

Vaka Analizi: Biyoinformatik

Case 1

Nanopore teknolojisi kullanılarak bir gölde yılın farklı zamanlarında, gölün farklı derinliklerinden örnek alınarak göldeki bakteriyel toplulukların zamansal ve derinliğe bağlı değişimlerini analiz etmek istiyorsunuz.

- Nasıl bir sekanslama yöntemi tercih edersiniz? (Lütfen tek bir yöntem seçiniz)
- Nasıl bir biyoinformatik iş akış yöntemi izlersiniz? (Lütfen seçtiğiniz yönteme uygun tek bir biyoinformatik iş akışı yazınız.)

Eğer bu göldeki arkeal değişimleri de incelemek isteseydiniz, birebir aynı yöntem ve biyoinformatik iş akışını kullanmaya devam edebilecek miydiniz?

VAKA ANALİZİ DEĞERLENDİRMESİ

Case 2

Sars-CoV-2 enfekte ve kontrol grupları arasındaki genetik ve biyolojik farklılıkları transkripsiyon verilerini kullanarak belirlemek istiyorsunuz. [Size verilen veri setinde](#), Sars-CoV-2 enfeksiyonuna maruz kalmış ve maruz kalmamış (mock) NHBE hücrelerindeki gen ifade verileri bulunmaktadır. Bu veriler, her bir koşul için üç tekrar içermektedir: mock_rep1, mock_rep2, mock_rep3 ve sars_cov_rep1, sars_cov_rep2, sars_cov_rep3. Sırasıyla aşağıdaki görevleri yerine getirmeniz bekleniyor.

1- Filtreleme: Hiçbir koşulda ifade göstermeyen (read içermeyen) genleri veri setinden çıkarın.

2- Diferansiyel İfade Analizi: İki koşul arasında anlamlı olarak değişen genleri belirleyin. Bunun için p-value 0.05 sınır değerini kullanabilirsiniz.

3- Enrichment Analizi: Anlamlı değişmiş gen listesine gProfiler (<http://biit.cs.ut.ee/gprofiler/gost>) kullanarak enrichment analizi uygulayın. GO:MF, GO:BP, KEGG ve REAC sonuçlarını raporlayın (Ekran görüntüsü olarak veya tablo olarak paylaşabilirsiniz).

4- Gen ID Dönüşümü: Anlamlı değişen genlerin Ensembl ID'lerini Entrez ID'lere dönüştürün. Bunun için size sunulan converter.tsv dosyasını kullanabilirsiniz ve pandas.merge komutunu uygulayabilirsiniz.

5- Sinyal Yolu Analizi: Enrichment analizi sonucundan elde edilen ilişkili sinyal yollarını listeleğin. Anlamlı değişen genleri bu sinyal yolları üzerinde KEGG Pathway (<https://www.genome.jp/kegg/pathway.html>) kullanarak haritalayın. Bir önceki görevde elde ettiğiniz Entrez ID'leri bu haritalama sürecinde kullanabilirsiniz.

6- Sonuçların Yorumlanması: Elde ettiğiniz sonuçları yorumlayın. Ayrıca aşağıdaki sorulardan bir veya birkaçını cevaplayabilirsiniz.

- Herhangi bir disease veya viral enfeksiyon ile ilgili bir terim elde ettiniz mi?
- Enfeksiyon sonrası hücrelerin metabolik süreçleri etkilenmiş mi?
- P-value sınır değerini arttırmak veya azaltmak sonuçlarınızı nasıl değiştirebilirdi?

VAKA ANALİZİ DEĞERLENDİRMESİ

Aday Bilgileri

İsim	Sanem
Soyisim	Coşkun

Case 1

Oxford Nanopore Technology web sitesindeki kaynakları taradığımda [1],[2] en verimli **sekanslama yönteminin Nanopore MinION™** olduğunu gördüm. Bu yöntem uzun okumalar ve taşınabilirliği ile göl örnekleri gibi çevresel,mikrobiyom DNA analizlerinde öne çıkmaktadır.

Biyoinformatik İş Akışı:

1- Örnek hazırlama ve DNA ekstraksiyonu

Gölün farklı derinliklerinden ve yılın farklı zamanlarında su örnekleri alınır.

DNA'yı ekstrakte etmek için uygun bir kit kullanılır.

2- Kütüphane Hazırlığı

Ekstrakte ettiğimiz DNA'yı Nanopore sekanslamaya hazırlamak için PCR ile 16S rRNA geninin tamamını amplifiye eden Barcoding kit kullanılır. Daha sonra ise sekanslama adaptörleri eklenir

3- Sekanslama

Sekanslama için kullanacağımız MinION cihazı hazırlanır ve hazırlanan DNA kütüphanesi cihazdaki Flow Cell' e yüklenir.

Basecaller kullanarak veriler toplanır.

4- Ham veri işleme

Okuma verilerini nükleotid dizilerine dönüştürmek için çeşitli basecalling yazılımı kullanılır ve düşük kaliteli okuma verileri filtrelenir.

5- Montaj (Assembly)

Temizlenmiş veriler montajlanarak contigler oluşturulur.

6- Taksonomik Sınıflandırma

Contigler bakteriler için uygun olan SILVA gibi referans veritabanlarına karşı hizalanarak sınıflandırılır. (QIIME2 gibi araçlar kullanılabilir)

7- Topluluk yapısı analizi:

Sınıflandırılmış veriler ile bakteriyel topluluk yapısı ve çeşitlilik analiz edilir.

Çeşitlilik indeksleri ve topluluk kompozisyonu analizleri yapılır.

NMDS gibi yöntemlerle zamansal ve derinliğe bağlı değişimler görselleştirilir.

Arkeal Değişim

Biyolojik iş akışında DNA arkeal 16S rRNA genine özgü primerler kullanarak amplifiye edilir aynı zamanda veri analizi yaparken kullanacağımız veritabanı arkealara özgü(SILVA Archaea) olmalı ve buna göre sınıflandırma yapılmalı.

Bunlar dışında iş akışı arkea toplulukları analizi içinde uygundur

Case 2

1-Filtreleme:

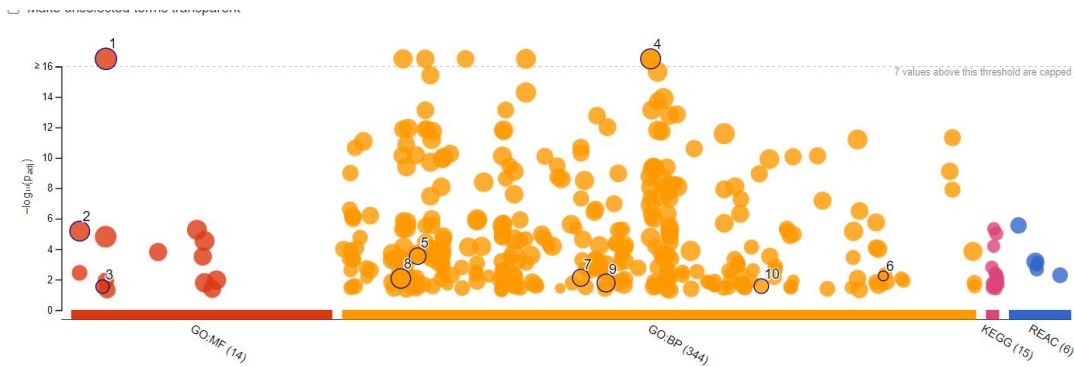
Python pandas kütüphanesi ile hiç bir koşulda ifade içermeyen genleri filtreledim. Filtrelediğim veri setini yeni .tsv uzantılı bir dosya olarak kaydettim

2-Diferansiyel İfade Analizi:

Filtrelediğim veri setini okuyarak mock ve sars sütunlarını birleştirdim. Her biri için ttest'i uyguladım ve p value değerlerini elde ettim. Bu değerler arasından 0.05 değerinden küçük olanları filtreledim.

3-Enrichment Analizi Sonuçlar:

Gprofiler sitesine giderek elde ettiğim anlamlı değişen gen listesini kopyaladım ve analiz ettim.



Örnek KEGG tablosunun ekran alıntısını koydum. Eğer diğer tablolara da erişmek istiyorsanız link ile tüm tabloları görüntüleyebilirsiniz. Daha doğru sonuçlar elde etmek için adjusted p value değerini kullanmalıyız. 2. Adımda bundan bahsetmediği için analizi sadece p value kullanarak yaptım.

<https://r.resimlink.com/8Wlr9Ai.png>

KEGG		stats	
<input type="checkbox"/> Term name	Term ID	P _{adj}	$-\log_{10}(P_{adj})$
<input type="checkbox"/> TNF signaling pathway	KEGG:04668	4.718×10^{-6}	
<input type="checkbox"/> Influenza A	KEGG:05164	9.665×10^{-6}	
<input type="checkbox"/> IL-17 signaling pathway	KEGG:04657	6.547×10^{-5}	
<input type="checkbox"/> NF-kappa B signaling pathway	KEGG:04064	1.564×10^{-3}	
<input type="checkbox"/> AGE-RAGE signaling pathway in diabetic complications	KEGG:04933	3.759×10^{-3}	
<input type="checkbox"/> Amoebiasis	KEGG:05146	4.398×10^{-3}	
<input type="checkbox"/> Rheumatoid arthritis	KEGG:05323	5.951×10^{-3}	
<input type="checkbox"/> Fluid shear stress and atherosclerosis	KEGG:05418	9.470×10^{-3}	
<input type="checkbox"/> Epstein-Barr virus infection	KEGG:05169	1.030×10^{-2}	
<input type="checkbox"/> Apoptosis	KEGG:04210	1.777×10^{-2}	
<input type="checkbox"/> Lipid and atherosclerosis	KEGG:05417	2.026×10^{-2}	
<input type="checkbox"/> Measles	KEGG:05162	2.519×10^{-2}	
<input type="checkbox"/> NOD-like receptor signaling pathway	KEGG:04621	2.815×10^{-2}	
<input type="checkbox"/> Necroptosis	KEGG:04217	3.486×10^{-2}	
<input type="checkbox"/> Legionellosis	KEGG:05134	4.320×10^{-2}	

4-Gen ID Dönüşümü:

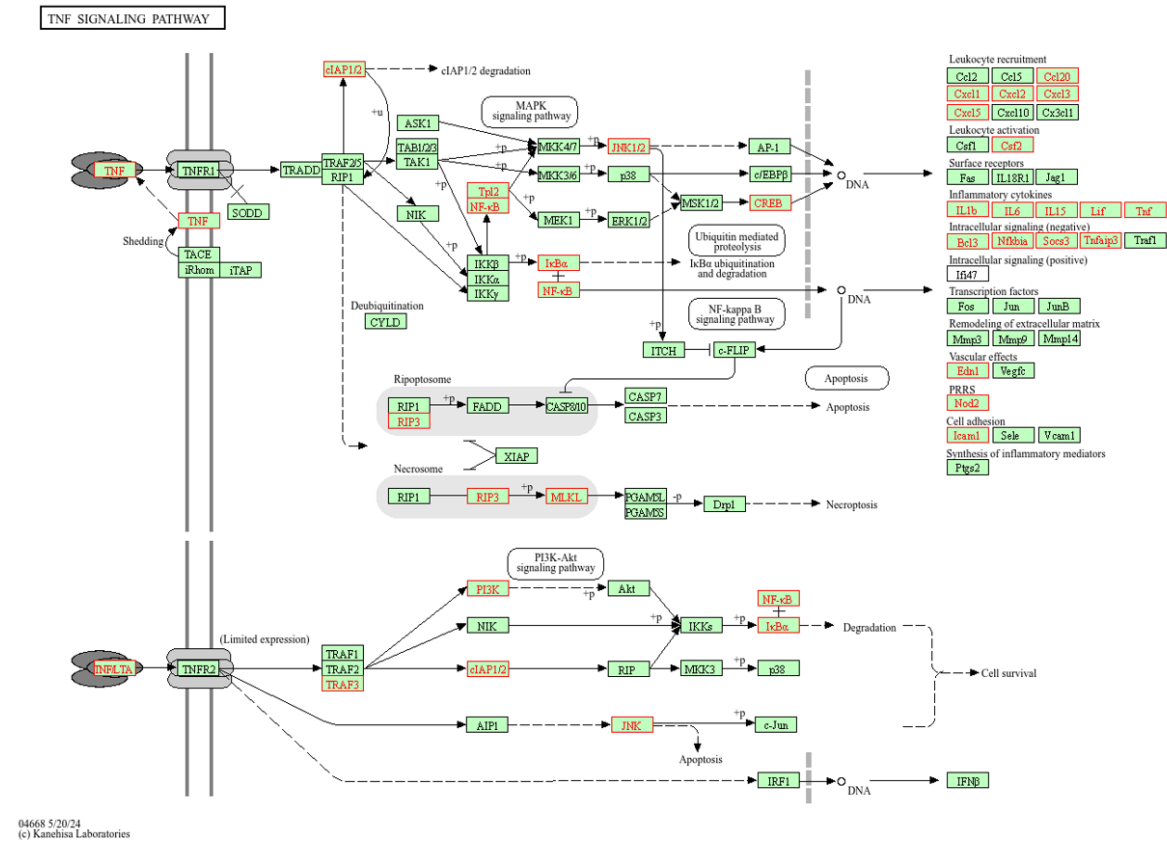
Galaxy bioinformatics' in [3] annotateMyID toolu ile elimdeki ENSEMBL ID'leri ENTREZ ID' lere çevirdim. Çevirdiğimde ENTREZ ID sayısında artış gözlemlendim ve filtreleyip incelediğimde bazı ENSEMBL ID' lerin farklı alt türler için birden fazla ENTREZ ID' ye sahip olduğunu gördüm.

ENSG00000288905 ID' si için 2 farklı ENTREZ ID çıkmıştır.

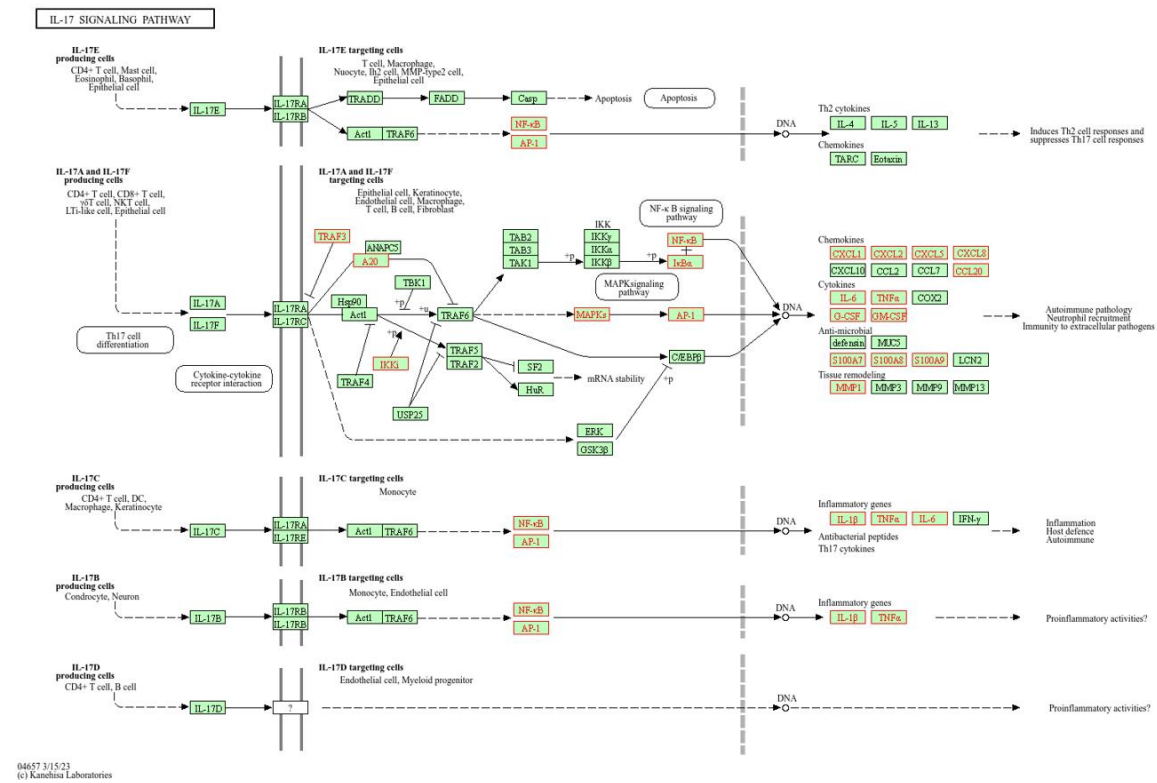
howing 25 of 2 Results for ENSG00000288905 Search Time: 0 ms

	Symbol ▲	Description	Category ▲ ?
1	LSP1P5	LSP1 Pseudogene 5	Pseudogene
2	LOC107985200	Uncharacterized LOC107985200	RNA Gene (ncRNA)

[4]

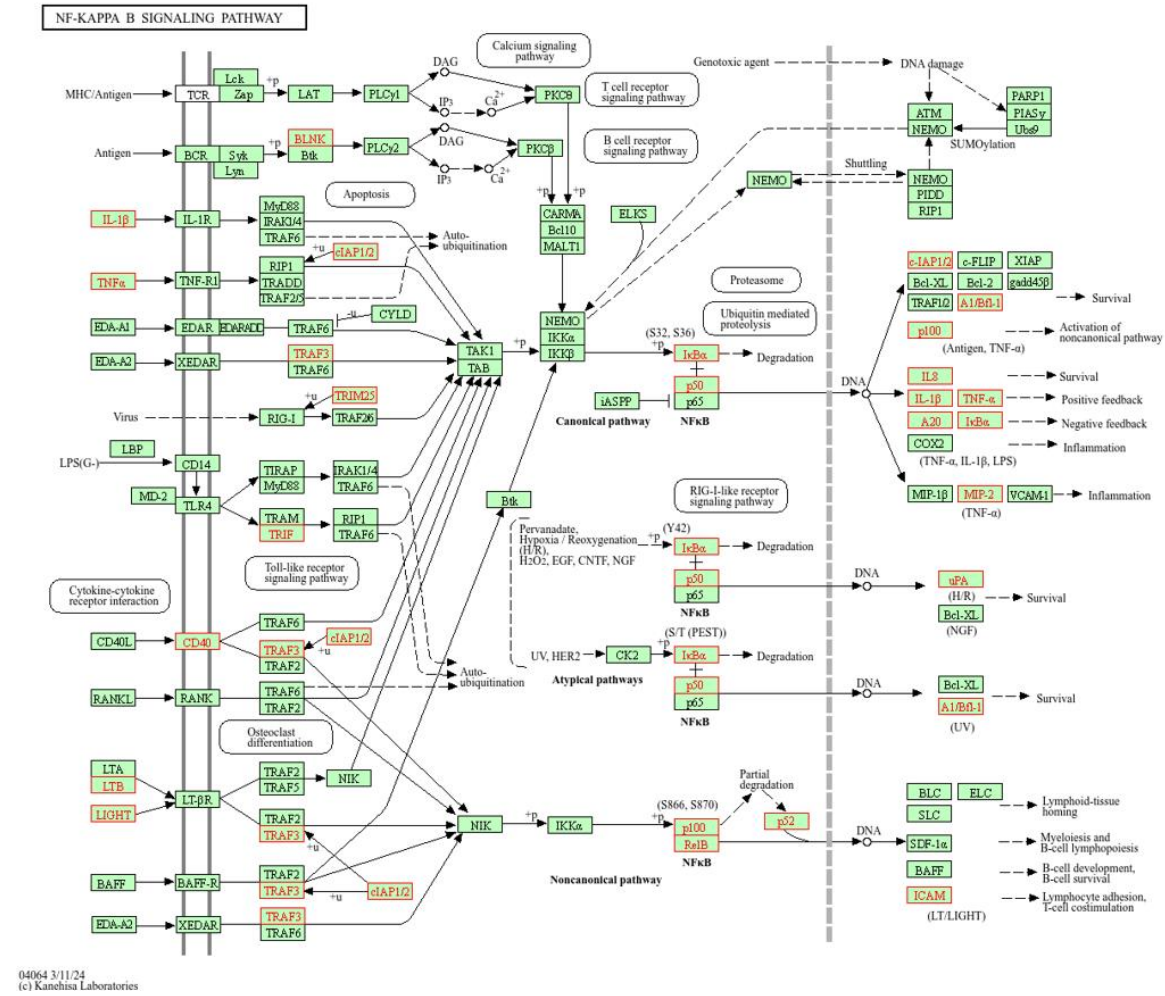
5- Sinyal Yolu Analizi:**KEGG:04668**
<https://r.resimlink.com/hiSnId.png>

KEGG:04657



<https://r.resimlink.com/BHf3EvmxW.png>

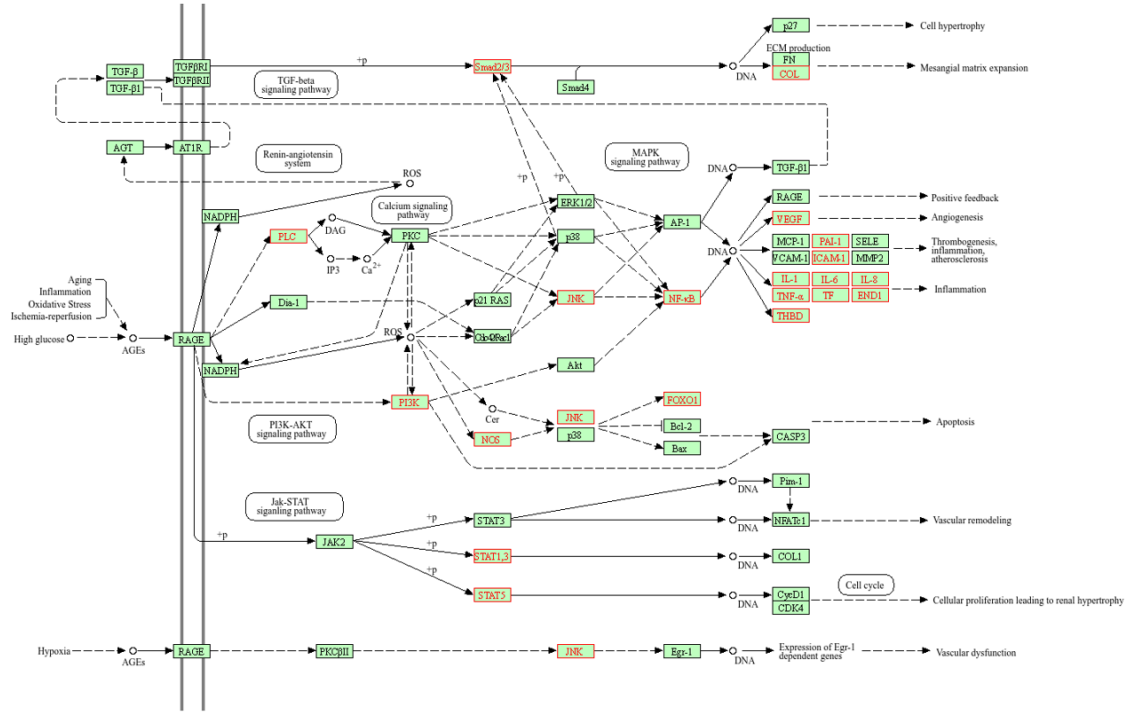
KEGG:04064



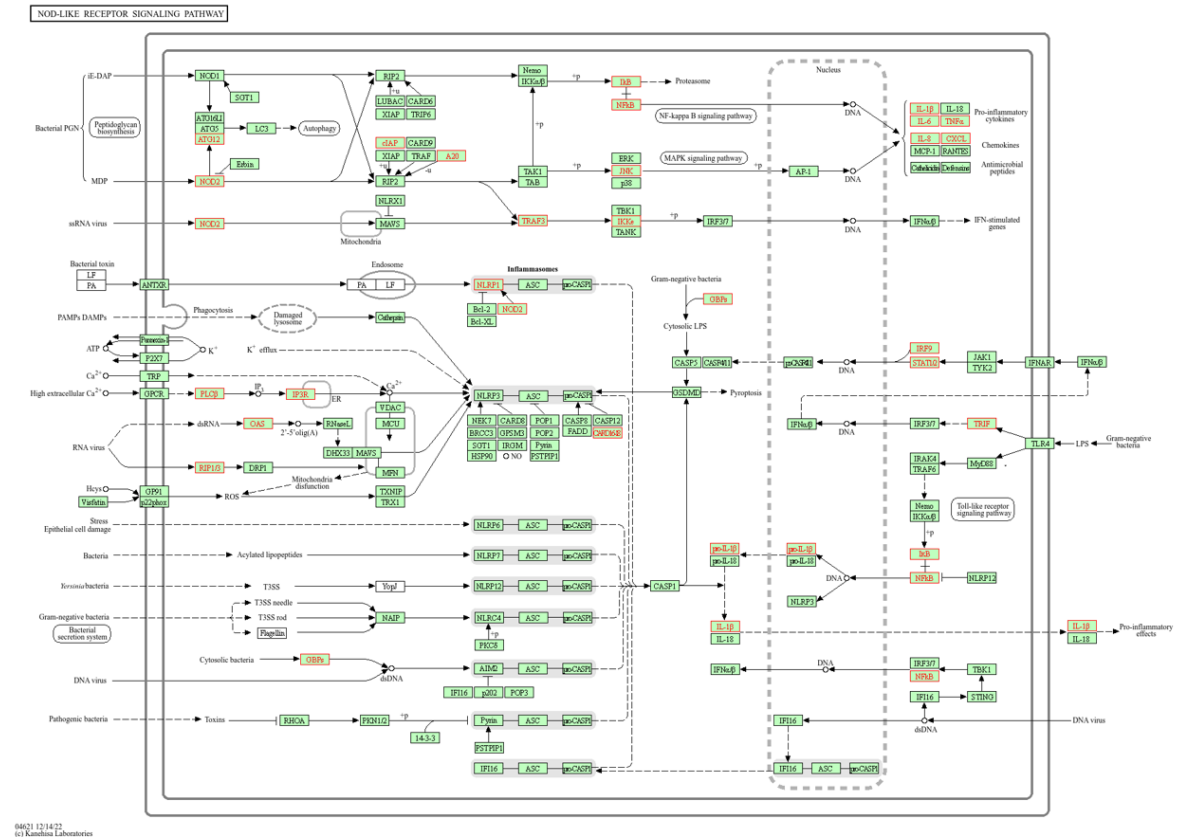
<https://r.resimlink.com/8SwO6p.png>

KEGG:04933

AGE-RAGE SIGNALING PATHWAY IN DIABETIC COMPLICATIONS



KEGG:04621



<https://r.resimlink.com/QwqnC8.png>

6-Sonuçların yorumlanması**AGE-RAGE Signaling Pathway in Diabetic Complications (KEGG:04933)**

Diyabetik komplikasyonlar sırasında gelişen oksidatif stres, inflamasyon ve vasküler disfonksiyon gibi süreçlerde rol oynar. Haritada görülen RAGE (Receptor for Advanced Glycation End-products) aktivasyonu, çeşitli inflamatuvar genlerin ve proteinlerin ekspresyonunu artırarak hücrel hasara ve vasküler remodeling'e neden olabilir.

NF-kappa B Signaling Pathway (KEGG:04064)

NF-kappa B, bağışıklık ve inflamasyon yanıtlarını düzenleyen önemli bir sinyal yoludur. Haritada, bu yolun aktivasyonu sonucu çeşitli inflamatuvar genlerin ve proteinlerin (örneğin, IL-1, TNF- α) ekspresyonu artar. Anlamli değişen genler, bu yolun aktivasyonunda veya baskılanmasında rol oynayabilir, böylece inflamatuvar hastalıkların gelişiminde etkili olabilirler.

IL-17 Signaling Pathway (KEGG:04657)

IL-17, inflamatuvar yanıtları düzenleyen bir sitokindir ve otoimmün hastalıklarda önemli bir rol oynar. Haritada, IL-17'nin aktivasyonu ile çeşitli inflamatuvar ve immün yanıt genlerinin ekspresyonu görülmektedir. Anlamli değişen genler, bu yolun aktivasyonunu artırabilir veya azaltabilir, böylece inflamasyon ve otoimmün yanıtların düzenlenmesinde kritik rol oynayabilirler.

TNF Signaling Pathway (KEGG:04668)

TNF (Tümör Nekroz Faktörü), inflamasyon ve hücre ölümü süreçlerini düzenleyen bir sitokindir. Haritada, TNF'nin aktivasyonu ile hücre ölümü (apoptosis) ve inflamatuvar yanıtların (örneğin, IL-6, IL-8) arttığı görülmektedir. Anlamli değişen genler, TNF sinyal yolunun çeşitli kademelerinde yer alarak inflamasyon ve hücre ölümünü etkileyebilirler.

NOD-like Receptor Signaling Pathway (KEGG:04621)

NOD-like reseptörler, hücrel stres ve patojenlere karşı yanıtları düzenleyen önemli sensörlerdir. Haritada, bu reseptörlerin aktivasyonu ile inflamatuvar yanıtlar ve hücrel stres yanıtları görülmektedir. Anlamli değişen genler, bu reseptörlerin sinyal yolundaki rollerini etkileyerek inflamasyon ve hücrel stres yanıtlarını düzenleyebilirler.

VAKA ANALİZİ DEĞERLENDİRMESİ

Genel Değerlendirme:

Elde ettiğimiz sonuçlar, inflamatuvar yanıtlar, hücresel stres ve diyabetik komplikasyonlar gibi biyolojik süreçlerin düzenlenmesinde rol oynayan sinyal yollarını ortaya koymaktadır. Anlamlı değişen genler, bu yolların çeşitli aşamalarında yer alarak inflamasyon, hücre ölümü ve hücresel stres yanıtlarını etkileyebilir. Bu sinyal yollarının anlaşılması, ilgili hastalıkların patogenezi anlamak ve potansiyel tedavi hedeflerini belirlemek açısından büyük önem taşır.

Bu sinyal yollarının her biri, belirli hastalık durumlarının gelişiminde kritik rol oynar ve elde edilen gen değişiklikleri bu süreçlerin nasıl etkilendiğini göstermektedir. Elde edilen veriler, biyolojik süreçlerin daha iyi anlaşılması ve potansiyel tedavi stratejilerinin geliştirilmesi için kullanılabilir. Elde edilen haritalar, biyolojik süreçlerin ve hastalıkların moleküler temelini anlamak için değerli bilgiler sunar. Anlamlı değişen genler, bu yolların çeşitli aşamalarında yer alarak biyolojik yanıtları modüle eder. Bu haritalar, hücresel ve moleküler düzeydeki etkileşimleri görselleştirerek, araştırmacılara hastalıkların patogenezi anlamada ve potansiyel tedavi hedefleri belirlemede yol gösterir.

Sonuç olarak, bu haritalar, inflamasyon, hücresel stres ve diyabetik komplikasyonlar gibi süreçlerin anlaşılmasında kritik öneme sahiptir. Elde edilen veriler, biyolojik süreçlerin daha derinlemesine incelenmesi ve tedavi stratejilerinin geliştirilmesi için kullanılabilir. Bu haritalar, genlerin ve proteinlerin biyolojik rolleri hakkında daha fazla bilgi sağlayarak, hastalık mekanizmalarının çözülmesine katkıda bulunur. Bu sayede, hastalıkların moleküler düzeydeki nedenleri ve bunların nasıl hedeflenebileceği konusunda daha kapsamlı ve detaylı bir anlayış geliştirilir.

VAKA ANALİZİ DEĞERLENDİRMESİ

Herhangi bir disease veya viral enfeksiyon ile ilgili bir terim elde ettiniz mi?

Enrichment analizi yaptığım sırada KEGG sonuçlarında disease ve viral hastalıklarla ilişkili terimlerle karşılaştım. Influenza A aralarında değeri en yüksek oldu. Influenza A, Epstein-Barr virus infection, Measles gibi viral enfeksiyonlarla, Amoebiasis gibi paraziter ve Legionellosis gibi bakteriyel enfeksiyonlarla karşılaştım. Buna ek olarak otoimmün bir hastalık olan Rheumatoid arthritis'i de sonuçlarımda gördüm.

Enfeksiyon sonrası hücrelerin metabolik süreçleri etkilenmiş mi?

Metabolik süreçlerin etkilenip etkilenmediğini değerlendirmek için sinyal yollarına bakabiliriz. **AGE-RAGE Signaling Pathway in Diabetic Complications (KEGG:04933)**, **NF-kappa B Signaling Pathway (KEGG:04064)**, **IL-17 Signaling Pathway (KEGG:04657)** sinyal yolları enfeksiyon sonrası metabolik süreçleri etkilendiğini gösterir. Bu yollara bakarak nasıl etkilendiğini de gözlemleyebiliriz.

P-value sınır değerini arttırmak veya azaltmak sonuçlarınızı nasıl değiştirebilirdi?

Eğer sınır değerini azaltırsak daha anlamlı değerler karşımıza çıkar ve bu da anlamlı yollar seçmemizi sağlar. Sonuçların güvenilirliğini artırabilir fakat bazı önemli yolların da gözden kaçmasına sebep olabilir.

Eğer sınır değerini artırırsak daha fazla anlamlı yolun kabul edilmesine yol açar fakat bu bazen yanıltıcı yolların dahil edilmesine sebep olabilir.

Özetle p-değeri sınırı sonuçları ve güvenilirliği doğrudan etkiler. Araştırmanın amacı ve verilere göre uygun bir değer seçmek bizim için çok önemlidir.

Ayrıca sonuçların daha güvenilir olması için **adjusted p-value** (düzeltilmiş p-değeri) kullanmak daha doğru bir yaklaşım olacaktır. Adjusted p-value, çoklu test düzeltmesi yapılarak elde edilen p-değeri ve yanlış pozitif sonuçların (false positives) azaltılmasına yardımcı olur.

Kaynakça

- [1]<https://gigascience.biomedcentral.com/articles/10.1186/s13742-016-0111-z>
- [2]<https://nanoporetech.com/resource-centre/oxford-nanopore-based-metagenomic-study-high-altitude-permafrost-microbiome>
- [3]https://usegalaxy.org/?tool_id=toolshed.g2.bx.psu.edu%2Frepos%2Fuc%2Fannotatemyids%2Fannotatemyids%2F3.18.0%2Bgalaxy0&version=latest
- [4]<https://www.genecards.org/Search/Keyword?queryString=ENSG00000288905>