 editingWindow = [4,5,6,7,8]
)
```

Argümanlar

Girdi argümanı	Açıklama
gRNAs	gRNA içeren string nesnesi. Dizilerin gRNA'lardan sonra eklenen PAM'ı içermesi gerekir. Örneğin ATCGAAATTCGAGCCAATCCCGG, burada CGG PAM'dir
BSgenomeName	Örneğin, hg19 için BSgenome.Hsapiens.UCSC.hg19, mm10 için BSgenome.Mmusculus.UCSC.mm10, ce6 için BSgenome.Celegans.UCSC.ce6, rn5 için BSgenome.Rnorvegicus.UCSC.rn5 ve dm3 için BSgenome.UCDmelanogaster.dm3
chromToSearch	Aranacak kromozomu belirtin, varsayılan olarak ["all"], yani tüm kromozomları arayın. Örneğin, ["chrX"] yalnızca X kromozomunda eşleşme aramayı belirtir
chromToExclude	chromToSearch tarafından belirtilen kromozomları aramak için, aranmayacak, varsayılan olarak hiçbiri olarak kromozomu belirtin. Örneğin, haplotip bloklarını hg19'da hedef dışı aramadan hariç tutmak için, chromToExclude öğesini ["chr17_ctg5_hap1", "chr4_ctg9_hap1", "chr6_apd_hap1", "chr6_cox_hap2", "chr6_dbb_hap3", "chrmc_dbb_hap3", "chrmc6_mann_hap4","chrmc6_mann_hap1", "chr6_ssto_hap7"] olarak ayarlanır
max_mismatch	Hedef dışı aramada izin verilen maksimum uyumsuzluğa, varsayılan 3.
PAM_size	PAM boyutu, varsayılan uzunluk 3
gRNA_size	gRNA boyutu, varsayılan uzunluk 20
PAM	Protospacer-bitişik motifin (PAM) düzenli ifadesi, spCas9 için varsayılan NGG. cpf1 için ^TTTN
PAM_pattern	PAM'nin normal ifadesi, spCas9 için varsayılan NGG. cpf1 için ^TTTN
allowed_mismatch_PAM	PAM dizisindeki dejeneratif bazların sayısı, NGG PAM için varsayılan olarak 1'dir

PAM_location	gRNA'ya göre PAM konumu. Örneğin, spCas9 PAM 3 üssünde bulunurken cpf1 PAM 5 üssünde bulunur
baseEditing	Temel düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtin. Varsayılan olarak False olarak ayarlayın True ise, baseEditing = True, targetBase ve editingWidow buna göre ayarlanır.
targetBase	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Temel düzenleme sistemleri için hedef tabanı belirtmek için kullanılır, CBE sisteminde C'yi T'ye dönüştürmek için varsayılan olarak C'dir. ABE sistemini kullanmayı düşünüyorsanız A olarak değiştirin.
editingWindow	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Yalnızca hedef dışı arama için dikkate alınacak etkili düzenleme penceresini belirtmek için kullanılır, orijinal CBE sistemi için varsayılan olarak 4 ila 8'dir. Kullandığınız sistemin farklı bir düzenleme penceresi varsa veya hedef tabanlı hedef dışı hedefleri daha büyük bir düzenleme penceresine dahil etmek istiyorsanız, buna göre değiştirilir.

3.2.3.11. convertPattern

Açıklama

IUPAC nükleotid belirsizlik kodlarını içeren kalıbı düzenli ifadeye çevirmektedir. Örneğin, 'Y' = ['C', 'T'], 'R' = ['A', 'G'], 'S' = ['G', 'C'], 'W' = ['A', 'T'], 'K' = ['T', 'U', 'G'], 'M' = ['A', 'C'], 'B' = ['C', 'G', 'T'], 'D' = ['A', 'G', 'T'], 'H' = ['A', 'C', 'T'], 'V' = ['A', 'C', 'G'] ve 'N' = ['A', 'C', 'T', 'G'].

Kullanım

```
convertPattern(pattern)
```

Argümanlar

Girdi argümanı	Açıklama
pattern	IUPAC nükleotid belirsizlik kodlarına sahip bir karakter vektörü

3.2.3.12. uniqueRestrictionEnzymes

Açıklama

Tanımlanan her gRNA için, yalnızca gRNA bölünme bölgelerini tanıyan çıkış restriksiyon enzimleri bulmaktadır.

Kullanım

```
uniqueRestrictionEnzymes(
 REcutDetails,
 BSgenomeName,
 summary,
 offTargets,
 scanUpstream = 100,
 scanDownstream = 100
)
```

Argümanlar

Girdi argümanı	Açıklama
REcutDetails	REcutDetails.xls içinde depolanan REcutDetails
BSgenomeName	Örneğin, hg19 için BSgenome.Hsapiens.UCSC.hg19, mm10 için BSgenome.Mmusculus.UCSC.mm10, ce6 için BSgenome.Celegans.UCSC.ce6, rn5 için BSgenome.Rnorvegicus.UCSC.rn5, Zv9 için BSgenome.Drerio.UCSCdan ve dm3 için BSgenome.Dmelanogaster.UCSC.dm3
summary	Summary.xls'de depolanan özet
offTargets	offTargets.xls içinde depolanan offTarget'lar
scanUpstream	gRNA başlangıcından yukarı akış ofseti, varsayılan 100
scanDownstream	gRNA ucundan aşağı akış ofseti, varsayılan 100

3.2.3.13. WritingHits2

Açıklama

Bu fonksiyon, sıralı aramanın isabetlerini bir dosyaya yaz, searchingHits tarafından kullanılan dahili bir fonksiyondur.

Kullanım

```
writingHits2(
 gRNA,
 seqname,
 BSgenomeName,
 matches,
 strand,
 file,
 chrom_len,
 gRNA_size = 20,
 PAM = "NGG",
 PAM_pattern = "NGG",
 max_mismatch = 4,
 appendd = False,
 PAM_location = "3prime",
 PAM_size = 3,
 Allowed_mismatch_PAM = 1,
 baseEditing = False,
 targetBase = "C",
 editingWindow = [4,5,6,7,8]
)
```

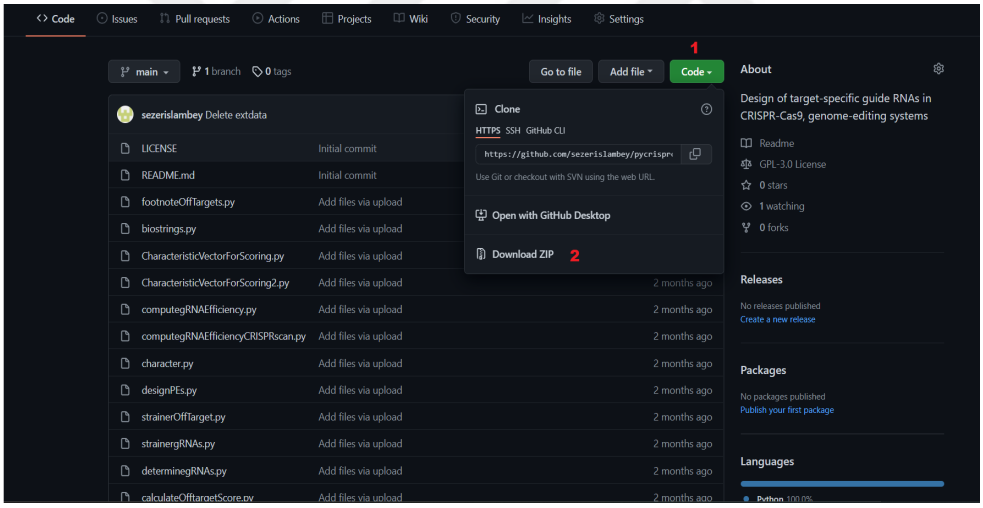
Argümanlar

Girdi argümanı	Açıklama
gRNA	PAM eklenmiş gRNA dizi nesnesi, ör. 20bp gRNA dizisine sahip ACGTACGTACGTACTGACGTCGG artı 3bp PAM dizisi CGG
seqname	karakter olarak kromozom adı, örn., Chr1
BSgenomeName	Örneğin, hg19 için BSgenome.Hsapiens.UCSC.hg19, mm10 için BSgenome.Mmusculus.UCSC.mm10, ce6 için BSgenome.Celegans.UCSC.ce6, rn5 için BSgenome.Rnorvegicus.UCSC.rn5, Zv9 için BSgenome.Drerio.UCSCdan. ve dm3 için BSgenome.Dmelanogaster.UCSC.dm3
matches	Eşleşen kromozom konumlarını depolayan nesne
strand	eşleşmenin dizisi, artı dizisi için + ve eksi dizisi için -
file	hits'in yazıldığı dosya yolu
chrom_len	eşleşen kromozomun uzunluğu
gRNA_size	gRNA boyutu, varsayılan uzunluk 20
PAM	gRNA'ya ekleme için normal ifade olarak PAM, SpCas9 için varsayılan NGG, cpf1 için TTTN olarak değiştirin.
PAM_pattern	İsabetleri filtrelemek için normal ifade olarak PAM, spCas9 için varsayılan NGG. cpf1 için ^TTTN
max_mismatch	filtrelemek için gRNA içinde izin verilen maksimum uyumsuzluk (PAM kısmı hariç),
appendd	Mevcut dosyaya eklenirse True, yeni bir dosya başlatılırsa False
PAM_location	gRNA'ya göre PAM konumu. Örneğin, spCas9 PAM 3 üssünde bulunurken cpf1 PAM 5 üssünde bulunur
PAM_size	PAM uzunluğu, varsayılan uzunluk 3
allowed_mismatch_PAM	PAM dizisiyle karşılaştırıldığında, hedeflerde izin verilen maksimum uyumsuzluk sayısı. NGG PAM için varsayılan 1'dir
baseEditing	Temel düzenleme için gRNA'ların tasarlanıp tasarlanmayacağını belirtin. Varsayılan olarak False. True ise, baseEditing = True, targetBase ve editingWidow uygun şekilde ayarlanır.
targetBase	Yalnızca baseEditing, True olarak ayarlandığında geçerlidir. Temel düzenleme sistemleri için hedef tabanı belirtmek için kullanılır, CBE sisteminde C'yi T'ye dönüştürmek için varsayılan olarak C'dir. ABE sistemini kullanmayı düşünüyorsanız bunu A olarak değiştirin.
editingWindow	Yalnızca baseEditing, TRUE olarak ayarlandığında geçerlidir. Yalnızca hedef dışı arama için dikkate alınacak etkili düzenleme penceresini belirtmek için kullanılır, orijinal CBE sistemi için varsayılan olarak 4 ila 8'dir. Kullandığınız sistemin farklı bir düzenleme penceresi varsa veya hedef tabanlı hedef dışı hedefleri daha büyük bir düzenleme penceresine dahil etmek istiyorsanız, buna göre değiştirin.

4. BULGULAR VE TARTIŞMA

Yapılan çalışma sonucunda oluşturulan pycrisprcas9 isimli Python kütüphanesi GitHUb'ta *https://github.com/sezerislambey/pycrisprcas9* adresinde dağıtılmaktadır. Yazılımı indirmek için Şekil 4.1'deki adımlar izlenmektedir.

- İndirilen zip dosyasından dosyalar çıkartılır.
- Python'un yüklü olduğu klasör içinde kütüphaneler klasörürün içine kopyalanır.
- Kopyalama işlemi tamamlandıktan sonra pycrisprcas9 kütüphanesi Python yorumlayıcısında veya Python geliştirme ortamlarında import edilerek kullanılabilir.



Şekil 4.1. pycrisprcas9 kütüphanesinin indirilmesi

4.1. pycrisprcas9 Kullanımı

İlk olarak pycrisprcas9 kütüphanesini ve çalışma esnasında kullanılacak veri dosyalarını çalışma ortamına yükleyebilmek için bazı kütüphaneleri çalışma

ortamına import etmek gerekir. Bu çalışma sonucunda oluşturulan pycrisprcas9 kütüphanesinin kullanıma yönelik bazı örnekler aşağıdaki şekildedir:

```
# pycrisprcas9 ve diğer kütüphanelerin import edilmesi
from pycrisprcas9 import *
import pandas as pd
import os
```

Uygulama 1: Restriksiyon enzimi kesim bölgeleri ile eşleştirilmiş gRNA'ları bulma

Uygun meseafe ve oryantasyondaki eşleştirilmiş gRNA'lar daha fazla özgüllük sağlar ve gRNA'lar restriksiyon enzimi kesim bölgeleriyle örtüşmesi, kesim izlemesini kolaylaştırır. Ana fonksiyon olan ***Non-TargetAnalysis*** fonksiyonunu findPairedgRNAOnly = True ve findgRNAsWithREcutOnly = True ile çağırmak, eşleştirilmiş konfigürasyonda olan ve çiftlerden en az biri bir restriksiyon enzimi kesme bölgesiyle örtüşen gRNA'ların aranması, puanlanması ve açıklanmasnı sağlar. Bir çift olarak kabul edilmek için, ileri gRNA ile karşılık gelen ters gRNA arasındaki boşluğun (min_gap, max_gap) kapsayıcı olması ve ters gRNA'nın ileri gRNA'dan önce oturması gerekir. Varsayılan min_gap = 0, max_gap = 20'dir. chromToSearch'ün hız amacıyla burada ["chrX"] olarak ayarlandı. Varsıyılan ["all"]'dır. Bir gRNA'nın restriksiyon enzimi kesim bölgesi ile örtüşme olarak kabul edilebilmesi için, enzim kesim modelinin, (overlap_gRNA_positions) belirtilen gRNA konumlarından biriyle örtüşmesi gerekir. Varsayılan overlap_gRNA_positions = [17, 18]. Hedef dışı bulma için izin verilen maksimum uyumsuzluğun varsayılan olarak max_mismatch = 4'tür. Daha yüksek bir maksimum uyumsuzluğa ayarlamanın aramayı önemli ölçüde yavaşlatacaktır. Org.Hs.eg, insan için org.Hs.eg.db paketindeki gen sembolü eşlemesinin entrezID'sidir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = True,
 findPairedgRNAOnly = True,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 chromToSearch = ["chrX"],
 min_gap = 0,
 max_gap = 20,
 txdb = "TxDb.Hsapiens.UCSC.hg19.knownGene",
 orgAnn = "org.Hs.eg",
 max_mismatch = 0,
 overlap_gRNA_positions = [17, 18],
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosya içeriği

```
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosya içeriği

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    48649563        48649586        Hsap_GATA1_ex2_gR7r     101.71982932957793
        -       48649568        48649570
chrX    48649586        48649609        Hsap_GATA1_ex2_gR40f    261.88175562825404
        +       48649602        48649604
chrX    48649585        48649608        Hsap_GATA1_ex2_gR39f    69.62952343741324
        +       48649601        48649603
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2              46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

```
Name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

inExon  inIntron        entrez_id       gene            score   n.mismatch
True                    2623            GATA1           100.0           0
True                    2623            GATA1           100.0           0
True                    2623            GATA1           100.0           0

mismatch.distance2PAM   alignment                       isCanonicalPAM
                        ....................            1
                        ....................            1
                        ....................            1

forViewInUCSC                   strand  chrom   chromStart      chromEnd
chrX:48649586-48649608          +       chrX    48649586        48649608
chrX:48649587-48649609          +       chrX    48649587        48649609
chrX:48649564-48649586          -       chrX    48649564        48649586

extendedSequence                        gRNAefficacy
TCTGGTGTCCTCCACACCAGAATCAGGGGT          0.0696295234374132
CTGGTGTCCTCCACACCAGAATCAGGGGTT          0.2618817556282540
GACACCAGAGCAGGATCCACAAACTGGGGG          0.1017198293295779

flankSequence
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

pairedgRNAx.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName                 gap
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f            0
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f            1
```

RecutDetails.xls rapor dosyası içeriği

ReversegRNAPlusPAM	ReversegRNAName	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	

ForwardgRNAPlusPAM	ForwardgRNAName	gap
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1

ForwardREcutgRNAName	ForwardREname	ForwardREpattern
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f	HinfI	GANTC
Hsap_GATA1_ex2_gR40f	HinfI	GANTC
Hsap_GATA1_ex2_gR40f	TfiI	GAWTC
Hsap_GATA1_ex2_gR40f	TfiI	GAWTC

ForwardREcutStart	ForwardREcutEnd	ReverseREcutgRNAName
13	23	Hsap_GATA1_ex2_gR7r
13	23	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
12	22	Hsap_GATA1_ex2_gR7r
12	22	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r

ReverseREname	ReverseREpattern	ReverseRECutStart	ReverseRECutEnd
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23

Summary.xls rapor dosyası içeriği

Names	forViewInUCSC	extendedSequence
Hsap_GATA1_ex2_gR7r	chrX:48649564-48649586	GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR39f	chrX:48649586-48649608	TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f	chrX:48649587-48649609	CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy	gRNAsPlusPAM	top5OfftargetTotalScore
0.10171982932957793	CCAGAGCAGGATCCACAAACNGG	NA
0.06962952343741324	GTGTCCTCCACACCAGAATCNGG	NA
0.26188175562825405	TGTCCTCCACACCAGAATCANGG	NA

top10OfftargetTotalScore	top1Hit(onTarget)MMdistance2PAM
NA	NMM
NA	NMM
NA	NMM

topOfftarget1MMdistance2PAM	topOfftarget2MMdistance2PAM
topOfftarget3MMdistance2PAM	topOfftarget4MMdistance2PAM
topOfftarget5MMdistance2PAM	topOfftarget6MMdistance2PAM
topOfftarget7MMdistance2PAM	topOfftarget8MMdistance2PAM
topOfftarget9MMdistance2PAM	topOfftarget10MMdistance2PAM

```
PairedgRNAName                                  REname
Hsap_GATA1_ex2_gR39f Hsap_GATA1_ex2_gR40f       BslI PflMI
Hsap_GATA1_ex2_gR7r                             BslI HinfI TfiI
Hsap_GATA1_ex2_gR7r                             BslI HinfI TfiI

uniqREin200     uniqREin100
```

Uygulama 2: Restriksiyon enzim kesim bölgesi olan/olmayan eşleştirilmiş gRNA'ların bulunması

Eşleştirilmiş bütün gRNA'ları bulmak için ***Non-TargetAnalysis*** fonksiyonunu findPairedgRNAOnly = True ve findgRNAsWithREcutOnly = False ile çağırmak, herhangi bir restriksiyon enzimi kesme bölgesiyle örtüşme gerektirmeden eşleştirilmiş konfigürasyondaki gRNA'ların aranması, puanlanması ve açıklanması ile sonuçlanır. Kullanıcıların daha sonra gözden geçirmeleri için gRNA'lara restriksiyon enzimi kesim bölgeleri eklenir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = False,
 findPairedgRNAOnly = True,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 chromToSearch = ["chrX"],
 txdb = "TxDb.Hsapiens.UCSC.hg19.knownGene",
 orgAnn = "org.Hs.eg",
 max_mismatch = 1,
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2              46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

```
Name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

inExon  inIntron        entrez_id       Gene            score   n.mismatch
True                    2623            GATA1           100.0   0
True                    2623            GATA1           100.0   0
True                    2623            GATA1           100.0   0

mismatch.distance2PAM   alignment               isCanonicalPAM
                        ....................    1
                        ....................    1
                        ....................    1

forViewInUCSC           strand  chrom   chromStart      chromEnd
chrX:48649586-48649608  +       chrX    48649586        48649608
chrX:48649587-48649609  +       chrX    48649587        48649609
chrX:48649564-48649586  -       chrX    48649564        48649586

extendedSequence                        gRNAefficacy
TCTGGTGTCCTCCACACCAGAATCAGGGGT          0.06962952343741324
CTGGTGTCCTCCACACCAGAATCAGGGGTT          0.26188175562825405
GACACCAGAGCAGGATCCACAAACTGGGGG          0.10171982932957793

flankSequence
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
```

```
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

pairedgRNAs.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName                 gap
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f           0
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f           1
```

RecutDetails.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName         gap
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
GTGTCCTCCACACCAGAATCAGG          Hsap_GATA1_ex2_gR39f   0
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f   1

ForwardREcutgRNAName    ForwardREname           ForwardREpattern
Hsap_GATA1_ex2_gR39f    BslI                    CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f    BslI                    CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f    HinfI                   GANTC
```

```
Hsap_GATA1_ex2_gR39f    HinfI                   GANTC
Hsap_GATA1_ex2_gR39f    TfiI                    GAWTC
Hsap_GATA1_ex2_gR39f    TfiI                    GAWTC
Hsap_GATA1_ex2_gR40f    BslI                    CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f    BslI                    CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f    HinfI                   GANTC
Hsap_GATA1_ex2_gR40f    HinfI                   GANTC
Hsap_GATA1_ex2_gR40f    TfiI                    GAWTC
Hsap_GATA1_ex2_gR40f    TfiI                    GAWTC
ForwardREcutStart       ForwardREcutEnd         ReverseREcutgRNAName
13                      23                      Hsap_GATA1_ex2_gR7r
13                      23                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
12                      22                      Hsap_GATA1_ex2_gR7r
12                      22                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r

ReverseREname   ReverseREpattern         ReverseRECutStart     ReverseRECutEnd
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
```

Summary.xls rapor dosyası içeriği

```
Names                   forViewInUCSC           extendedSequence
Hsap_GATA1_ex2_gR7r     chrX:48649564-48649586  GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR39f    chrX:48649586-48649608  TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f    chrX:48649587-48649609  CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy            gRNAsPlusPAM            top5OfftargetTotalScore
0.10171982932957793     CCAGAGCAGGATCCACAAACNGG NA
0.06962952343741324     GTGTCCTCCACACCAGAATCNGG NA
0.26188175562825405     TGTCCTCCACACCAGAATCANGG NA


top10OfftargetTotalScore        top1Hit(onTarget)MMdistance2PAM
NA                              NMM
NA                              NMM
NA                              NMM
```

```
topOfftarget1MMdistance2PAM     topOfftarget2MMdistance2PAM


topOfftarget3MMdistance2PAM     topOfftarget4MMdistance2PAM


topOfftarget5MMdistance2PAM     topOfftarget6MMdistance2PAM


topOfftarget7MMdistance2PAM     topOfftarget8MMdistance2PAM


topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM


PairedgRNAName                                  REname
Hsap_GATA1_ex2_gR39f Hsap_GATA1_ex2_gR40f       BslI PflMI
Hsap_GATA1_ex2_gR7r                             BslI HinfI TfiI
Hsap_GATA1_ex2_gR7r                             BslI HinfI TfiI

uniqREin200             uniqREin100
```

Uygulama 3: Restriksiyon enzimi kesim bölgelerine sahip tüm gRNA'ları bulma

Restriksiyon enzimi kesim bölgelerine sahip bütün gRNA'ları bulmak için ***Non-TargetAnalysis*** fonksiyonunu findPairedgRNAOnly = False ve findgRNAsWithREcutOnly = True ile çağırmak, tüm gRNA'ların (eşleştirilmiş ve eşlenmemiş) örtüşmesi restriksiyon enzimi kesim bölgeleri ve hedef dışı arama, puanlama ve açıklama eklenmesiyle sonuçlanır. Kullanıcıların daha sonra gözden geçirmeleri için gRNA'lara eşleştirilmiş bilgiler eklenir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = True,
 findPairedgRNAOnly = False,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 chromToSearch = ["chrX"],
 txdb = "TxDb.Hsapiens.UCSC.hg19.knownGene",
 orgAnn = "org.Hs.eg",
 max_mismatch = 1,
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR17f
CCAGTTTGTGGATCCTGCTCTGG
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR20r
GGTGTGGAGGACACCAGAGCAGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeirği

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    48649563        48649586        Hsap_GATA1_ex2_gR7r     101.71982932957793
        -       48649568        48649570
chrX    48649586        48649609        Hsap_GATA1_ex2_gR40f    261.88175562825404
        +       48649602        48649604
chrX    48649585        48649608        Hsap_GATA1_ex2_gR39f    69.62952343741324
        +       48649601        48649603
chrX    48649576        48649599        Hsap_GATA1_ex2_gR20r    142.091080678942
        -       48649581        48649583
chrX    48649563        48649586        Hsap_GATA1_ex2_gR17f    106.54794221610265
        +       48649579        48649581
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2              46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       17..18
                     /note="gRNAf"
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(19..20)
                     /note="gRNAr"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

```
name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR17f    CCAGTTTGTGGATCCTGCTCNGG         CCAGTTTGTGGATCCTGCTCTGG
Hsap_GATA1_ex2_gR20r    GGTGTGGAGGACACCAGAGCNGG         GGTGTGGAGGACACCAGAGCAGG
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

inExon  inIntron        entrez_id       gene    score   n.mismatch
True                    2623            GATA1   100.0   0
True                    2623            GATA1   100.0   0
True                    2623            GATA1   100.0   0
True                    2623            GATA1   100.0   0
True                    2623            GATA1   100.0   0

mismatch.distance2PAM   alignment               isCanonicalPAM
                        ....................    1
                        ....................    1
                        ....................    1
                        ....................    1
                        ....................    1

forViewInUCSC           strand  chrom   chromStart      chromEnd
chrX:48649564-48649586  +       chrX    48649564        48649586
chrX:48649577-48649599  -       chrX    48649577        48649599
chrX:48649586-48649608  +       chrX    48649586        48649608
chrX:48649587-48649609  +       chrX    48649587        48649609
chrX:48649564-48649586  -       chrX    48649564        48649586

extendedSequence                        gRNAefficacy
TCCCCCAGTTTGTGGATCCTGCTCTGGTGT          0.10654794221610266
TTCTGGTGTGGAGGACACCAGAGCAGGATC          0.142091080678942
TCTGGTGTCCTCCACACCAGAATCAGGGGT          0.06962952343741324
CTGGTGTCCTCCACACCAGAATCAGGGGTT          0.26188175562825405
```

```
GACACCAGAGCAGGATCCACAAACTGGGGG          0.10171982932957793

flankSequence
ATGGGGAGGTGGGAAGGAGAAATATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAG
GATGTGGGGTGGAAGGATTTCTGTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTG
GCCTGGGGTCCCTGGGGACCTCAGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTT
TTCTTCCCCTCTGGGCCTGAGGGCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACT
GGCCTACTACAGGGACGCTGAGGCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTG
TTGGGGTT
GGATCTCCATGGCAACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCC
TCAGCGTCCCTGTAGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCC
CTCAGGCCCAGAGGGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTG
AGGTCCCCAGGGACCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACA
GAAATCCTTCCACCCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATAT
TTCTCCTT
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

pairedgRNAs.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName                 gap
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f            0
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f            1
```

RecutDetails.xls rapor dosyası

gRNAPlusPAM	REcutgRNAName	REname
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	LpnPI
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	MspJI
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	MspJI
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	Aco12261II
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	LpnPI
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	MspJI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	BslI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	HinfI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	TfiI
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	BslI
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	HinfI
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	TfiI
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	BslI
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	PflMI

REpattern	RecutStart	REcutEnd
CCDG	14	17
YNNG	16	13
CNNR	14	17
CCRGAG	14	19
CCDG	14	17
YNNG	16	13
CCNNNNNNNGG	13	23
GANTC	16	20
GAWTC	16	20
CCNNNNNNNGG	12	22
GANTC	15	19
GAWTC	15	19
CCNNNNNNNGG	13	23
CCANNNNNTGG	13	23

Summary.xls rapor dosyası içeriği

names	forViewInUCSC	extendedSequence
Hsap_GATA1_ex2_gR17f	chrX:48649564-48649586	TCCCCCAGTTTGTGGATCCTGCTCTGGTGT
Hsap_GATA1_ex2_gR7r	chrX:48649564-48649586	GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR20r	chrX:48649577-48649599	TTCTGGTGTGGAGGACACCAGAGCAGGATC
Hsap_GATA1_ex2_gR39f	chrX:48649586-48649608	TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f	chrX:48649587-48649609	CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy	gRNAsPlusPAM	top5OfftargetTotalScore
0.10654794221610266	CCAGTTTGTGGATCCTGCTCNGG	NA
0.10171982932957793	CCAGAGCAGGATCCACAAACNGG	NA
0.142091080678942	GGTGTGGAGGACACCAGAGCNGG	NA
0.06962952343741324	GTGTCCTCCACACCAGAATCNGG	NA
0.26188175562825405	TGTCCTCCACACCAGAATCANGG	NA

top10OfftargetTotalScore	top1Hit(onTarget)MMdistance2PAM
NA	NMM
NA	NMM
NA	NMM
NA	NMM

```
NA                                  NMM

topOfftarget1MMdistance2PAM     topOfftarget2MMdistance2PAM




topOfftarget3MMdistance2PAM     topOfftarget4MMdistance2PAM




topOfftarget5MMdistance2PAM     topOfftarget6MMdistance2PAM




topOfftarget7MMdistance2PAM     topOfftarget8MMdistance2PAM




topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM




REname                  uniqREin200             uniqREin100
LpnPI MspJI MspJI
BslI PflMI              PflMI                   PflMI
Aco12261II LpnPI        MspJI
BslI HinfI TfiI         HinfI TfiI              HinfI TfiI
BslI HinfI TfiI         HinfI TfiI              HinfI TfiI
```

Uygulama 4: Tüm gRNA'ları bulma

Bütün gRNA'ların bulunması için ***Non-TargetAnalysis*** fonksiyonunu findPairedgRNAOnly = False ve findgRNAsWithREcutOnly = False ile çağırmak, tüm gRNA'ları ve hedef dışı öğeleri arama, puanlama ve açıklama ekleme ile sonuçlanır. gRNA'lara, kullanıcıların daha sonra gözden geçirmeleri için

eşleştirilmiş bilgiler ve restriksiyon enzimi kesme bölgeleri eklenir. Bu arama yukarıda belirtilen tüm arama türleri arasında en yavaş olanıdır.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = False,
 findPairedgRNAOnly = False,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 chromToSearch = ["chrX"],
 txdb = "TxDb.Hsapiens.UCSC.hg19.knownGene",
 orgAnn = "org.Hs.eg",
 max_mismatch = 1,
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR17f
CCAGTTTGTGGATCCTGCTCTGG
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR20r
GGTGTGGAGGACACCAGAGCAGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeriği

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    48649563        48649586        Hsap_GATA1_ex2_gR7r     101.71982932957793
        -       48649568        48649570
chrX    48649586        48649609        Hsap_GATA1_ex2_gR40f    261.88175562825404
        +       48649602        48649604
chrX    48649585        48649608        Hsap_GATA1_ex2_gR39f    69.62952343741324
        +       48649601        48649603
chrX    48649576        48649599        Hsap_GATA1_ex2_gR20r    142.091080678942
        -       48649581        48649583
chrX    48649563        48649586        Hsap_GATA1_ex2_gR17f    106.54794221610265
        +       48649579        48649581
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2           46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       17..18
                     /note="gRNAf"
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(19..20)
                     /note="gRNAr"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

name	gRNAPlusPAM			OffTargetSequence
Hsap_GATA1_ex2_gR17f	CCAGTTTGTGGATCCTGCTCNGG			CCAGTTTGTGGATCCTGCTCTGG
Hsap_GATA1_ex2_gR20r	GGTGTGGAGGACACCAGAGCNGG			GGTGTGGAGGACACCAGAGCAGG
Hsap_GATA1_ex2_gR39f	GTGTCCTCCACACCAGAATCNGG			GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f	TGTCCTCCACACCAGAATCANGG			TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r	CCAGAGCAGGATCCACAAACNGG			CCAGAGCAGGATCCACAAACTGG

inExon	inIntron	entrez_id	gene	score	n.mismatch
True		2623	GATA1	100.0	0
True		2623	GATA1	100.0	0
True		2623	GATA1	100.0	0
True		2623	GATA1	100.0	0
True		2623	GATA1	100.0	0

mismatch.distance2PAM	alignment	isCanonicalPAM
	.....	1
	.....	1
	.....	1
	.....	1
	.....	1

forViewInUCSC	strand	chrom	chromStart	chromEnd
chrX:48649564-48649586	+	chrX	48649564	48649586
chrX:48649577-48649599	-	chrX	48649577	48649599
chrX:48649586-48649608	+	chrX	48649586	48649608
chrX:48649587-48649609	+	chrX	48649587	48649609
chrX:48649564-48649586	-	chrX	48649564	48649586

extendedSequence	gRNAefficacy
TCCCCCAGTTTGTGGATCCTGCTCTGGTGT	0.10654794221610266
TTCTGGTGTGGAGGACACCAGAGCAGGATC	0.142091080678942
TCTGGTGTCCTCCACACCAGAATCAGGGGT	0.06962952343741324
CTGGTGTCCTCCACACCAGAATCAGGGGTT	0.26188175562825405
GACACCAGAGCAGGATCCACAAACTGGGGG	0.10171982932957793

```
flankSequence
ATGGGGAGGTGGGAAGGAGAAATATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAG
GATGTGGGGTGGAAGGATTTCTGTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTG
GCCTGGGGTCCCTGGGGACCTCAGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTT
TTCTTCCCCTCTGGGCCTGAGGGCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACT
GGCCTACTACAGGGACGCTGAGGCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTG
TTGGGGTT
GGATCTCCATGGCAACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCC
TCAGCGTCCCTGTAGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCC
CTCAGGCCCAGAGGGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTG
AGGTCCCCAGGGACCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACA
GAAATCCTTCCACCCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATAT
TTCTCCTT
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

pairedgRNAs.xls rapor dosyası

ReversegRNAPlusPAM	ReversegRNAName	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
ForwardgRNAPlusPAM	ForwardgRNAName	gap
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1

RecutDetails.xls rapor dosyası içeriği

gRNAPlusPAM	REcutgRNAName	REname
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	LpnPI
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	MspJI
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f	MspJI
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	Aco12261II
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	LpnPI
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r	MspJI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	BslI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	HinfI
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	TfiI
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	BslI
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	HinfI

```
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f            TfiI
CCAGAGCAGGATCCACAAACTGG          Hsap_GATA1_ex2_gR7r             BslI
CCAGAGCAGGATCCACAAACTGG          Hsap_GATA1_ex2_gR7r             PflMI

REpattern                        REcutStart                      RecutEnd
CCDG                             14                              17
YNNG                             16                              13
CNNR                             14                              17
CCRGAG                           14                              19
CCDG                             14                              17
YNNG                             16                              13
CCNNNNNNNGG                      13                              23
GANTC                            16                              20
GAWTC                            16                              20
CCNNNNNNNGG                      12                              22
GANTC                            15                              19
GAWTC                            15                              19
CCNNNNNNNGG                      13                              23
CCANNNNNTGG                      13                              23
```

Summary.xls rapor dosyası içeriği

```
names                    forViewInUCSC            extendedSequence
Hsap_GATA1_ex2_gR17f     chrX:48649564-48649586   TCCCCCAGTTTGTGGATCCTGCTCTGGTGT
Hsap_GATA1_ex2_gR7r      chrX:48649564-48649586   GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR20r     chrX:48649577-48649599   TTCTGGTGTGGAGGACACCAGAGCAGGATC
Hsap_GATA1_ex2_gR39f     chrX:48649586-48649608   TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f     chrX:48649587-48649609   CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy             gRNAsPlusPAM             top5OfftargetTotalScore
0.10654794221610266      CCAGTTTGTGGATCCTGCTCNGG  NA
0.10171982932957793      CCAGAGCAGGATCCACAAACNGG  NA
0.142091080678942        GGTGTGGAGGACACCAGAGCNGG  NA
0.06962952343741324      GTGTCCTCCACACCAGAATCNGG  NA
0.26188175562825405      TGTCCTCCACACCAGAATCANGG  NA

top10OfftargetTotalScore         top1Hit(onTarget)MMdistance2PAM
NA                               NMM
NA                               NMM
NA                               NMM
NA                               NMM
NA                               NMM

topOfftarget1MMdistance2PAM      topOfftarget2MMdistance2PAM




topOfftarget3MMdistance2PAM      topOfftarget4MMdistance2PAM




topOfftarget5MMdistance2PAM  topOfftarget6MMdistance2PAM
```

```
topOfftarget7MMdistance2PAM  topOfftarget8MMdistance2PAM

topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM

RErname                  uniqREin200             uniqREin100
LpnPI MspJI MspJI
BslI PflMI               PflMI                   PflMI
Aco12261II               LpnPI MspJI
BslI HinfI TfiI          HinfI TfiI              HinfI TfiI
BslI HinfI TfiI          HinfI TfiI              HinfI TfiI
```

Uygulama 5: Hedef veya hedef dışı analiz olmadan hızlı gRNA bulma

Hedef veya hedef dışı analiz yapmadan hızlı bir şekilde gRNA'ları bulmak için chromToSearch = [""] ile ***Non-TargetAnalysis*** fonksiyonunun çağrılması, hedef üzerinde ve hedef dışı analiz gerçekleştirmeden hızlı gRNA aramasıyla sonuçlanır. Fonksiyonun findgRNAsWithREcutOnly ve findPairedgRNAOnly parametreleri, gRNA'ların aranmasının restriksiyon enzimi kesim bölgeleriyle örtüşüp örtüşmediğini ve eşleştirilmiş konfigürasyonda gRNA'ların aranıp aranmadığını gösterecek şekilde ayarlanabilir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = True,
 findPairedgRNAOnly = True,
 chromToSearch = [""],
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2              46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

pairedgRNAs.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName                 gap
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f            0
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f            1
```

RecutDetails.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
```

ForwardgRNAPlusPAM	ForwardgRNAName	gap
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1

ForwardREcutgRNAName	ForwardREname	ForwardREpattern
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f	HinfI	GANTC
Hsap_GATA1_ex2_gR40f	HinfI	GANTC
Hsap_GATA1_ex2_gR40f	TfiI	GAWTC
Hsap_GATA1_ex2_gR40f	TfiI	GAWTC

ForwardREcutStart	ForwardREcutEnd	ReverseREcutgRNAName
13	23	Hsap_GATA1_ex2_gR7r
13	23	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
16	20	Hsap_GATA1_ex2_gR7r
12	22	Hsap_GATA1_ex2_gR7r
12	22	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r
15	19	Hsap_GATA1_ex2_gR7r

ReverseREname	ReverseREpattern	ReverseRECutStart	ReverseRECutEnd
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23
BslI	CCNNNNNNNGG	13	23
PflMI	CCANNNNNTGG	13	23

Uygulama 6: gRNA etkinlik tahmini ile hızlı gRNA bulma

Hızlı gRNA araması yapabilmek için ***Non-TargetAnalysis*** fonksiyonu max_mismatch = 0 atamasıyla çağrılır. Bu işlem hedef dışı analiz olmaksızın gRNA etkinlik tahmini ile sonuçlanır. Fonksiyonun bir parametresi olan rule_set, 2015'te yayınlanan CRISPRscan (Moreno-Mateos ve ark., 2015) veya 2014'te yayınlanan kural seti 1 kullanılarak etkinliği hesaplamak için Root_RuleSet1_2014 olarak ayarlanabilir. Varsayılan olarak, gRNA etkinliği 2014'te yayınlanan kural seti 1 kullanılarak tahmin edilir. Ek olarak, findgRNAsWithREcutOnly ve findPairedgRNAOnly parametreleri, gRNA'ların aranmasının restriksiyon enzimi kesim bölgeleriyle örtüşüp örtüşmediğini ve eşleştirilmiş konfigürasyonda gRNA'ların aranıp aranmadığını gösterecek şekilde ayarlanabilir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = True,
 annotateExon = False,
 findPairedgRNAOnly = True,
 chromToSearch = ["chrX"],
 max_mismatch = 0,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeriği

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    48649563        48649586        Hsap_GATA1_ex2_gR7r     101.71982932957793
        -       48649568        48649570
chrX    48649586        48649609        Hsap_GATA1_ex2_gR40f    261.88175562825404
        +       48649602        48649604
chrX    48649585        48649608        Hsap_GATA1_ex2_gR39f    69.62952343741324
        +       48649601        48649603
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2              46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

```
name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

score   n.mismatch      mismatch.distance2PAM           alignment
100.0   0                                               ..................
100.0   0                                               ..................
100.0   0                                               ..................

isCanonicalPAM          forViewInUCSC                   strand          chrom
1                       chrX:48649586-48649608          +               chrX
1                       chrX:48649587-48649609          +               chrX
1                       chrX:48649564-48649586          -               chrX

chromStart      chromEnd        extendedSequence
48649586        48649608        TCTGGTGTCCTCCACACCAGAATCAGGGGT
48649587        48649609        CTGGTGTCCTCCACACCAGAATCAGGGGTT
48649564        48649586        GACACCAGAGCAGGATCCACAAACTGGGGG

gRNAefficacy
0.06962952343741324
0.26188175562825405
0.10171982932957793
```

```
flankSequence
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

pairedgRNAs.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName               gap
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f          0
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f          1
```

RecutDetails.xls rapor dosyası içeriği

```
ReversegRNAPlusPAM              ReversegRNAName
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG         Hsap_GATA1_ex2_gR7r

ForwardgRNAPlusPAM              ForwardgRNAName
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG         Hsap_GATA1_ex2_gR39f
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG         Hsap_GATA1_ex2_gR40f
```

```
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG          Hsap_GATA1_ex2_gR40f

gap     ForwardREcutgRNAName            ForwardREname   ForwardREpattern
0       Hsap_GATA1_ex2_gR39f            BslI            CCNNNNNNNGG
0       Hsap_GATA1_ex2_gR39f            BslI            CCNNNNNNNGG
0       Hsap_GATA1_ex2_gR39f            HinfI           GANTC
0       Hsap_GATA1_ex2_gR39f            HinfI           GANTC
0       Hsap_GATA1_ex2_gR39f            TfiI            GAWTC
0       Hsap_GATA1_ex2_gR39f            TfiI            GAWTC
1       Hsap_GATA1_ex2_gR40f            BslI            CCNNNNNNNGG
1       Hsap_GATA1_ex2_gR40f            BslI            CCNNNNNNNGG
1       Hsap_GATA1_ex2_gR40f            HinfI           GANTC
1       Hsap_GATA1_ex2_gR40f            HinfI           GANTC
1       Hsap_GATA1_ex2_gR40f            TfiI            GAWTC
1       Hsap_GATA1_ex2_gR40f            TfiI            GAWTC

ReverseREname   ReverseREpattern        ReverseRECutStart     ReverseRECutEnd
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
```

Summary.xls rapor dosyası içeriği

```
names                   forViewInUCSC           extendedSequence
Hsap_GATA1_ex2_gR7r     chrX:48649564-48649586  GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR39f    chrX:48649586-48649608  TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f    chrX:48649587-48649609  CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy            gRNAsPlusPAM            top5OfftargetTotalScore
0.10171982932957793     CCAGAGCAGGATCCACAAACNGG NA
0.06962952343741324     GTGTCCTCCACACCAGAATCNGG NA
0.26188175562825405     TGTCCTCCACACCAGAATCANGG NA


top10OfftargetTotalScore        top1Hit(onTarget)MMdistance2PAM
NA                              NMM
NA                              NMM
NA                              NMM

topOfftarget1MMdistance2PAM     topOfftarget2MMdistance2PAM
```

```
topOfftarget3MMdistance2PAM     topOfftarget4MMdistance2PAM



topOfftarget5MMdistance2PAM     topOfftarget6MMdistance2PAM



topOfftarget7MMdistance2PAM     topOfftarget8MMdistance2PAM



topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM



PairedgRNAName                                      REname
Hsap_GATA1_ex2_gR39f Hsap_GATA1_ex2_gR40f       BslI PflMI
Hsap_GATA1_ex2_gR7r                                 BslI HinfI TfiI
Hsap_GATA1_ex2_gR7r                                 BslI HinfI TfiI

uniqREin200             uniqREin100
```

"CRISPRscan" kural setini kullanmak için rule_set = "CRISPRscan" olarak seçilir.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = True,
 annotateExon = False,
 findPairedgRNAOnly = True,
 chromToSearch = ["chrX"],
 max_mismatch = 0,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 rule_set = "CRISPRscan",
 baseBeforegRNA = 6,
 baseAfterPAM = 6,
 featureWeightMatrixFile = pd.read_csv(os.getcwd() +
 os.sep + os.sep.join(["extdata", "Morenos-Mateo.csv"])),
 outputDir = os.getcwd(),
 overwrite = True)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği (rule_set = "CRISPRscan")

```
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeriği (rule_set = "CRISPRscan")

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    4864956348649586Hsap_GATA1_ex2_gR7r             -       4864956848649570
chrX    4864958648649609Hsap_GATA1_ex2_gR40f            +       4864960248649604
chrX    4864958548649608Hsap_GATA1_ex2_gR39f            +       4864960148649603
```

gbk uzantılı Genbank veri dosyası içeriği (rule_set = "CRISPRscan")

```
LOCUS       Hsap_GATA1_ex2               46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği (rule_set = "CRISPRscan")

```
name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

score   n.mismatch  mismatch.distance2PAM       alignment
100.0   0                                       ....................
100.0   0                                       ....................
100.0   0                                       ....................

isCanonicalPAM  forViewInUCSC                   Strand          chrom
1               chrX:48649586-48649608          +               chrX
1               chrX:48649587-48649609          +               chrX
1               chrX:48649564-48649586          -               chrX
```

chromStart	chromEnd	extendedSequence	gRNAefficacy
48649586	48649608	TCTGGTGTCCTCCACACCAGAATCAGGGGT	57.7301209
48649587	48649609	CTGGTGTCCTCCACACCAGAATCAGGGGTT	45.1233547
48649564	48649586	GACACCAGAGCAGGATCCACAAACTGGGGG	50.1092327

pairedgRNAs.xls rapor dosyası içeriği (rule_set = "CRISPRscan")

ReversegRNAPlusPAM	ReversegRNAName	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
ForwardgRNAPlusPAM	ForwardgRNAName	gap
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1

RecutDetails.xls rapor dosyası içeriği (rule_set = "CRISPRscan")

ReversegRNAPlusPAM	ReversegRNAName	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r	
ForwardgRNAPlusPAM	ForwardgRNAName	gap
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f	0
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f	1
ForwardREcutgRNAName	ForwardREname	ForwardREpattern
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	HinfI	GANTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR39f	TfiI	GAWTC
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG
Hsap_GATA1_ex2_gR40f	BslI	CCNNNNNNNGG

```
Hsap_GATA1_ex2_gR40f    HinfI                   GANTC
Hsap_GATA1_ex2_gR40f    HinfI                   GANTC
Hsap_GATA1_ex2_gR40f    TfiI                    GAWTC
Hsap_GATA1_ex2_gR40f    TfiI                    GAWTC
ForwardREcutStart       ForwardREcutEnd         ReverseREcutgRNAName
13                      23                      Hsap_GATA1_ex2_gR7r
13                      23                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
16                      20                      Hsap_GATA1_ex2_gR7r
12                      22                      Hsap_GATA1_ex2_gR7r
12                      22                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r
15                      19                      Hsap_GATA1_ex2_gR7r

ReverseREname   ReverseREpattern        ReverseRECutStart     ReverseRECutEnd
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
BslI            CCNNNNNNNGG             13                    23
PflMI           CCANNNNNTGG             13                    23
```

Summary.xls rapor dosyası içeriği (rule_set = "CRISPRscan")

```
names                   forViewInUCSC           extendedSequence
Hsap_GATA1_ex2_gR7r     chrX:48649564-48649586  GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR39f    chrX:48649586-48649608  TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f    chrX:48649587-48649609  CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy    gRNAsPlusPAM            top5OfftargetTotalScore
57.7301209      CCAGAGCAGGATCCACAAACNGG NA
45.1233547      GTGTCCTCCACACCAGAATCNGG NA
50.1092327      TGTCCTCCACACCAGAATCANGG NA

top10OfftargetTotalScore        top1Hit(onTarget)MMdistance2PAM
NA                              NMM
NA                              NMM
NA                              NMM

topOfftarget1MMdistance2PAM     topOfftarget2MMdistance2PAM




topOfftarget3MMdistance2PAM     topOfftarget4MMdistance2PAM
```

```
topOfftarget5MMdistance2PAM     topOfftarget6MMdistance2PAM

topOfftarget7MMdistance2PAM     topOfftarget8MMdistance2PAM

topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM

PairedgRNAName                                          REname
Hsap_GATA1_ex2_gR39f Hsap_GATA1_ex2_gR40f               BslI PflMI
Hsap_GATA1_ex2_gR7r                                     BslI HinfI TfiI
Hsap_GATA1_ex2_gR7r                                     BslI HinfI TfiI

uniqREin200                     uniqREin100
```

Uygulama 7: Hedef dışı hedefler için kesim frekansı belirleme (CFD) puanı

Hedef dışı hedefleri CFD puanına göre kesim frekanslarının belirlenmesi için, score_method = "CFDscore" ve PAM_pattern = "NNG$|NGN$" olarak ayarlıyken ***Non-TargetAnalysis*** fonksiyonunu çağırmak, Doench ve ark (2016), tarafından geliştirilen hem uyumsuzluk pozisyonunun hem de uyumsuzluk tipinin kesme frekansı üzerindeki etkilerini modelleyen algoritmayı kullanarak CFD puanı verecektir. Varsayılan olarak, score_method, yalnızca uyumsuzluk konumunun etkisini modelleyen Hsu-Zhang'a ayarlanmıştır.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 annotatePaired = False,
 scoring_method = "CFDscore",
 PAM_pattern = "NNG$|NGN$",
 chromToSearch = ["chrX"],
 annotateExon = False,
 max_mismatch = 2,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 outputDir = os.getcwd(),
 overwrite = True
)
```

Çıktı:

gRNA'ları içeren allgRNAs.fa dosyası içeriği

```
>Hsap_GATA1_ex2_gR17f
CCAGTTTGTGGATCCTGCTCTGG
>Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG
>Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG
>Hsap_GATA1_ex2_gR20r
GGTGTGGAGGACACCAGAGCAGG
>Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG
```

Genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası içeriği

```
track name="gRNA sites" description="pycrisprcas9" visibility=2 useScore=1
itemRgb="On"
chrX    48649563        48649586        Hsap_GATA1_ex2_gR7r     101.71982932957793
        -       48649568        48649570
chrX    48649586        48649609        Hsap_GATA1_ex2_gR40f    261.88175562825404
        +       48649602        48649604
chrX    48649585        48649608        Hsap_GATA1_ex2_gR39f    69.62952343741324
        +       48649601        48649603
chrX    48649576        48649599        Hsap_GATA1_ex2_gR20r    142.091080678942
        -       48649581        48649583
chrX    48649563        48649586        Hsap_GATA1_ex2_gR17f    106.54794221610265
        +       48649579        48649581
```

gbk uzantılı Genbank veri dosyası içeriği

```
LOCUS       Hsap_GATA1_ex2               46 bp    dna     linear   UNK
DEFINITION  pycrisprcas9 output for inputseq sequence
ACCESSION   unknown
FEATURES             Location/Qualifiers
     misc_bind       17..18
                     /note="gRNAf"
     misc_bind       39..40
                     /note="gRNAf"
     misc_bind       40..41
                     /note="gRNAf"
     misc_bind       complement(19..20)
                     /note="gRNAr"
     misc_bind       complement(6..7)
                     /note="gRNAr"
ORIGIN
        1 ccagtttgtg gatcctgctc tggtgtcctc cacaccagaa tcaggg
//
```

Non-TargetAnalysis.xls rapor dosyası içeriği

```
name                    gRNAPlusPAM                     OffTargetSequence
Hsap_GATA1_ex2_gR17f    CCAGTTTGTGGATCCTGCTCNGG         CCAGTTTGTGGATCCTGCTCTGG
Hsap_GATA1_ex2_gR20r    GGTGTGGAGGACACCAGAGCNGG         GGTGTGGAGGACACCAGAGCAGG
Hsap_GATA1_ex2_gR39f    GTGTCCTCCACACCAGAATCNGG         GTGTCCTCCACACCAGAATCAGG
Hsap_GATA1_ex2_gR40f    TGTCCTCCACACCAGAATCANGG         TGTCCTCCACACCAGAATCAGGG
Hsap_GATA1_ex2_gR7r     CCAGAGCAGGATCCACAAACNGG         CCAGAGCAGGATCCACAAACTGG

score   n.mismatch  mismatch.distance2PAM       alignment
1       0                                       ......................
1       0                                       ......................
1       0                                       ......................
1       0                                       ......................
1       0                                       ......................

isCanonicalPAM  forViewInUCSC                   strand          chrom
1               chrX:48649564-48649586          +               chrX
1               chrX:48649577-48649599          -               chrX
1               chrX:48649586-48649608          +               chrX
1               chrX:48649587-48649609          +               chrX
1               chrX:48649564-48649586          -               chrX

chromStart      chromEnd                extendedSequence
48649564        48649586                TCCCCCAGTTTGTGGATCCTGCTCTGGTGT
48649577        48649599                TTCTGGTGTGGAGGACACCAGAGCAGGATC
48649586        48649608                TCTGGTGTCCTCCACACCAGAATCAGGGGT
48649587        48649609                CTGGTGTCCTCCACACCAGAATCAGGGGTT
48649564        48649586                GACACCAGAGCAGGATCCACAAACTGGGGG

gRNAefficacy
0.10654794221610266
0.142091080678942
0.06962952343741324
0.26188175562825405
0.10171982932957793
```

```
flankSequence
ATGGGGAGGTGGGAAGGAGAAATATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAG
GATGTGGGGTGGAAGGATTTCTGTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTG
GCCTGGGGTCCCTGGGGACCTCAGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTT
TTCTTCCCCTCTGGGCCTGAGGGCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACT
GGCCTACTACAGGGACGCTGAGGCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTG
TTGGGGTT
GGATCTCCATGGCAACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCC
TCAGCGTCCCTGTAGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCC
CTCAGGCCCAGAGGGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTG
AGGTCCCCAGGGACCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACA
GAAATCCTTCCACCCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATAT
TTCTCCTT
TATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCT
GTGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTC
AGAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGG
GCTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAG
GCCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCT
TGGCTAGG
ATGGAGACTGAGGTGATGGAGTGGGAGGAGGGGGAAAGGAGGAAGAGGAGCAGGTGAAAGGATGTGGGGTGGAAGGATTTCTG
TGTCTGAGGACCCCTTCTGTCCTCGCAGGTTAATCCCCAGAGGCTCCATGGAGTTCCCTGGCCTGGGGTCCCTGGGGACCTCA
GAGCCCCTCCCCCAGTTTGTGGATCCTGCTCTGGTGTCCTCCACACCAGAATCAGGGGTTTTCTTCCCCTCTGGGCCTGAGGG
CTTGGATGCAGCAGCTTCCTCCACTGCCCCGAGCACAGCCACCGCTGCAGCTGCGGCACTGGCCTACTACAGGGACGCTGAGG
CCTACAGACACTCCCCAGGTAACTCCATTGAGTGGCTGTCTTGGCATTGGCTGAGTGCTGTTGGGGTTGCCATGGAGATCCTT
GGCTAGGT
AACCCCAACAGCACTCAGCCAATGCCAAGACAGCCACTCAATGGAGTTACCTGGGGAGTGTCTGTAGGCCTCAGCGTCCCTGT
AGTAGGCCAGTGCCGCAGCTGCAGCGGTGGCTGTGCTCGGGGCAGTGGAGGAAGCTGCTGCATCCAAGCCCTCAGGCCCAGAG
GGGAAGAAAACCCCTGATTCTGGTGTGGAGGACACCAGAGCAGGATCCACAAACTGGGGGAGGGGCTCTGAGGTCCCCAGGGA
CCCCAGGCCAGGGAACTCCATGGAGCCTCTGGGGATTAACCTGCGAGGACAGAAGGGGTCCTCAGACACAGAAATCCTTCCAC
CCCACATCCTTTCACCTGCTCCTCTTCCTCCTTTCCCCCTCCTCCCACTCCATCACCTCAGTCTCCATATTTCTCCTTCCCAC
CTCCCCAT
```

RecutDetails.xls rapor dosyası içeriği

gRNAPlusPAM	REcutgRNAName
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f
CCAGTTTGTGGATCCTGCTCTGG	Hsap_GATA1_ex2_gR17f
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r
GGTGTGGAGGACACCAGAGCAGG	Hsap_GATA1_ex2_gR20r
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f
GTGTCCTCCACACCAGAATCAGG	Hsap_GATA1_ex2_gR39f
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f
TGTCCTCCACACCAGAATCAGGG	Hsap_GATA1_ex2_gR40f
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r
CCAGAGCAGGATCCACAAACTGG	Hsap_GATA1_ex2_gR7r

REname	REpattern	RECutStart	RECutEnd
LpnPI	CCDG	14	17
MspJI	YNNG	16	13
MspJI	CNNR	14	17
Aco12261II	CCRGAG	14	19

```
LpnPI           CCDG                    14                      17
MspJI           YNNG                    16                      13
BslI            CCNNNNNNNGG             13                      23
HinfI           GANTC                   16                      20
TfiI            GAWTC                   16                      20
BslI            CCNNNNNNNGG             12                      22
HinfI           GANTC                   15                      19
TfiI            GAWTC                   15                      19
BslI            CCNNNNNNNGG             13                      23
PflMI           CCANNNNNTGG             13                      23
```

Summary.xls rapor dosyası içeriği

```
names                   forViewInUCSC           extendedSequence
Hsap_GATA1_ex2_gR17f    chrX:48649564-48649586  TCCCCCAGTTTGTGGATCCTGCTCTGGTGT
Hsap_GATA1_ex2_gR7r     chrX:48649564-48649586  GACACCAGAGCAGGATCCACAAACTGGGGG
Hsap_GATA1_ex2_gR20r    chrX:48649577-48649599  TTCTGGTGTGGAGGACACCAGAGCAGGATC
Hsap_GATA1_ex2_gR39f    chrX:48649586-48649608  TCTGGTGTCCTCCACACCAGAATCAGGGGT
Hsap_GATA1_ex2_gR40f    chrX:48649587-48649609  CTGGTGTCCTCCACACCAGAATCAGGGGTT

gRNAefficacy            gRNAsPlusPAM            top5OfftargetTotalScore
0.10654794221610266     CCAGTTTGTGGATCCTGCTCNGG NA
0.10171982932957793     CCAGAGCAGGATCCACAAACNGG NA
0.142091080678942       GGTGTGGAGGACACCAGAGCNGG NA
0.06962952343741324     GTGTCCTCCACACCAGAATCNGG NA
0.26188175562825405     TGTCCTCCACACCAGAATCANGG NA

top10OfftargetTotalScore        top1Hit(onTarget)MMdistance2PAM
NA                                      NMM
NA                                      NMM
NA                                      NMM
NA                                      NMM
NA                                      NMM

topOfftarget1MMdistance2PAM     topOfftarget2MMdistance2PAM




topOfftarget3MMdistance2PAM     topOfftarget4MMdistance2PAM




topOfftarget5MMdistance2PAM     topOfftarget6MMdistance2PAM
```

```
topOfftarget7MMdistance2PAM     topOfftarget8MMdistance2PAM


topOfftarget9MMdistance2PAM     topOfftarget10MMdistance2PAM


RErname                         uniqREin200                     uniqREin100
LpnPI MspJI MspJI
BslI PflMI                      PflMI                           PflMI
Aco12261II LpnPI MspJI
BslI HinfI TfiI                 HinfI TfiI                      HinfI TfiI
BslI HinfI TfiI                 HinfI TfiI                      HinfI TfiI
```

Uygulama 8: Temel düzenleme sistemleri için gRNA'lar tasarlayın

Sitozin veya adenin baz düzenleyicileri (CBE'ler veya ABE'ler), spesifik DNA C'den T'ye veya A'dan G'ye değişiklikleri sağlayabilir. Varsayılan olarak, CBE sistemi için targetBase C'ye, editWindow = [4, 8] olarak ayarlanmıştır.

Komut:

```
results = Non-TargetAnalysis(
 inputFilePath,
 findgRNAsWithREcutOnly = False,
 findPairedgRNAOnly = False,
 annotatePaired = False,
 BsgenomeName = "Bsgenome.Hsapiens.UCSC.hg19",
 chromToSearch = ["chrX"],
 txdb = "TxDb.Hsapiens.UCSC.hg19.knownGene",
 orgAnn = "org.Hs.eg",
 max_mismatch = 4,
 outputDir = os.getcwd(),
 overwrite = True,
 PAM_location = "5prime",
 PAM = "TGT",
 PAM_pattern = "^TRN",
 allowed_mismatch_PAM = 2,
 subPAM_position = [1,2],
 baseEditing = True,
 editingWindow = [10, 20],
 targetBase = "A"
)
```

Çıktı:

Herhangi bir gRNA bulunmadığında programın verdiği çıktı

```
Validating input ...

Searching for gRNAs ...

No gRNAs found in the input sequence ['Hsap_GATA1_ex2']
no gRNAs found!
None
```

Kullanım için verilen örneklerin sonuçlarına bakıldığında pycrisprcas9 kütüphanesi gRNA, gRNA verimliliği, gRNA'nın bulunduğu kromozom, başlangıç ve bitiş konumları, uyumsuzluklar ve konumları, restriksiyon enzimleri, restriksiyon enzimlerinin deseni gibi detaylı bir analiz gerçekleştirmektedir. Bunlara ek olarak genomik bölgeleri koordinatlar ve ilişkili ek açıklamaları içeren bed dosyası, gRNA'ları içeren fasta dosyası ve genbank veri dosyalarınıda oluşturmaktadır.

4.2. Tartışma

Çalışma kapsamında geliştirilen pycrisprcas9 kütüphanesinin performansı temel olarak CRISPRseek paketindeki veriler baz alınarak test edilmiş ve CRISPRseek ile aynı sonuçları ürettiği görülmüştür. Ayrıca aynı verilerle CRISPRseek yazılımı ile de test edilerek işlem zamanı açısında yapılan karşılaştırma sonuçlar Çizelge 4.1'de verilmiştir.

Çizelge 4.1. pycrisprcas9 kütüphanesi ile CRISPRseek paketinin karşılaştırılması

	Yazılım	***İşlem süresi (saniye)***
Uygulama 1	*pycrisprcas9*	*98.25*
	CRISPRseek	*8.33*
Uygulama 2	*pycrisprcas9*	*151.6*
	CRISPRseek	*9.04*
Uygulama 3	*pycrisprcas9*	*150.8*
	CRISPRseek	*9.33*
Uygulama 4	*pycrisprcas9*	*212.43*
	CRISPRseek	*10.79*
Uygulama 5	*pycrisprcas9*	*0.440*
	CRISPRseek	*0.555*
Uygulama 6	*pycrisprcas9*	*68.30*
	CRISPRseek	*4.90*
Uygulama 7	*pycrisprcas9*	*108.11*
	CRISPRseek	*4.79*
Uygulama 8	*pycrisprcas9*	*0.0107*
	CRISPRseek	*0.019*

Çizelge 4.1'de görüntülenen sonuçlara göre pycrisprcas9 kütüphanesinin R'deki CRISPRseek paketinden daha yavaş çalıştığı görülmektedir. Yapılacak güncellemeler ile multiproccessing (çoklu işleme) kütüphanelerinden yararlanılarak hız bakımından iyileştirme planlanmaktadır.

5. SONUÇLAR VE ÖNERİLER

Mevcut CRISPR-Cas9 türetilmiş sistemler için hızla değişen bilgiler ve bu türlerin yeni türlerden sürekli olarak geliştirilmesi, gRNA'ların ve hedef dışı bölgelerin tanımlanması ve değerlendirilmesinde en son bilgileri esnek bir şekilde barındırabilen bir yazılıma ihtiyaç duymaktadır. Özellikle son yıllarda popülerliği artan Python'da da bulunan bu alandaki eksikliğin giderilmesi amacıyla, bu çalışmada CRISPR-Cas9 sisteminin etkili bir şekilde uygulanması adına hedefe özgü gRNA'ların tasarımını kolaylaştırmak için açık kaynaklı bir kütüphane olan pycrisprcas9 adı verilen bir Python yazılımı geliştirilmiştir. Yazılım herhangi bir türü mevcut sekansla arama ve karakterize edilmiş herhangi bir bakteri türünden CRISPR-Cas9 sistemleri için hedef sekans tercihlerini kullanma yeteneği / esnekliği sunmaktadır.

gRNA tasarımı için geliştirilen biyoenformatik araçların kendi içinde barındırdıkları bazı artı ve eksi yönleri bulunmaktadır. Cass-OFFinder'ı komut satırında kullanmak için OpenCL yazılımı gerektirmesinden dolayı yüksek depolama alanı gerektirmektedir. Ayrıca Cas-OFFinder 0-10 arası uyumsuzluklara izin vermektedir ve NGG, NRG, NNAGAAW, NNNNGMTT PAM bölgelerini desteklemektedir (Bae ve ark., 2014). CRISPRscan NGG, TTTV ve TTTN PAM bölgelerini desteklemektedir (Moreno-Mateos ve ark., 2015). CRISPy-web ise maksimum 4 uyumsuzluğa izin vermekle birlikte yalnızca PAM bölgesi olarak NGG'yi desteklemektedir (Ronda ve ark., 2014). CRISPR Design PAM bölgesi olarak NGG ve NAG'yi desteklemektedir (Hsu ve ark., 2013). Çevrimdışı olan SSFinder'ın kullanımı çok yavaştır. Birçok araçta bulunan bu tür eksikleri gidermek adına pycrisprcas9 kullanıcı tanımlı uyumsuzluk sayısını ve PAM bölgelerini desteklemektedir. Ayrıca bütün genomda arama yaparak genomlardaki bütün hedefleri ek açıklamalar ile birlikte raporlayabilmektedir.

Geliştirilen yazılım Github'da *https://github.com/sezerislambey/pycrisprcas9* adresinde GNU General Public License 3 lisansı ile açık kaynak ve açık erişimle dağıtılmaktadır. Geliştirilen pycrisprcas9 kütüphanesindeki ***Non-TargetAnalysis*** fonksiyonu temel olarak ihtiyaç duyulan tüm CRISPR işlemlerini gerçekleştirilebilmektedir. Bundan sonraki aşamada kullanım kolaylığı sağlamak için yazılımın grafik kullanıcı arayüzü (GUI) sürümü geliştirilmesi planlanmıştır. Ayrıca ilgili iki sekanstan birini veya her iki sekansı hedefleyen gRNA'ların tanımlanmasına olanak sağlayacak yeni bir fonksiyonun yanında Kim ve ark (2018) tarafından açıklanan derin öğrenmeye dayalı gRNA aktivitesinin tahmini için gerekli algoritmaların eklenmesiyle kütüphanenin fonksiyonel açıdan daha yetenekli hale getirilmesi hedeflenmektedir.

KAYNAKLAR

Alkhnbashi, O. S., Meier, T., Mitrofanov, A., Backofen, R., ve Voß, B., 2020. CRISPR-Cas bioinformatics. Methods, 172, 3-11.

Bae, S., Park, J., ve Kim, J. S., 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics, 30(10), 1473-1475.

Baker, M., 2012. Gene-editing nucleases: precise ways to modify the genome arose from unexpected places. Nature methods, 9(1), 23-27.

Baliou, S., Adamaki, M., Kyriakopoulos, A. M., Spandidos, D. A., Panayiotidis, M., Christodoulou, I., ve Zoumpourlis, V., 2018. CRISPR therapeutic tools for complex genetic disorders and cancer. International journal of oncology, 53(2), 443-468.

Barrangou, R., van der Oost, J., 2013. CRISPR-Cas Systems: RNA-mediated Adaptive Immunity in Bacteria and Archaea. Springer Press, Heidelburg, s.1–129.

Berg, J. M., 1990. Zinc finger domains: hypotheses and current knowledge. Annual review of biophysics and biophysical chemistry, 19(1), 405-421.

Biswas, A., Staals, R. H., Morales, S. E., Fineran, P. C., ve Brown, C. M., 2016. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC genomics, 17(1), 1-14.

Blin, K., Pedersen, L. E., Weber, T., ve Lee, S. Y., 2016. CRISPy-web: an online resource to design sgRNAs for CRISPR applications. Synthetic and systems biotechnology, 1(2), 118-121.

Bolotin, A., Quinquis, B., Sorokin, A., ve Ehrlich, S. D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.

Chari, R., Mali, P., Moosburner, M., ve Church, G. M., 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nature methods, 12(9), 823-826.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., ve Kim, J. S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome research, 24(1), 132-141.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., ... ve Voytas, D. F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics, 186(2), 757-761.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... ve Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science, 339(6121), 819-823.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., ... ve Pourcel, C., 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic acids research, 46(W1), W246-W251.

Cui, Y., Xu, J., Cheng, M., Liao, X., ve Peng, S., 2018. Review of CRISPR/Cas9 sgRNA design tools. Interdisciplinary Sciences: Computational Life Sciences, 10(2), 455-465.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... ve Root, D. E., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature biotechnology, 34(2), 184-191.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., ... ve Root, D. E., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nature biotechnology, 32(12), 1262-1267.

Dong, C., Hao, G. F., Hua, H. L., Liu, S., Labena, A. A., Chai, G., ... ve Guo, F. B. 2018. Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins. Nucleic acids research, 46(D1), D393-D398.

Edgar, R. C., 2007. PILER-CR: fast and accurate identification of CRISPR repeats. BMC bioinformatics, 8(1), 1-6.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... ve Searle, S. M., 2014. Ensembl 2014. Nucleic acids research, 42(D1), D749-D755.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., ve Sander, J. D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, *31*(9), 822-826.

Graham, D. B., ve Root, D. E., 2015. Resources for the design of CRISPR gene editing experiments. Genome biology, 16(1), 1-21.

Grissa, I., Vergnaud, G., ve Pourcel, C. 2007a. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic acids research, 35(suppl_2), W52-W57.

Grissa, I., Vergnaud, G., ve Pourcel, C., 2007b. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC bioinformatics, 8(1), 1-10.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., ... ve Concordet, J. P., 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome biology, 17(1), 1-12.

Hodgkins, A., Farne, A., Perera, S., Grego, T., Parry-Smith, D. J., Skarnes, W. C., ve Iyer, V., 2015. WGE: a CRISPR database for genome engineering. Bioinformatics, 31(18), 3078-3080.

Horvath, P., ve Barrangou, R,. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, *327*(5962), 167-170.

Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., ve Thomson, J. A., 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences, 110(39), 15644-15649.

Hsu P.D., Lander E.S., Zhang F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157(6):1262–1278.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... ve Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology, 31(9), 827-832.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., ... ve Joung, J. K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology, 31(3), 227-229.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., ve Nakata, A., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, *169*(12), 5429-5433.

Jacobs, T. B., ve Martin, G. B., 2016. High-throughput CRISPR vector construction and characterization of DNA modifications by generation of tomato hairy roots. Journal of visualized experiments: JoVE, (110).

Jiang, F., ve Doudna, J. A., 2017. CRISPR–Cas9 structures and mechanisms. Annual review of biophysics, 46, 505-529.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., ve Doudna, J., 2013. RNA-programmed genome editing in human cells. elife, 2, e00471.

Kaur, K., Tandon, H., Gupta, A. K., ve Kumar, M., 2015. CrisprGE: a central hub of CRISPR/Cas-based genome editing. Database, 2015.

Kim, Y. G., Cha, J., ve Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences, 93(3), 1156-1160.

Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., ... ve Kim, H. H., 2018. Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. Nature biotechnology, 36(3), 239-241.

Koonin, E. V., ve Makarova, K. S., 2009. CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 biology reports*, *1*.

Labaj, W., Papiez, A., Polanski, A., ve Polanska, J., 2017. Comprehensive analysis of MILE gene expression data set advances discovery of leukaemia type and subtype biomarkers. Interdisciplinary Sciences: Computational Life Sciences, 9(1), 24-35.

Langmead, B., Trapnell, C., Pop, M., ve Salzberg, S. L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 10(3), 1-10.

Lee, C. M., Cradick, T. J., Fine, E. J., ve Bao, G., 2016. Nuclease target site selection for maximizing on-target activity and minimizing off-target effects in genome editing. Molecular Therapy, 24(3), 475-487.

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., ... ve Liu, M., 2013. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nature biotechnology, 31(8), 681-683.

Li, H., ve Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.

Mali, P., Esvelt, K. M., ve Church, G. M., 2013a. Cas9 as a versatile tool for engineering biology. Nature methods, 10(10), 957-963.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., ... ve Church, G. M., 2013b. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature biotechnology, 31(9), 833-838.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... ve Church, G. M., 2013c. RNA-guided human genome engineering via Cas9. Science, 339(6121), 823-826.

Mojica, F. J., Díez-Villaseñor, C., Soria, E., ve Juez, G., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, *36*(1), 244-246.

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., ve Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174-182.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., ve Valen, E., 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic acids research, 42(W1), W401-W407.

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., ve Giraldez, A. J., 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nature methods, 12(10), 982-988.

Naito, Y., Hino, K., Bono, H., ve Ui-Tei, K., 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics, 31(7), 1120-1123.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., ... ve Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell, 156(5), 935-949.

Parkhi, V., Bhattacharya, A., Choudhary, S., Pathak, R., Gawade, V., Palan, B., ... ve Char, B., 2018. Demonstration of CRISPR-cas9-mediated pds gene editing in a tomato hybrid parental line. Indian J. Genet, 78(1), 132-137.

Pourcel, C., Salvignol, G., ve Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and

provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., ve Zhang, F., 2013a. Genome engineering using the CRISPR-Cas9 system. Nature protocols, 8(11), 2281-2308.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., ... ve Zhang, F., 2013b. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 154(6), 1380-1389.

Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T., ve Kildegaard, H. F., 2014. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnology and bioengineering, 111(8), 1604-1616.

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., ve Weiss, D. S., 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, *497*(7448), 254-257.

Sander, J. D., ve Joung, J. K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology, 32(4), 347-355.

Scholze, H., ve Boch, J., 2010. TAL effector-DNA specificity. Virulence, 1(5), 428-432.

Shah, S. A., Hansen, N. R., ve Garrett, R. A., 2009. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., ... ve Skarnes, W. C., 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nature methods, 11(4), 399-402.

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., ve Mateo, J. L., 2015. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. PloS one, 10(4), e0124633.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., ... ve Joung, J. K., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature biotechnology, 32(6), 569-576.

Terns, M.P., ve Terns, R.M., 2011. CRISPR-based adaptive immune systems. Current Opinion in Microbiology, 14(3):321–327.

Upadhyay, S. K., ve Sharma, S., 2014. SSFinder: high throughput CRISPR-Cas target sites prediction tool. BioMed research international, 2014.

Upadhyay, S. K., Kumar, J., Alok, A., ve Tuli, R., 2013. RNA-guided genome editing for target gene mutations in wheat. *G3: Genes, Genomes, Genetics*, *3*(12), 2233-2238.

Wang, H., La Russa, M., ve Qi, L. S., 2016. CRISPR/Cas9 in genome editing and beyond. Annual review of biochemistry, 85, 227-264.

Wilson, L. O., O'Brien, A. R., ve Bauer, D. C., 2018. The current state and future of CRISPR-Cas9 gRNA design tools. Frontiers in pharmacology, 9, 749.

Xiao, A., Wu, Y., Yang, Z., Hu, Y., Wang, W., Zhang, Y., ... ve Zhang, B., 2012. EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering. Nucleic acids research, 41(D1), D415-D422.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., ve Zhang, B., 2014. CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics, 30(8), 1180-1182.

Xie, S., Shen, B., Zhang, C., Huang, X., ve Zhang, Y., 2014. sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PloS one, 9(6), e100448.

Xiong, Y., Xie, X., Wang, Y., Ma, W., Liang, P., Songyang, Z., ve Dai, Z., 2017. pgRNAFinder: a web-based tool to design distance independent paired-gRNA. Bioinformatics, 33(22), 3642-3644.

Zhu, L. J., 2020. CRISPRseek Workshop.

Zhu, L. J., Holmes, B. R., Aronin, N., ve Brodsky, M. H., 2014. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PloS one, 9(9), e108424.

Zou, H., ve Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.

ÖZGEÇMİŞ

Gaziantep İnci Konukoğlu Anadolu Lisesinden mezun oldu. 2012 yılında Fırat Üniversitesi Biyomühendislik Mühendisliği Bölümü'nde başladığı lisans öğrenimini, 2017 yılında tamamladı. 2020 yılının şubat ayında Çukurova Üniversitesi Fen Bilimleri Enstitüsü Biyoteknoloji Anabilim Dalı'nda yüksek lisans eğitimine başladı. 2020 haziran ayında Tezkim Tarımsal Kimya A.Ş.'nde Biyomühendis olarak işe başladı. 2021 yılının haziran ayından itibaren Türkiye Sağlık Enstitüleri Başkanlığı'nın (TÜSEB) Türkiye Biyoteknoloji Enstitüsünde Proje Personeli olarak çalışmaya devam etmektedir. Çukurova Üniversitesi Biyoteknoloji Anabilim Dalı'nda devam ettiği yüksek lisans öğrenimini 2022 yılının şubat ayı itibari ile tamamlamıştır.