

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK ÖLÇEKLİ GENOMİK VERİNİN ÇOK YÖNLÜ
ANALİZİ**

DUYGU DEDE

YÜKSEK LİSANS TEZİ

2013

YÜKSEK ÖLÇEKLİ GENOMİK VERİNİN ÇOK YÖNLÜ ANALİZİ

MULTIWAY ANALYSIS OF HIGH THROUGHPUT GENOMIC DATA

DUYGU DEDE

Başkent Üniversitesi
Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin
BİLGİSAYAR Mühendisliği Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2013

Yüksek Ölçekli Genomik Verinin Çok Yönlü Analizi başlıklı bu çalışma, jürimiz tarafından, 23./07/2013 tarihinde, **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI** 'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan : Doç. Dr. Mustafa KOCAKULAK

Üye (Danışman) : Doç. Dr. Hasan OĞUL

Üye : Yrd. Doç. Dr. Emre SÜMER

ONAY

.../.../.....

Prof. Dr. Emin AKATA
Fen Bilimleri Enstitüsü Müdürü

TEŐEKKÜR

Sayın Doç. Dr. Hasan OĐUL' a (tez danışmanı), alıőmanın sonuca ulaőtırılmasında ve karşılaşılan güçlüklerin aőtılmasında her zaman yardımcı ve yol gösterici olduđu için...

Canım oda arkadaşım Didem Ölçer'e, her zaman kurtarıcım olduđu için...

Biricik hayat arkadaşım Çađrı Őener'e, her zaman yardımcı olduđu için...

Ve ok deđerli aileme, her zaman yanımda oldukları için...

Bu tez alıőması Baőkent Üniversitesi Araőtırma Projesi kapsamında BA12/FM-10 nolu proje ve kısmen TUBİTAK 110E160 nolu proje ile desteklenmiőtir.

ÖZ

YÜKSEK ÖLÇEKLİ GENOMİK VERİNİN ÇOK YÖNLÜ ANALİZİ

Duygu DEDE

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Gelişen mikroçip teknolojileri sayesinde gen ifade (gene expression) verilerinin ölçümleri yapılarak yüksek ölçekli (high throughput) veriler elde edilebilmekte ve bu veriler çok boyutlu matrislerde saklanabilmektedir. Yüksek ölçekli verinin analizi, gerek işlemsel karmaşıklık yönüyle, gerekse anlamlı bilgi çıkarımındaki zorluklar nedeniyle üzerinde çalışılması gereken önemli bir konudur. Son yıllarda özellikle yaşam bilimleri alanında bu tür veri üretiminin fazla olması ve genellikle bu verilerin analizlerinin gözle veya standart yöntemlerle yapılamaması nedeniyle, problem daha kritik bir hale gelmiştir. Bu tür veriler, yapılan deneyler ve kullanılan düzenekler gereği, çok yönlü ve/veya çok kaynaklı bir yapıya sahip olabilmektedir. Bu çalışmada belirtilen motivasyondan yola çıkılarak, akıllı veri analizi yeterliliklerini artıracak bir yöntemin geliştirilmesi hedeflenmiş ve bir yazılım aracının gerçekleştirimi yapılmıştır. Geliştirilen uygulamada iki veya daha fazla organizmadan (örneğin; insan, fare ve maymun) alınan, aynı koşullarda elde edilmiş mRNA (mesajcı RNA) ifade verileriyle türler arası (çok kaynaklı) bir analiz hedeflenmiştir. Bunun için yeni bir üç-yönlü kümeleme yöntemi (TriClustering) geliştirilmiş ve bu yöntem geliştirilen yazılım aracı ile NCBI'S GEO veri merkezinden alınan üç farklı veri kümesine uygulanmıştır. Gen ontoloji (GO) terimi zenginleşme analizi ve Dunn indeks (DI) metriği kullanılarak sonuçların biyolojik ve istatistiksel değerlendirilmesi yapılmıştır. Deneysel sonuçlara göre TriClustering yöntemi anlamlı üçlü kümeler (tricluster) bulabilen, türler arası analiz için kullanışlı bir araçtır.

ANAHTAR SÖZCÜKLER: üç-yönlü kümeleme, ikili kümeleme, eş zamanlı kümeleme, çok kaynaklı, çok yönlü analiz, biyolojik veri analizi.

Danışman: Doç.Dr. Hasan OĞUL, Başkent Üniversitesi, Bilgisayar Mühendisliği Bölümü.

ABSTRACT

MULTIWAY ANALYSIS OF HIGH THROUGHPUT GENOMIC DATA

Duygu DEDE

Baskent University Institute of Science and Engineering

Department of Computer Engineering

High throughput data can be generated by developing techniques and represented in large matrices. Analysis of such data has become one of the major tool but it has brought along many challenges for biological data mining such as process complexities and difficulties in information retrieval. In recent years these challenges have become critic since huge amount of data is produced especially in life sciences and standard techniques may not be used in analysing of such data. This kind of data structure may be multi-way and/or multi-source by used test and devices. In this study based on the specified motivation, a novel method and software tool are developed to increase proficiency of data analysis techniques. The developed method aims to make cross species (multi sources) analysis using mRNA expression values obtained from different organisms (human, mouse, monkey etc.) under same conditions. To achieve this goal a novel three way clustering method named TriClustering is introduced and the method has been applied to three different gene expression data obtained from NCBI's GEO data collection. Biological and statistical significance of the results are evaluated using Gene Ontology (GO) term enrichment analysis and Dunn index (DI) metric, respectively. The experimental results indicate that TriClustering on multi-organism data can be resulted with better gene clusters in comparison to biclustering on single-organism data. The method also promote a useful tool for cross species gene regulation analysis.

KEYWORDS; triclustering, three-way clustering, biclustering, simultaneous clustering, , multi-source, multi-way analysis, biological data analysis.

Advisor: Assoc. Prof. Dr. Hasan OĞUL, Başkent University, Department of Computer Engineering.

İÇİNDEKİLER LİSTESİ

Sayfa

ÖZ	i
ABSTRACT	ii
SİMGELER VE KISALTMALAR LİSTESİ	v
ÇİZELGELER LİSTESİ	vii
1. GİRİŞ	1
1.1 Gen İfadesi (Gene Expression)	2
1.2 Mesajcı RNA (mRNA)	3
1.3 Yüksek Ölçekli Veri (High Throughput Data)	3
1.4 Çok Yönlü Analiz (Multi-Way Analysis)	4
1.5 Çok Kaynaklı Analiz (Multi-Source Analysis)	5
1.6 Kümeleme (Clustering) Yöntemi	6
1.7 İkili Kümeleme-Çift-Yönlü Kümeleme (Biclustering) Yöntemi	9
1.8 GO (Gene Ontology) Analizi	10
2. ÖNCEKİ ÇALIŞMALAR	14
2.1 Gen İfade Veri Analizi	14
2.2 Kümeleme Yöntemi	14
2.3 İkili Kümeleme (Biclustering) Yöntemi	15
2.4 Çok-Yönlü Kümeleme Yöntemleri	19
2.5 Türler Arası Analiz	21
3. YÖNTEMLER	24
3.1 Üçlü Küme (Tricluster) Modeli	24
3.2 TriClustering Algoritması	25
3.2.1 Algoritma-1: Tek bileşen silinmesi (Single node deletion)	26
3.2.2 Algoritma-2: Çoklu bileşen silinmesi (Multiple node deletion)	27

3.2.3	Algoritma-3: Bileşen eklenmesi (Node addition).....	28
3.2.2	Algoritma-4: TriClustering algoritması	29
4.	GELİŞTİRİLEN ARAÇ	30
4.1	İşlevler ve Kullanıcı Arayüzleri	30
4.2	Teknik Altyapı	38
5.	VERİ KÜMELERİ	39
6.	SONUÇLAR	41
7.	TARTIŞMA VE GELECEK ÇALIŞMALAR	45
	KAYNAKLAR LİSTESİ	47
	EKLER LİSTESİ.....	52

SİMGELER VE KISALTMALAR LİSTESİ

DNA	Deoksiribonükleik asit
RNA	Ribonükleik asit
mRNA	Mesajcı RNA
tRNA	Taşıyıcı RNA
rRNA	Ribozomal RNA
cDNA	Bütünleyici DNA
mİRNA	MikroRNA
GO	Gen Ontolojisi
BAP	Bilimsel Araştırma Projesi
TÜBİTAK	Türkiye Bilimsel ve Teknolojik Araştırma Kurumu

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1. 1	Örnek mikroçip deneyi görünümü	3
Şekil 1. 2	Gen ifade matrisi.....	4
Şekil 1. 3	Tek-yönlü, Tek-kaynaklı analiz.....	5
Şekil 1. 4	Çok-yönlü analiz	5
Şekil 1. 5	Çok-kaynaklı veri	6
Şekil 1. 6	Çok yönlü-Çok kaynaklı analiz birleşimi.....	6
Şekil 1. 7	Kümeleme sonucu oluşan üç farklı kümenin farklı renkler ile gösterimi ...	7
Şekil 1. 8	Yığınsal ve Ayrıştırıcı Kümeleme yöntemlerinin dendogram gösterimi	8
Şekil 1. 9	k-ortalamlar kümeleme örneği.....	9
Şekil 1. 10	İkili kümeleme örneği [9]	10
Şekil 1. 11	GO açıklaması dosya örneği [11].....	11
Şekil 1. 12	GO yapısının diyagram gösterimi.....	13
Şekil 2. 1	BicAT yazılım aracı	17
Şekil 2. 2	BicOverlapper yazılım aracı	18
Şekil 2. 3	STEM yazılım aracı.....	19
Şekil 2. 4	BiNGO yazılım aracı	20
Şekil 2. 5	Türler arası analiz örneği [33].....	23
Şekil 3. 1	TriClustering Algoritmasının örnek bir veri kümesine uyarlanması	26
Şekil 4. 1	Triclustering giriş arayüzü	30
Şekil 4. 2	Orijinal verinin gen ifade verisi görünümü	31
Şekil 4. 3	Ön işlenmiş verinin gen ifade verisinin görünümü.....	32
Şekil 4. 4	Orijinal verinin ısı haritası görünümü.....	33
Şekil 4. 5	TriClustering algoritmasının çalıştırma arayüzü	33
Şekil 4. 6	Oluşan üçlü kümelerden birinin gen ifade verisi görünümü.....	34
Şekil 4. 7	Oluşan üçlü kümelerden birinin ısı haritası görünümü	35
Şekil 4. 8	Oluşan üçlü kümelerden birinin analiz görünümü	36
Şekil 4. 9	Oluşan üçlü kümelerden birinin (Tricuster1) GO analiz görünümü.....	37
Şekil 4. 10	Üçlü kümelerde gen araması yapılması	37

ÇİZELGELER LİSTESİ

Sayfa

Çizelge 6. 1 Parametrelerin deęiřimi	42
Çizelge 6. 2 GO terimi zenginleřen gen gruplarının veri kümelerine göre yüzdesi..	43
Çizelge 6. 3 Örnek bir gen grubunun GO analiz sonuçları	44
Çizelge 6. 4 Dunn indeks sonuçları	44

1. GİRİŞ

Genler canlıların üretim şifresini taşıyan biyolojik yapılardır. Canlılar hakkında merak edilen önemli soruların cevabı gen ifade verilerinin analizi ile verilebilmektedir. Gen ifade verisi analizi, türler hakkında detaylı bilgi edinme, türler arası benzerlikleri ortaya çıkarma, çeşitli hastalıkların sebebini araştırma, bu hastalıklara uygun tedavi yöntemlerinin tespiti amacı ile sıklıkla kullanılan bir yöntem haline gelmiştir. Bu amaçlar doğrultusunda çok sayıda gen ifade verisi analizi yöntemi geliştirilmiştir.

Bu tezde önceki yöntemlerden farklı olarak sadece bir türün, organizmanın (örneğin; insan, fare ve maymun) gen ifade verisini baz alarak değil birden fazla organizmanın gen ifade bilgisini alarak türler arası analizin yapılması, benzerliklerin açığa çıkarılması için bir yöntemin ve yazılım aracının geliştirilmesi hedeflenmektedir. Geliştirilen araç yaşam bilimlerinden elde edilen gerçek veriler üzerinde test edilmiştir. Elde edilen analiz sonuçları biyolojik anlamda tahmin yeterliliği açısından değerlendirilmiştir. Böylece projenin biyolojik motivasyonunu oluşturan; farklı ortamlarda farklı kaynaklardan elde edilen gen ifade verilerinin biyolojik olarak anlamlandırılması problemi için bir çözüm gerçekleştirilmiştir. Türler arası analiz için üç yönlü kümelemeye dayalı ilk yöntem olma özelliğine sahip olacak bu çalışma kanser araştırmaları ve ilaç tasarımı gibi biyotıp çalışmalarına katkı sağlayacaktır.

Proje en genel tanımıyla, biyolojik bazı sürecin anlaşılabilmesi için matematiksel modellerin tanımlanmasını ve buna bağlı olarak bilişimsel bazı araçların geliştirilmesini hedefleyen bir biyobilişim çalışmasıdır.

Tez raporu yedi bölümden oluşmaktadır. Birinci bölümde tezin motivasyonundan, katkılarından bahsedilmiş ve temel bilgiler verilmiştir. İkinci bölümde önceki çalışmalar hakkında bilgi verilirken, üçüncü bölümde yöntemler anlatılmıştır. Dördüncü bölümde geliştirilen araç detaylı olarak anlatılırken, beşinci bölümde kullanılan veri kümelerine değinilmiş, altıncı bölümde sonuçlar verilmiştir. Son bölümde ise tartışma ve gelecek çalışmalar hakkında bilgi verilerek rapor sonlandırılmıştır.

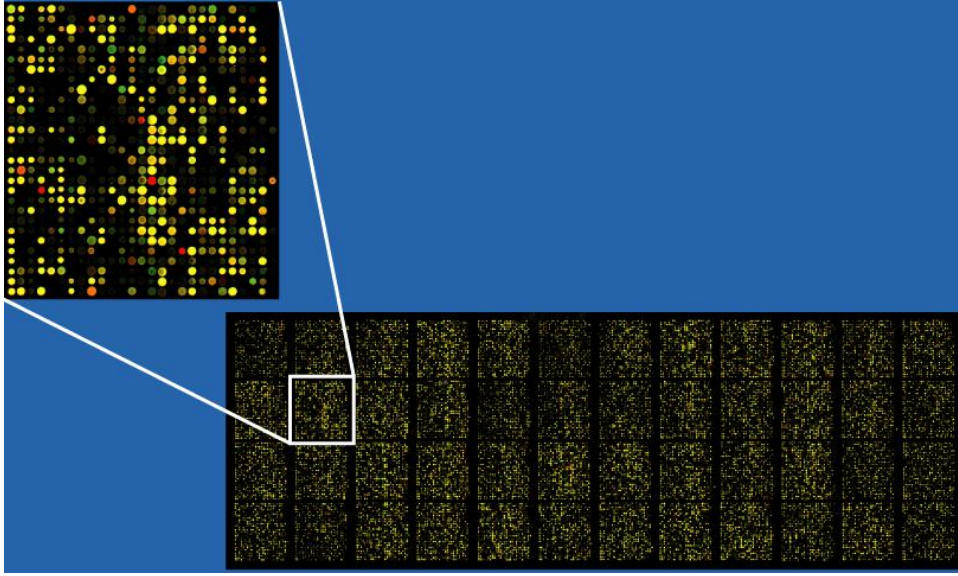
1.1 Gen İfadesi (Gene Expression)

Gen ifadesi ya da gen ekspresyonu, genlerin fonksiyonel gen ürünlerine dönüşmesi süreci için kullanılan bir terimdir. Bu ürünler genelde proteinler, ribozomal RNA (rRNA)'lar, taşıyıcı RNA (tRNA)'lar olmaktadır. Basitçe, bu durum genlerin veya karşılık gelen mRNA'ların açık (aktif) olup olmadıkları olarak da tanımlanabilir.

Gen ifadesi üretimi işlemi makro moleküler yaşam hücrelerinin üretilmesi için tüm canlılar (ökaryotlar, prokaryotlar vb.) tarafından yapılan bir işlemdir. Binlerce genin ifade değerleri mikroçip teknolojileri sayesinde paralel olarak ölçülebilmektedir. Mikroçip deneylerinin iki temel tipi vardır; cDNA mikroçipleri [1] ve oligonükleotit çipleri [2]. İki tipin deneysel protokol farklılıkları olsa da temelde üç aşamayı içermektedirler; Birinci aşama çip üretimi aşamasıdır. Mikroçip, küçük boyutta çiplerin (kimyasal kaplanmış cam, naylon zar veya silikondan yapılmış) katı bir yüzeyde birleştirilmesinden oluşturulmuş bir cihazdır. Katı yüzeyin her bir kare hücresi bir DNA dizilimine (sekans) eşittir. Hedef hazırlanması, etiketleme ve melezleme ikinci aşamadır, bu aşamada; iki mRNA örnekleme (test ve kontrol örnekleme) cDNA'lara (hedefler) ters kopyalanır, flüorışılmalı boya veya radyoaktif izotoplar kullanılarak etiketlenmesi yapılır ve sonrasında çipin yüzeyindeki problemler ile karmaştırılması (melezleştirme) yapılır. Tarama işlemi aşaması ise son aşamadır, bu aşamada; Çipler etiketlenmiş ve melezleştirilmiş hedeflerin sinyal yoğunluğunu okumak için taranır.

Genellikle, cDNA mikroçip ve oligo çip deneyleri test ve kontrol örneklemleri arasındaki sinyal yoğunluğu oranı ile her bir DNA diziliminin gen ifade seviyesini ölçmektedir, bu yüzden her iki yöntem sonucu elde edilen veri kümeleri biyolojik olarak aynı anlamı taşımaktadırlar [3].

Örnek bir mikroçip deney görünümü Şekil 1.1' de verilmiştir. Bu şekilde 40.000 örneklem için oluşturulmuş mikroçip ve her bir noktanın karşılık gelen genin ifade değerine göre renklendirilmesi görülmektedir.



Şekil 1. 1 Örnek mikroçip deneyi görünümü ¹

1.2 Mesajcı RNA (mRNA)

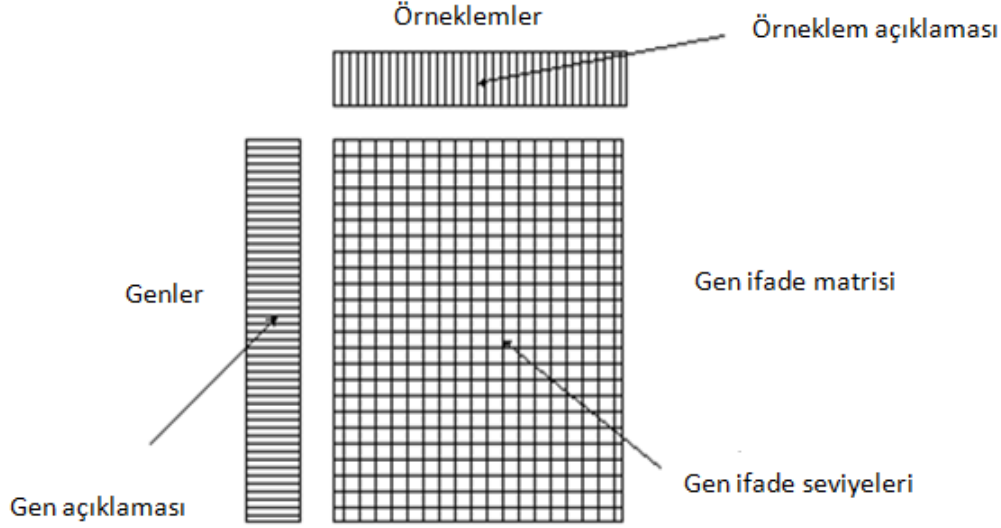
Mesajcı RNA, sentezlenecek bir proteinin amino asit dizisine karşılık gelen kimyasal şifreyi taşıyan bir moleküldür. mRNA, bir DNA kalıptan transkripsiyon yoluyla sentezlenir ve protein sentez yeri olan ribozomlara, protein kodlayıcı bilgiyi taşır. Burada, çevirim (translasyon) süreci sonucu, RNA polimerindeki bilgi ile bir amino asit polimeri üretilir. Nükleik asitlerin (mRNA ve DNA) amino asit dizilerine karşılık gelen bölgelerindeki her üç baz, proteindeki bir amino asite karşılık gelir. Bu üçlülere kodon denir, her biri bir amino asit kodlar, bitiş kodonu ise protein sentezini durdurur. Bu işlem iki diğer RNA türünü daha gerektirir: tRNA kodonun tanınmasına aracılık eder ve ona karşılık gelen amino asiti getirir; rRNA ise ribozomdaki protein imalat mekanizmasının kataliz merkezidir.

1.3 Yüksek Ölçekli Veri (High Throughput Data)

Aynı anda binlerce varlık değerinin (örneğin, gen ifadeleri) farklı örneklemeler üzerinden ölçülüp elde edildiği çok boyutlu matrislerde tutulduğu veri türüdür. Yüksek ölçekli veri örneği olan gen ifade matrisinde satırlar genleri, sütunlar ise durumları veya örneklemeleri temsil etmektedir (Şekil1.2). Örneklemeler farklı zaman noktalarına ve farklı deneylere karşılık geldiği gibi farklı organlardan, dokulardan, bireylerden alınan örnekleri de temsil edebilmektedir. Gen ifade matrisinin her bir

¹ <http://eng.thesaurus.rusnano.com/wiki/article813>

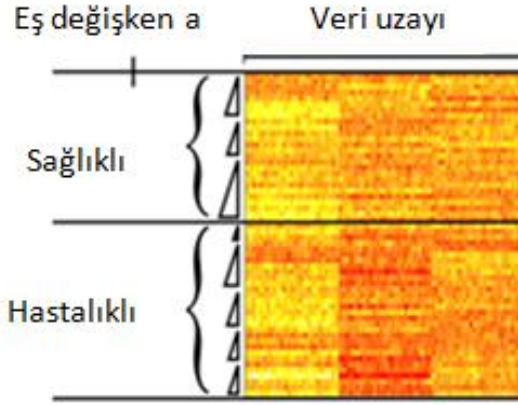
hücreyi bir genin belirli durumda veya örnekteki gen ifade seviyesinin (gene expression level) ölçüm sonucunun sayısal karşılığını içermektedir.



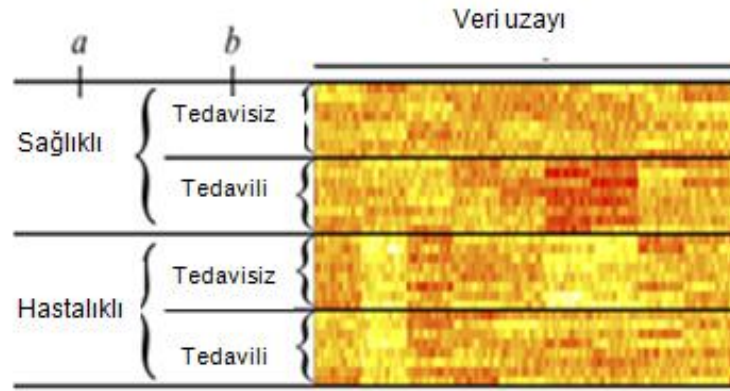
Şekil 1. 2 Gen ifade matrisi

1.4 Çok Yönlü Analiz (Multi-Way Analysis)

Veri analizi tek-yönlü (Şekil 1.3), çift-yönlü veya genelde çok-yönlü (Şekil 1.4) analizler olarak tanımlanabilir. Tasarımdaki faktörler, örneğin bir deneyde cinsiyet, yaş grubu, tip gibi, analizin kaç yönlü olacağını belirlemektedir. Tek yönlü analizde; örneğin bir farenin böbreğinden RNA örnekleri 3 farklı (düşük, orta, yüksek) zehir miktarını kapsıyor ise burada 3 seviyeli bir faktörden bahsedilir. Her bir seviye için birden fazla fareden örnekler alınarak analiz yapılır. Çift yönlü analizde iki faktör vardır. Eğer ilk faktör k_1 seviye içeriyor, ikinci faktör k_2 seviye içeriyorsa çift yönlü faktör yapısından $k_1 \cdot k_2$ kadar seviye kombinasyonları bulunmaktadır. Örneğin; ilk faktör iki fare türünü (mutant ve yabani) içeriyor ve diğer faktörde zehir miktarını (3 seviyeli) içermektedir. Çok yönlü analiz; ikiden fazla faktörün bulunduğu analiz türüdür. Fare örneği için cinsiyet, yaş grubu, fare tipi ve vb karakteristikler bu analizde ele alınabilecek çoklu faktörlerdir.



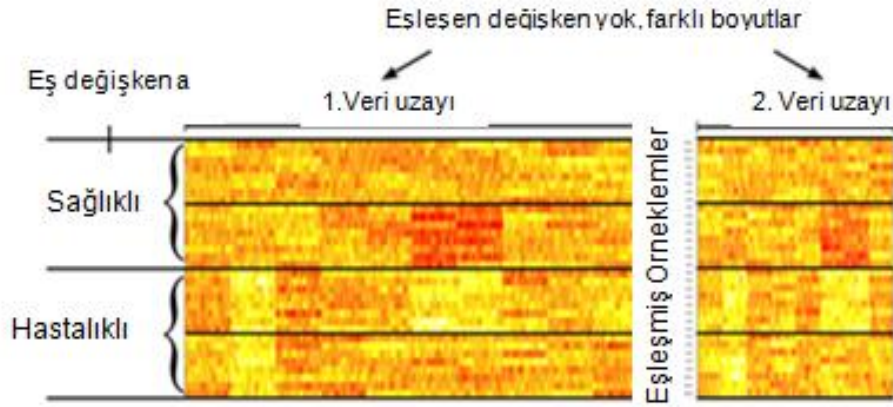
Şekil 1. 3 Tek-yönlü, Tek-kaynaklı analiz



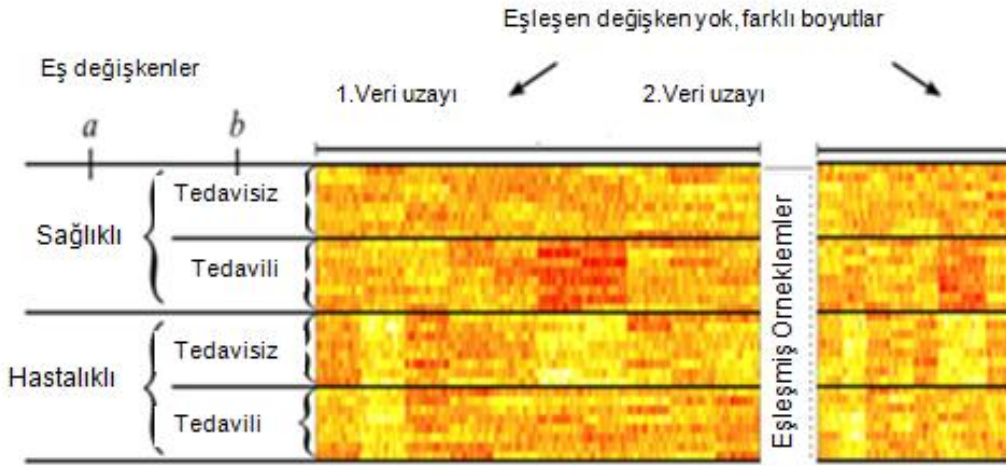
Şekil 1. 4 Çok-yönlü analiz

1.5 Çok Kaynaklı Analiz (Multi-Source Analysis)

Çok kaynaklı veri (Şekil 1.5); birden fazla kaynaktan alınan karmaşık, heterojen, dinamik, dağınık ve yüksek ölçekli veridir. Çok kaynaklı verinin analizi, tasarımı ve diğer kullanıcıların kullanımına hazır hale getirilmesi kaynakların farklı karakteristikleri nedeniyle, tek kaynaklı veri analiz işlemlerinden daha zordur. Çok kaynaklı ve çok yönlü analizin birleşimi Şekil 1.6' da verilmiştir.



Şekil 1. 5 Çok-kaynaklı veri



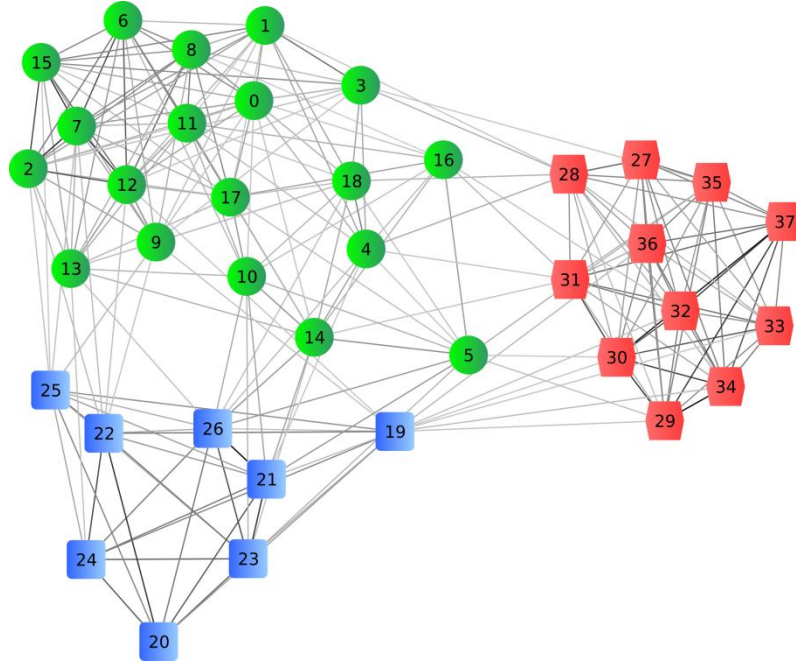
Şekil 1. 6 Çok yönlü-Çok kaynaklı analiz birleşimi

1.6 Kümeleme (Clustering) Yöntemi

Kümeleme yöntemi benzer objeleri bir grupta toplamak için kullanılan bir güdümsüz öğrenme (unsupervised learning) yöntemidir. Veri madenciliği araştırmasında temel konulardan biri olan kümeleme yöntemi, istatistiksel veri analizi, makine öğrenimi, örüntü tanıma, imge analizi, biyobilişim gibi çalışma alanlarında sıklıkla kullanılan bir yöntemdir.

Kümeleme analizinde belirli bir algoritma yoktur, çözülecek probleme göre farklı algoritmalar, çözüm yöntemleri mevcuttur. Kümelemede temel amaç, kümeleme yapılacak veri kümesinden bulunacak olan kümelerin birbirine en yakın (benzer olmaları), diğer kümelere uzak mesafede olmalarını sağlamaktır. Şekil 1.7'de bir

veri kümesinin kümeleme sonrası oluşan üç kümesinin farklı renkler ile görünümü verilmiştir.



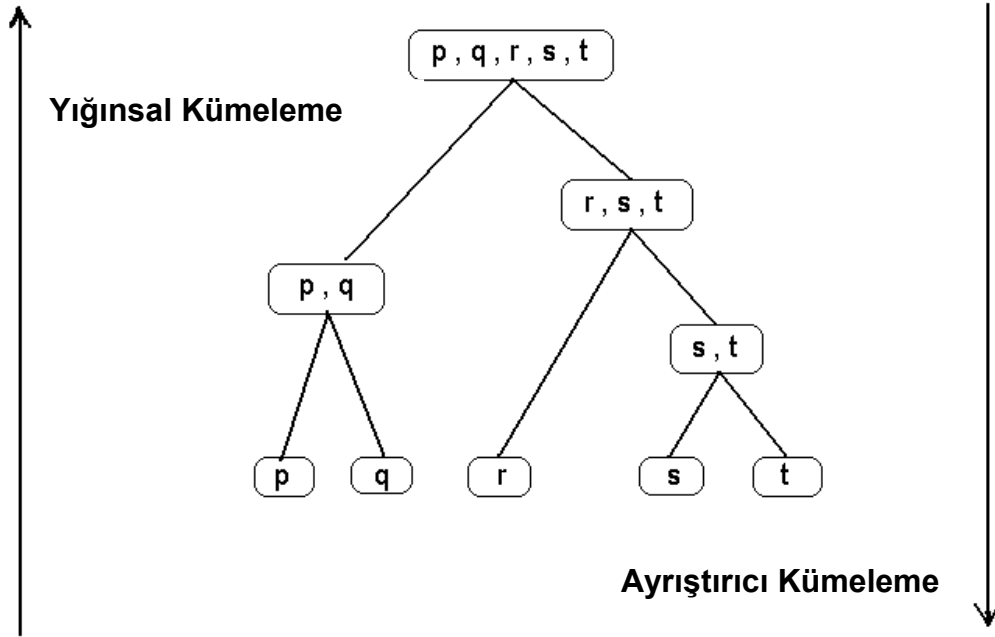
Şekil 1. 7 Kümeleme sonucu oluşan üç farklı kümenin farklı renkler ile gösterimi²

Kümeleme biyobilişim alanında gen ifade verilerinin analizinde en sık kullanılan yöntemlerden birisidir. Kümelemenin temel mantığı; benzer ifade örüntülerine sahip genleri (aynı transkripsiyon faktör (TF) tarafından aktive edilen, aynı biyolojik işlemlerde yer alan genler) kümelemek, ilişkilendirmektir. En sık kullanılan kümeleme yöntemlerinde (k-means, hierachical vb.) her bir gen sadece bir kümeye konulmakta, tüm genlerin tüm durumlarda benzer davranış gösterdiği yaklaşımı esas alınmaktadır. En sık kullanılan kümeleme yöntemleri hiyerarşik ve k-ortalamlar (k-means) kümeleme yöntemleridir.

Hiyerarşik kümeleme (Hierarchical clustering) yöntemi bağlantı temelli (connectivity-based) bir yöntem olup, birbirine yakın objeler uzak olanlara göre daha ilişkili olanlardır mantığı ile kümeleme yapmaktadır. Ayrıca objeler arası

² <http://espin086.wordpress.com/2011/02/27/clustering-to-improve-merchandise-allocation-testing-and-forecasting-an-application-of-the-k-medians-algorithm/>

uzaklığı ve hiyerarşiyi göstermek için dendogram yapısı kullanılmaktadır. Dendogram yapısında y eksenini gruplar arası uzaklığı gösterirken, x eksenini ise objelerin gruplara dağılışını göstermektedir. Hiyerarşik kümeleme yığınsal (agglomerative) ve ayırıcı (divisive) kümeleme olmak üzere iki temel gruba ayrılmaktadır. Yığınsal kümeleme aşağıdan yukarı doğru işlem yapan bir algoritmadır; her obje kendi grubu ile başlayıp grup çiftleri birleşerek büyüyerek hiyerarşinin en üst seviyesini oluşturmaktadır. Ayırıcı kümeleme ise yukarıdan aşağı doğru çalışan bir algoritmadır; tüm objeler tek bir grupta toplanmış hiyerarşinin en alt seviyesine gidene kadar farklı gruplara bölünmektedirler [4]. Her iki hiyerarşik kümeleme yönteminin dendogram gösterimi Şekil 1.8’ de verilmiştir.



Şekil 1. 8 Yığınsal ve Ayrıştırıcı Kümeleme yöntemlerinin dendogram gösterimi³

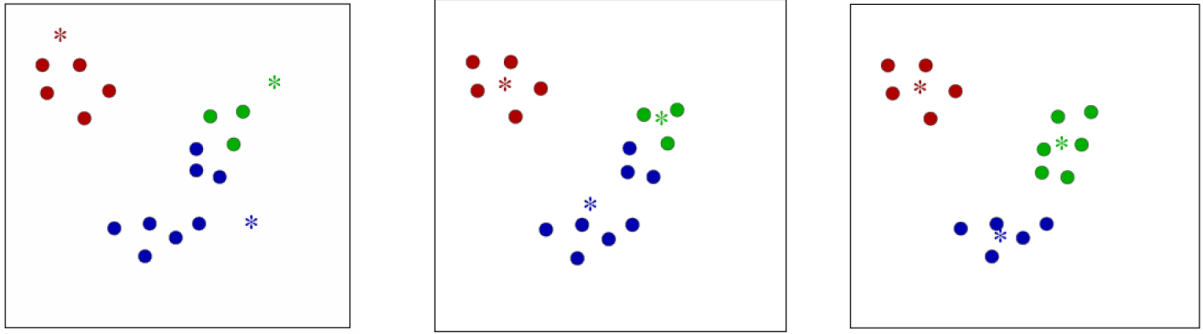
En sık kullanılan kümeleme yöntemlerinden bir diğeri ise merkez tabanlı (centroid-based) bir yöntem olan k-ortalamlar kümeleme yöntemidir. Bu yöntemde gruplar merkez olarak seçilen bir vektöre (veri kümesinin elemanı olmak zorunda değil) göre oluşturulmaktadır. Burada amaç veri kümesindeki elemanları en yakın

³ <http://www.solver.com/hierarchical-clustering-intro>

oldukları merkezin grubuna dahil ederek sonuçta k adet küme elde etmektir. Algoritma temel olarak 4 aşamadan oluşmaktadır;

1. Sınıf merkezlerinin belirlenmesi,
2. Merkez dışındaki örneklerin mesafelerine göre sınıflandırılması,
3. Yapılan sınıflandırmaya göre yeni merkezlerin belirlenmesi (veya eski merkezlerin yeni merkeze kaydırılması),
4. Kararlı hale (stable state) gelinene kadar 2. ve 3. adımların tekrarlanması [5].

Temel üç aşama Şekil 1.9' da gösterilmiştir. Örnekte k-ortalamlar kümeleme yöntemi sonucunda üç farklı küme elde edilmiştir.



Şekil 1. 9 k-ortalamlar kümeleme örneği⁴

1.7 İkili Kümeleme-Çift-Yönlü Kümeleme (Biclustering) Yöntemi

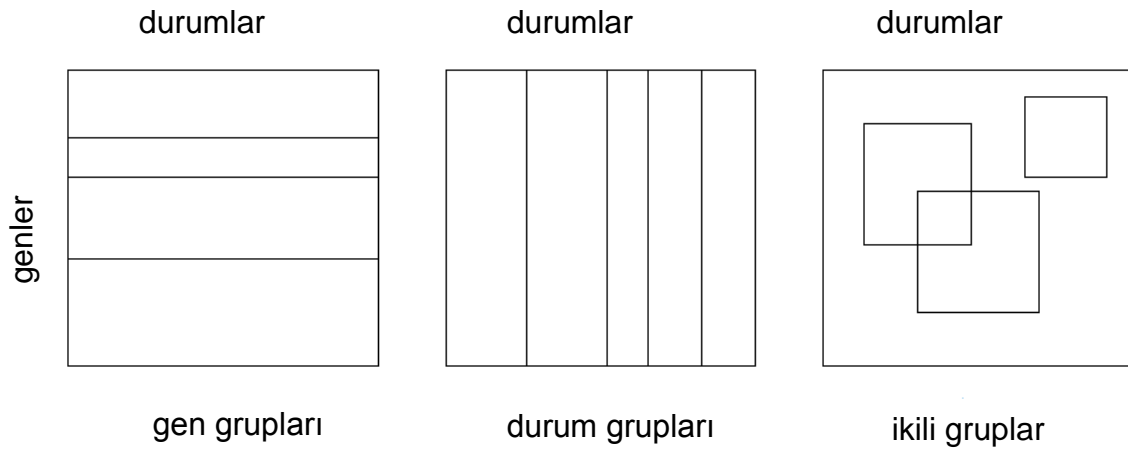
Gen ifade verisi analizinde önemli olan temel amaçlar vardır, bunlar; belirli durumlara göre gen ifade değerlerini kullanarak genleri, belirli genlere bağlı olarak durumları kümelemektir. Kümeleme yöntemleri tek yönlü kümeleme yaptığı için (sadece genleri veya sadece durumları kümeleme) gen ifade veri analizinde belirtilen amaçlara karşılık verememektedir.

Kümeleme yönteminin gen ifade veri analizindeki yetersizliği sebebiyle ikili kümeleme (biclustering) yöntemi geliştirilmiştir. Co-clustering, two-mode clustering olarak da bilinen ikili kümeleme yöntemi, diğer kümeleme yöntemlerinden türetilmiş bir kümeleme yöntemidir. Kümeleme yöntemi ile bir matrisin sadece satır

⁴ <http://www.cs.uvm.edu/~xwu/kdd/Slides/Kmeans-ICDM06.pdf>

veya sütunları kümelendirilirken, ikili kümeleme yöntemi ile satır ve sütunlar eş zamanlı kümelendirilmektedir. İlk olarak Cheng ve Church [6] araştırmacıları tarafından tanımlandığı bilinse de bu yöntemi ilk kullanan J.A Hartigan'dır [7]. İkili kümeleme yöntemi ile küçük matris grupları üretilmekte ve gruplarda hangi gen kümesinin hangi durum kümesinde veya hangi durum kümesinin hangi gen grubunda aktif olduğu bilgisi elde edilebilmektedir. Şekil 1.10'dan da görüleceği üzere ikili kümeleme yöntemleri gen ifade matrisi üzerinde hem genleri hem durumları içeren alt matrisler bulmaktadır. İkili kümeleme yöntemi kümeleme yönteminden farklı olarak veri matrisinin iki boyutunda kümeleme yapan bir yöntemdir. Diğer bir deyişle kümeleme küresel bir model olarak, ikili kümeleme ise yerel bir model olarak işlem yapmaktadır [8].

Sonuç olarak kümeleme algoritmalarından farklı olarak ikili kümeleme algoritmaları, belirli durum veya koşulda benzer davranışı gösteren gen gruplarını veya belirli gen grubuna bağlı olarak aynı grupta yer alan durumları tanımlayabilen algoritmalarlardır. Bu yüzden ikili kümeleme algoritmalarını gen ifade veri analizinde kullanmak daha elverişli sonuçlar vermektedir.



Şekil 1. 10 İkili kümeleme örneği [9]

1.8 GO (Gene Ontology) Analizi

Biyologlar bilimsel araştırmalarının her bir adımında uygun veriye ulaşmak için çok zaman harcamaktadırlar. Farklı veritabanlarında, çeşitli terminolojilerin kullanılıyor olması bilgiye erişim konusunda çeşitli sıkıntılar yaratmaktadır. Örneğin; eğer farklı sekanslara (gen dizilimine) sahip, bakteriyel protein sentezlerinde yer alan

antibiyotikler için yeni hedefler (target) araştırılıyorsa bu araştırmanın farklı veri tabanlarının kullanılarak yapılması aramayı zorlaştırmaktadır. Çünkü bir veritabanı aranan molekülleri “çevirim” (translation), diğer veritabanı ise “protein sentezleme” (protein synthesis) görevinde yer alıyor olarak tanımlayabilmektedir. Bu durum araştırmacıların sonuca ulaşmalarında zorluklar yaratmaktadır. GO projesi bahsedilen sıkıntıları çözmek için geliştirilmiş, gen ve gen ürünlerinin davranışlarını, türleri temel alarak standartlaşmak ve temsil etmek için geliştirilmiş bir biyobilişim projesidir [10].

Bu projede genlerin karakteristiklerini tanımlamak için kontrollü bir terimler sözlüğü (ontolojiler) ve gen açıklama (gene annotation) verisi geliştirilmiştir. Bir başka deyişle; GO, türden bağımsız olarak tasarlanmış ve prokaryotları (bakteriler, mavi yeşil alg vb. canlılar), ökaryotları (hücreli organizmalar), tek ve çok hücreli organizmaları içeren terimlere sahip bir sözlüktür.

Gen açıklaması (gene annotation), GO terimlerinin gen ürünlerine atanması işlemidir. Şekil 1.11 'de gösterildiği üzere türe bağlı oluşturulan açıklama dosyalarında terim isimleri, GO ID'leri, tanımları, ontoloji türleri vb. bilgiler yer almaktadır [10].

id: GO:0006094	unique GO ID
name: gluconeogenesis	term name
namespace: process	ontology
def: The formation of glucose from noncarbohydrate precursors, such as pyruvate, amino acids and glycerol.	definition
[http://cancerweb.ncl.ac.uk/omd/index.html]	
exact_synonym: glucose biosynthesis	synonym
xref_analog: MetaCyc:GLUCONEO-PWY	database ref
is_a: GO:0006006	
is_a: GO:0006092	parentage

Şekil 1. 11 GO açıklaması dosya örneği [11]

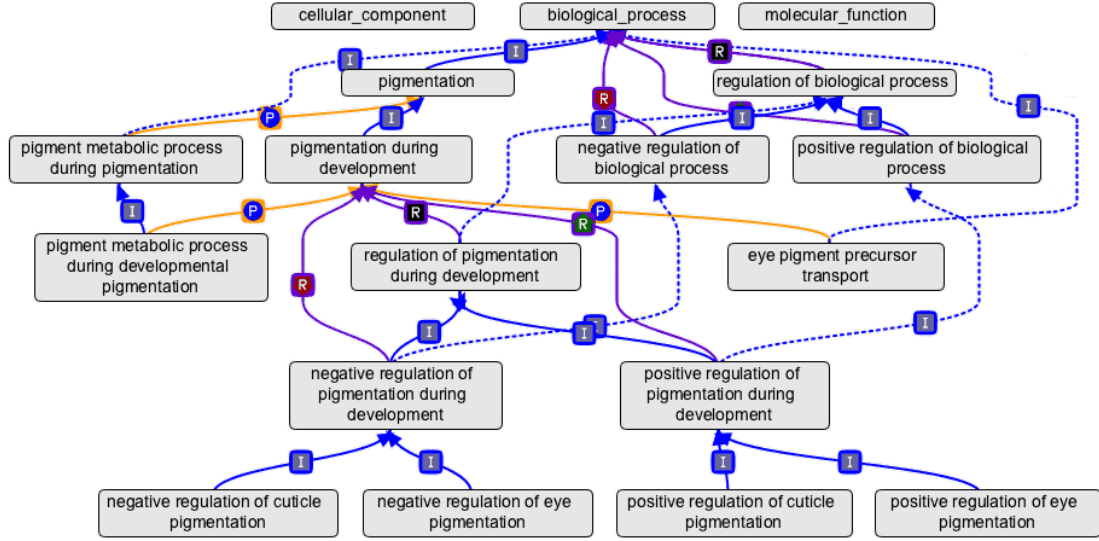
Projede gen ürünlerinin türden bağımsız bir şekilde biyolojik işlemlerine (biological process), hücresel bileşenlerine (cellular component) ve moleküler fonksiyonlarına (molecular function) göre tanımlayan üç farklı ontoloji geliştirilmiştir. Hücresel bileşenler, bir hücrenin kısımlarını veya hücrenin dış ortamını, moleküler fonksiyonlar bir genin moleküler seviyedeki ana aktivitelerini (bağlama, kataliz vb.),

biyolojik işlemler ise başlangıcı ve sonu tanımlı olan moleküler olayların işlemlerini hücreler, dokular, organlar ve organizmalar gibi canlı yapılara göre tanımlanmasını kapsamaktadır.

GO ontolojisi yönlü, çevrimsel olmayan bir çizge (directed acyclic graph) olarak yapılandırılmış; her bir terim aynı ve bazen de farklı alanlardaki diğer terimlerle ilişkilendirilmiştir. Şekil 1.12'de GO yapısının diyagram olarak görünümü verilmiştir. Bu şekle göre her bir terim bir düğüm olarak, her bir ilişki düğümler arası bağlantı olarak tanımlanmıştır. GO yapısındaki ilişkiler yönsüz ilişkilerdir; örneğin mitokondri bir organeldir, fakat bir organelin mitokondri olmaması ilişkilerin yönsüz olduğunu göstermektedir. Aynı zamanda ilişkilerin çevrimsel olmayışı çizgede döngülerin olmadığını göstermektedir.

Ontolojiler çocuk terimlerin daha detaylı gösterildiği, ana terimlerin çok detaylandırılmadığı, bir terimin birden fazla ana teriminin olduğu hiyerarşik yapılardır. Ontolojiler üç farklı seviyede yapılandırılmıştır; örneğin; fare genomunda "signal transduction" işleminde yer alan tüm gen ürünlerini aramak için veya tüm "tyrosine kinases"(bir enzim türü) alıcılarını yakından incelemek için GO kullanılabilir.

GO açıklamaları yüksek ölçekli gen verilerinin analizlerinde sıklıkla kullanılan önemli bir araç haline gelmiştir. GO terimleri zenginleştirme (GO term enrichment) analizi için geliştirilmiş çeşitli yazılım araçları ve paketleri bulunmaktadır. Bu araçlar ve paketler gen ifade verisi analizi sonucu elde edilen gen gruplarının biyolojik olarak anlamlı olup olmadıklarını tespit etmek için kullanılmaktadırlar.



Şekil 1. 12 GO yapısının diyagram gösterimi

Yapılan analizin başarılı olup olmadığı kullanılan araç veya paketlerde kullanıcıya çıktı olarak sunulmaktadır. Kümeleme tekniklerinde sonuç analizinde başarılı olma kriterleri, aynı kümede yer alan genlerin benzer gen ifade verisine sahip olmaları, benzer davranış göstermeleri (up-down regulation) veya benzer hücresel aktivitelerde yer almalarıdır. Bu başarının sayısal doğrulaması GO terimlerinin istatistiksel analizi ile yapılmaktadır. Analiz şu şekilde yapılmaktadır; deney sonucu elde edilen gen gruplarında Hipergeometrik testi kullanılarak, zenginleşen GO terimlerinin olma olasılığı bulunmaktadır. Hipergeometrik dağılımı, olasılık kuramı ve istatistik bilim kollarında, hipergeometrik dağılım sonlu bir ana kütle içinden tekrar geri koymadan seri halinde, n tane nesnenin çekilmesi şeklinde bir işlem için başarı sayısının dağılımını bir ayrık olasılık dağılımı şeklinde betimlemektedir. Tüm GO terimleri için ayrı ayrı Hipergeometrik dağılım testi uygulanmakta ve Bonferroni doğrulanmış p değeri (Bonferroni corrected p-values) elde edilmektedir. Elde edilen ortalama p değeri ne kadar küçük olursa analizi yapılan gen gruplarında zenginleşmiş GO terimi sayısı fazladır demektir. Bu da birbirleriyle ilişkili olan gen gruplarının elde edildiğini yani analizin başarılı olduğunu göstermektedir. Diğer bir deyişle p değeri analiz sonucunun istatistiksel olarak anlamlı olup olmadığını göstermektedir.

2. ÖNCEKİ ÇALIŞMALAR

Bu bölümde gen ifade veri analizi, kümeleme, ikili kümeleme ve çok yönlü kümeleme yöntemleri ile ilgili daha önce yapılan çalışmalar ve geliştirilen yazılım araçları hakkında bilgi verilecektir.

2.1 Gen İfade Veri Analizi

Gen ifade veri matrisleri, aynı anda birçok örneklemden elde edilen binlerce genin ifade verisini içeren çok boyutlu yüksek ölçekli verilerdir. Mikroçip teknolojileri sayesinde yüksek ölçekli gen ifade verilerinin analizleri yapılabilmektedir. Bu tür verilerin analizi için yeni ve etkili yöntemlere ihtiyaç duyulmaktadır [12].

Gen ifade verisinin analizinde çeşitli öğretici ile öğrenme (supervised) ve eğitici (unsupervised) öğrenme teknikleri kullanılmaktadır. Bu teknikler de kendi içlerinde farklı yöntem kategorilerini barındırmaktadır. Analiz edilecek gen ifade verisine uygun olan yöntemin seçilmesi, verinin ön işleme (preprocessing) işlemlerinin yapılması, elde edilen sonuçların biyolojik ve istatistiksel doğrulamasının yapılması gen ifade veri analizinin önemli adımlarıdır.

2.2 Kümeleme Yöntemi

Gen ifade veri analizinde sıklıkla kullanılan eş güdümsüz öğrenme tekniğinden birisi kümeleme yöntemidir. Kümeleme yöntemleri sayesinde benzer ifade düzenlerine (expression pattern) sahip, benzer biyolojik görevlerde yer alan genler gruplanmaktadır. En sık kullanılan kümeleme yöntemlerinden birisi k-ortalamlar (k-means) kümeleme yöntemidir. Bu yöntemde herhangi bir sınıf bilgisi olmadan veriler k sayıdaki kümeye gruplanmaktadır. Kümeleme işlemi, ilgili kümenin merkez değeri ile veri kümesindeki her objenin arasındaki farkın kareleri toplamının minimumu bulunarak yapılmaktadır. Bir diğer kümeleme yöntemi nesnelerin iç içe kümeleme ilişkisini dendogram yapısı ile gösteren Hiyerarşik (Hierarchical) kümeleme yöntemidir. Eisen vd. [13] çalışmalarında bir birleştirici (agglomerative) kümeleme algoritması geliştirmişler ve gen ifade veri matrisine uygulamışlardır. Bu yöntemde gen ifade veri matrisinin her bir hücresi flüorışıma oranlamasına (fluorence ratio) göre renklendirilmiş ve matris satırları hiyerarşik dendogram yapısına göre sıralandırılarak kümeleme işlemi yapılmıştır.

2.3 İkili Kümeleme (Biclustering) Yöntemi

Çift yönlü kümeleme yöntemleri, biyoloji ve biyobilişim alanlarında yüksek ölçekli verilerden anlamsal bilgi çıkarımında ve gen ifade matrislerinin analizlerinde en sık kullanılan yöntemlerden birisidir. 1996 yılında Mirkin [14] tarafından keşfedilen bu matematiksel yöntemin, ilk kez gen ifade verileri üzerinde uygulanması 1996 yılında Cheng ve Church (CC) tarafından yapılmıştır.

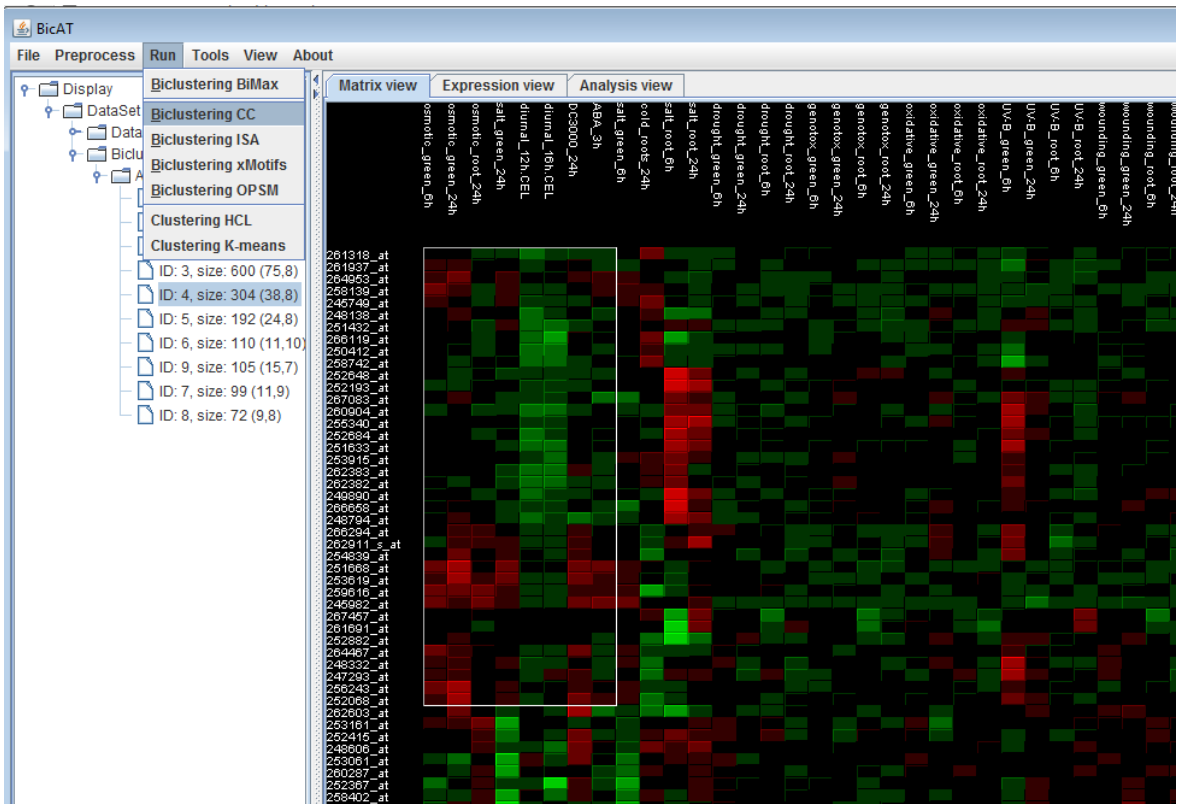
Çift yönlü kümeleme yöntemi, gen ifade verilerinin matrisi üzerinde uygulandığında kümeleme yöntemlerinden farklı olarak gen ve durumları eş zamanlı olarak kümelemekte, alt matrisler bularak hangi genlerin hangi durumlarda aktif olduğu veya hangi durumlarda hangi genlerin benzer davranış gösterdiği bilgisini sağlamaktadır. Kümeleme yöntemleri veri matrisinin satır ve sütunlarına ayrı ayrı uygulanırken, çift yönlü kümeleme yöntemleri bu iki boyuta eş zamanlı olarak uygulanabilmektedir. Kümeleme algoritmaları uygulandığında bir gen grubundaki genin tüm durumlara bağlı olarak tanımlaması yapılır, aynı şekilde bir durum grubundaki durumun da tanımlaması tüm genlere bağlı kalınarak yapılmaktadır. Bu yaklaşım doğru bir yaklaşım değildir; örneğin bir gen grubundaki gen tüm durumlar yerine, ilişkili olduğu durumlara (alt gruplara) bağlı olarak tanımlanmalıdır. Bunun yanında bir gen/durum birden fazla grupta yer alabilmektedir. Kümeleme yöntemlerindeki bu eksiklikler geliştirilen çift yönlü kümeleme yöntemlerinde yoktur.

Çift yönlü kümeleme yönteminin gen verilerinin analizinde kullanılması ile birlikte birçok çift yönlü kümeleme yöntemi geliştirilmiştir [6; 15; 16; 17; 18; 19]. Cheng ve Church geliştirdikleri ilk çift yönlü kümeleme yönteminde, matrislerin satır ve sütunlarını orta kare değerine (mean squared residue-MSR) yani temel benzerlik bulma amacına göre eş zamanlı olarak kümelemektedir. Algoritma birbirine benzer genlerin (coexpressed genes) olduğu büyük boyutlu ikili grupları bulmayı amaçlamaktadır. Cheng ve Church'un geliştirmiş olduğu bu yöntem birçok çalışmanın temelini oluşturmuştur. Bunlardan birisi Gremalschi ve Altun'un [16] geliştirmiş olduğu; MSR değerine üst sınır koyularak, CC algoritmasının performansını iyileştirmek amacıyla yapılan çalışmadır. Gremalschi ve Altun'un [17] bir diğer çalışmalarında ise önceki çalışmalarında geliştirdikleri algoritmaya

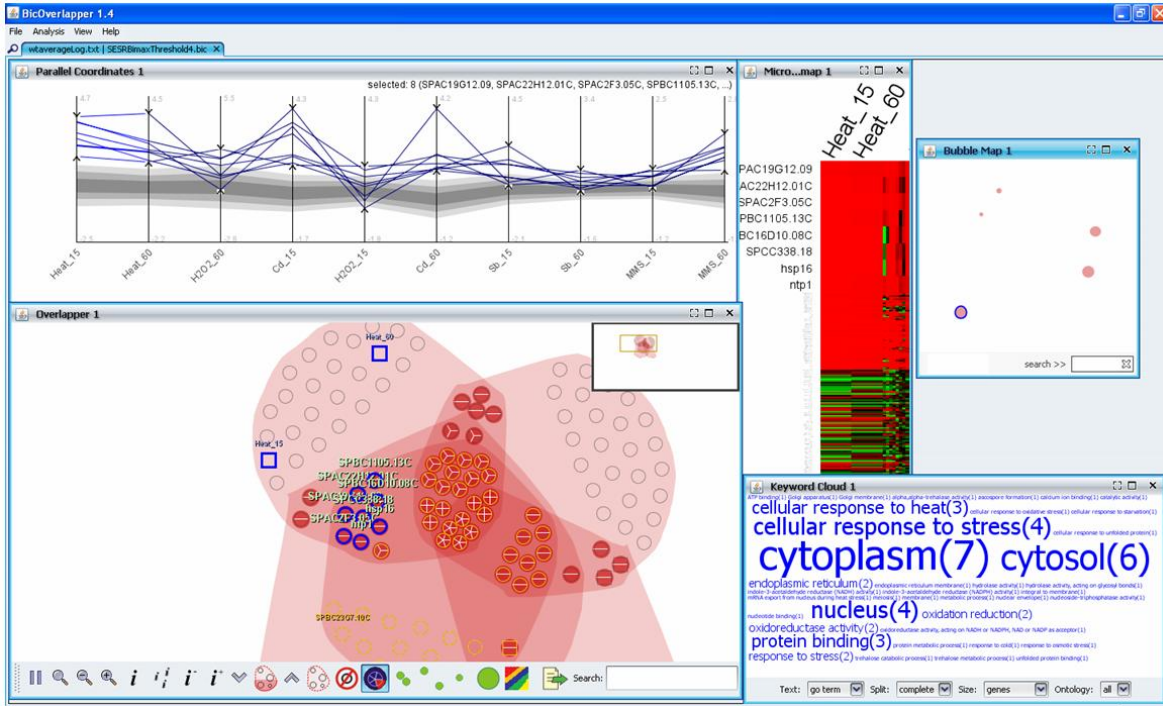
bazı yeni özellikler ekleyerek CC algoritmasından daha performanslı bir algoritma geliştirmişlerdir. Diğer çalışmalardan farklı olarak satır işlemlerinin tutulduğu satır durumu değişkeni eklenerek ve kayıp değerli matris hücrelerine rastgele değer atanması yerine bu değerler yok sayılarak işlemler yapılmıştır. Çift yönlü kümeleme yöntemlerinden bir diğeri Plaid model yöntemidir [20]. Bu yöntemde veri matrisindeki her bir eleman katman olarak adlandırılan yapıların toplamı olarak ifade edilmektedir. Veri matrisi ise karşılık geldikleri ikili kümenin katmanlarının lineer bir fonksiyonu olarak tanımlanmaktadır. Bu yöntemde birden fazla ikili küme arasında etkileşim tanımlanarak ve belirlenen metrik minimize edilerek istenilen sayıda ikili kümelerin bulunması hedeflenmektedir. Bunun yanında CTWC (Couple Two-Way Clustering) [21] yöntemi de gen ifade verisini analiz etmede sıklıkla kullanılan bir yöntemdir. Bu yöntemde matrisin satır ve sütunlarına tek-yönlü kümeleme yapılmakta ve sütun kümelemesinde satırların sabit gruplarını, satır kümelemesinde sütunların sabit gruplarını özellik olarak alarak işlem yapılmaktadır. Bu işlemler yapılırken kullanılacak tek-yönlü kümeleme yönteminin seçimi ve sabit grupların tanımı belirli bir kurala bağlı kalınmadan yapılmaktadır. Bu alanda yapılan bir diğer çalışmada ise hiyerarşik generatif bir çift yönlü kümeleme yöntemi geliştirilmiş ve bir miRNA veri kümesi üzerinde uygulanmıştır [22].

Yüksek ölçekli veri kümelerinin analizi için geliştirilmiş çeşitli yazılım araçları mevcuttur. Bu yazılım araçları ile büyük ölçekli gen ifade verisi üzerinde farklı tek yönlü, çift yönlü kümeleme yöntemleri uygulanabilmekte, sonuçların biyolojik ve istatistiksel yorumlanması yapılabilmektedir. Bunlardan birisi Şekil 2.1 de ana sayfası görülen BicAT (Biclustering Analysis Toolbox) yazılım aracıdır. BicAT ile farklı kümeleme ve ikili kümeleme yöntemleri (CC, ISA, xMotifs, OPSM, HCL, K-means) yüksek ölçekli gen ifade matrislerine uygulanabilmekte, elde edilen kümeleme sonuçlarının matris, ısı haritası ve grafiksel gösterimleri yapılabilmektedir [23]. Bir diğer yazılım aracı ise Şekil 2.2'de ana sayfası görülen, gen ifade veri matrislerinden farklı yöntemleri kullanarak ikili gruplar elde eden, kullanılan yöntemlerinin sonuçlarının kıyaslanmasına olanak sağlayan BicOverlapper aracıdır [24]. Bu araç ile aynı zamanda sonuçların biyolojik ve istatistiksel analizi de yapılabilmektedir. Şekil 2.3' de anasayfası görülen STEM (Short Time-series Expression Miner) ise zaman serili gen ifade matrislerinin (time series gene

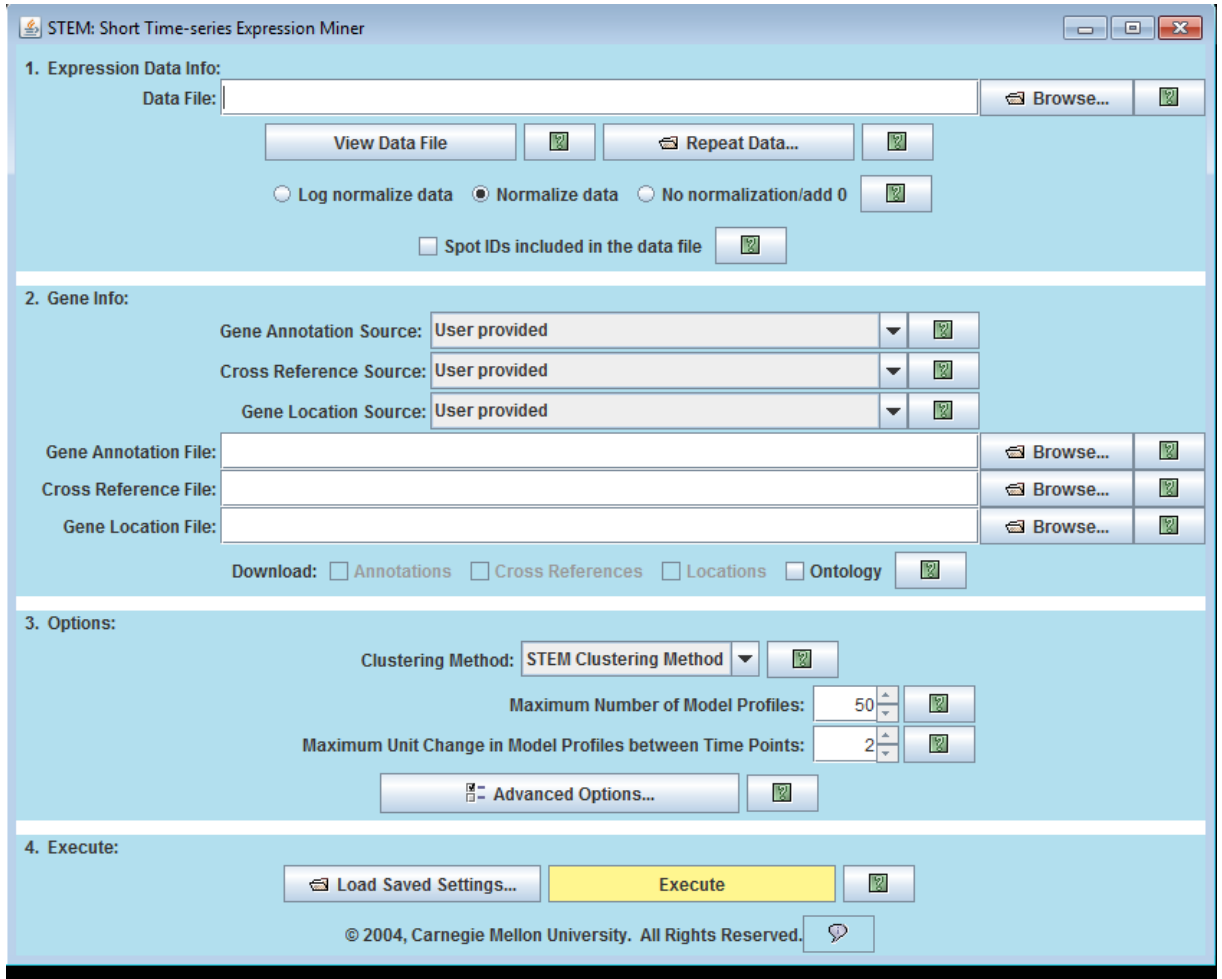
expression matrix) analizi için kullanılan bir yazılım aracıdır [25]. Bu araçta kümeleme yöntemleri uygulanabilmekte, kıyaslanabilmekte ve verinin görselleştirilmesi yapılabilmektedir. Bu araçlara ek olarak, Şekil 2.4'de ana sayfası görülen Cytoscape [26] yazılım aracının bir eklentisi olan BiNGO [27] ise gen gruplarının biyolojik ve istatistiksel analizini yapmada kullanılan bir araçtır. Bu araç ile çeşitli istatistiksel testlerin, referans kümelerinin veya ontoloji dosyalarının (genlerin görev aldığı biyolojik işlemlerin tanımlarının olduğu dosyalar) seçimi yapılarak analiz edilecek gen grubunun anlamlı olup olmadığı test edilebilmektedir.



Şekil 2. 1 BicAT yazılım aracı



Şekil 2. 2 BicOverlapper yazılım aracı

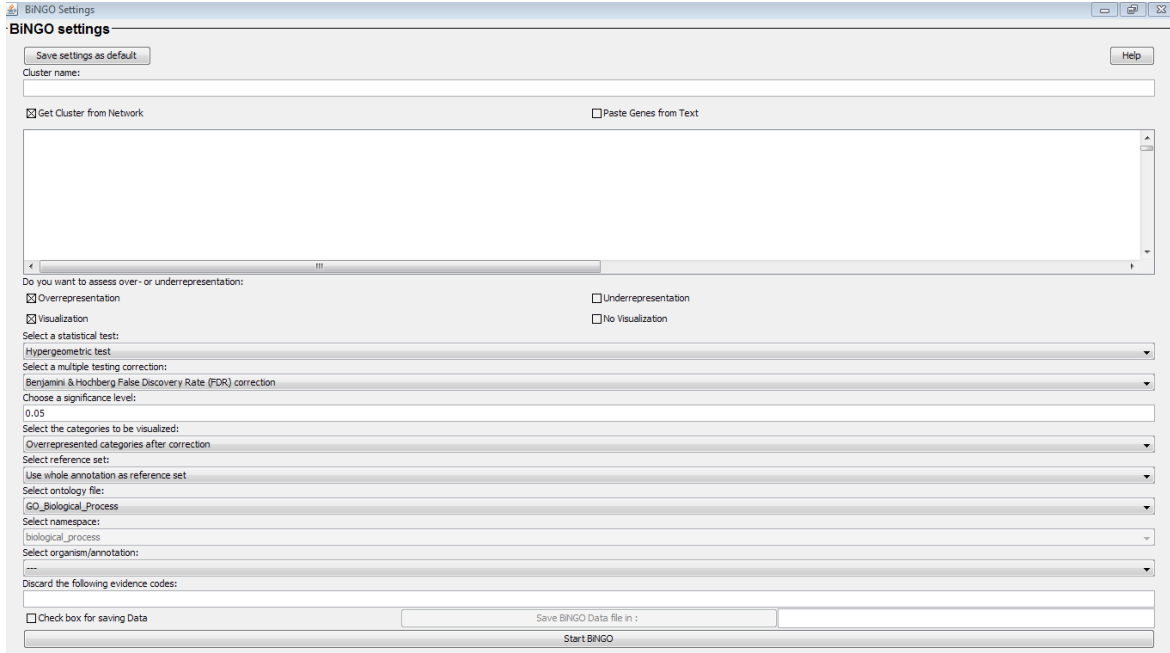


Şekil 2. 3 STEM yazılım aracı

2.4 Çok-Yönlü Kümeleme Yöntemleri

Çift yönlü kümeleme yöntemleri, biyoloji ve biyobilişim alanlarında yüksek ölçekli verilerden anlamsal bilgi çıkarımında sıklıkla kullanılan yöntemler olmasına rağmen iki boyutlu diziler veya matrisler verinin tüm içeriğini kapsama konusunda yetersiz kalmakta, standart iki-boyut analiz teknikleri çok kipli (multimodal) veri setlerindeki temel yapıları bulmada başarısız olmaktadır. 2009 yılında Acar ve Yener [28] tarafından yapılan bir çalışmada çeşitli çok yönlü (multi-way) veri analiz yöntemleri tanıtılmış, karşılaştırmaları yapılmış ve uygulama alanlarından örnekler verilmiştir. Bu tür, verinin birden fazla biçiminin olduğu veri setlerini analiz etmek için çok yönlü veri analiz teknikleri yaygın bir şekilde kullanılmaktadır. Çok kaynaklı, çok değişkenli biyolojik veri analizi için, Huopaniemi ve ekibi [29] tarafından yapılan bir çalışmada birleşik bir Bayes makine öğrenim yöntemi

geliştirilerek tek bir modelde çok yönlü ve çok kaynaklı verinin analizi hedeflenmiştir.



Şekil 2. 4 BiNGO yazılım aracı

Bunun yanında 2007 yılında Vermut'un [30] üç-yönlü veri kümesinin analizi için geliştirmiş olduğu hiyerarşik model ve iki deneysel veri üzerinde bu modeli uygulaması yapılan çalışmalara örnek olarak gösterilebilir. Üç yönlü kümeleme için geliştirilmiş olan bir diğer yöntem ise üç boyutlu zaman serili gen ifade matrisini kullanarak, gen-örneklem-zaman (gene-sample-time) boyutlarında gruplar bulan TRICLUSTER yöntemidir [31]. Tchaganag vd. de [32] Order Preserving Triclustering adında bir algoritma geliştirerek, üç boyutlu zaman serisi verisinin analizini yapmışlardır.

Üç yönlü kümeleme ile ilgili çeşitli çalışmalar olsa da [30; 31; 32], türler arası analize olanak sağlayacak bir kümeleme yöntemi ve yazılım aracı henüz geliştirilmemiştir. Türler arası benzerliklerin, farklılıkların ortaya çıkmasına, tıbbi çalışmalarda kullanabilecek analizlerin yapılmasına olanak sağlayacak üç yönlü bir kümeleme yöntemi ve yazılım aracına ihtiyaç vardır. Bu tezde belirtilen ihtiyacı kapatacak bir araç geliştirilmeye çalışılmıştır. Geliştirilen araç, gen-durum-tür (gene-condition-organism) boyutlarında analiz yapmayı hedefleyen yeni bir üç yönlü kümeleme yöntemi olan TriClustering (Three-way clustering) yöntemini

içermektedir. Geliştirilen araç ile aynı zamanda elde edilen üçlü kümelerin biyolojik ve istatistiksel analizi yapılabilmektedir.

2.5 Türler Arası Analiz

Biyolojik sistemler çeşitli organizmalarda, türlerde benzer davranış göstermektedirler ve bu sistemlerde görev yapan genler türler arasında korunmaktadır [33]. Sekans (gen dizilimi) benzerliği yeni genlerin tespitinde kullanılan önemli kaynaklardan biridir [34]. Aynı zamanda sekans korunumu (sequence conservation) analizi yeni miRNA (microRNA)'ların tanımlanmasını da sağlamaktadır [35]. Etkileşimli verinin (interaction data) türler arası analizi ise türlere göre farklılık gösteren etkileşim modüllerinin, regülör programlarının (regulator programmes) tespitini kapsamaktadır [36; 37].

Sekans verisi biyolojik sistemlerin işlem zamanı süresince farklılık göstermemektedir. Etkileşimler ise durum ve zamana bağlı değişiklik gösterirken, neredeyse tüm etkileşim verileri bir zamana ve duruma bağlı olarak sabit kalmaktadır [38; 39]. Bu yüzden hangi genlerin hangi biyolojik işlemlerde yer aldığını bulmada sadece bu tür verilerin kullanılması birtakım zorluklar çıkarmaktadır. Ayrıca, sekanslarda ve etkileşimli verilerdeki büyük değişiklikler bazı durumlarda fonksiyonları az etkilerken, bazı durumlarda ise iki gen arasındaki ufak bir sekans değişikliği yapıların fonksiyonelliğinde büyük değişimlere sebep olabilmektedir [40].

Belirtilen zorluklara karşılık araştırmacılar daha sağlıklı analizler yapabilmek için veri elde etmede farklı teknikleri kullanmışlardır. Mikroçip teknolojilerini kullanarak birbirleriyle etkileşimli, yüzlerce proteini içeren biyolojik sistemlerin dinamik, duruma bağlı tepkilerini ölçerek mikroçip verilerini elde etmişlerdir. Elde edilen veriler hücre döngüsü (cycle), bağışıklık ve diğer stres tepkileri, günlük ritim (circadian rhythm), gelişim işlemlerini kapsamaktadır. Birçok tür arasında paylaşılan bu işlemlerin analiz edilmesi, karşılaştırılması ile birçok gen grubunun tanımlanması yapılabilmektedir. Bu genler türler arası olan, sekans ve ifade verilerinde saklanan ve biyolojik sistemlerin araştırılması konusunda anahtar bileşen olan genlerdir.

Sekans ve etkileşimli verinin türler arası analizinde çok sayıda teknik kullanılmaktadır. Sekans verisinin dizilimi için çeşitli kombinasyonel ve olasılıksal yöntemler [41;42] kullanılırken, tüm genom (Bir organizmanın kalıtım materyalinde bulunan genetik şifrelerin tamamı) dizilimleri için ise çizge-tabanlı algoritmalar kullanılmaktadır [43]. Son yıllarda ise araştırmacılar etkileşimli ağların türler arası analizi için karmaşık hesaplamalı yöntemler geliştirmişlerdir.

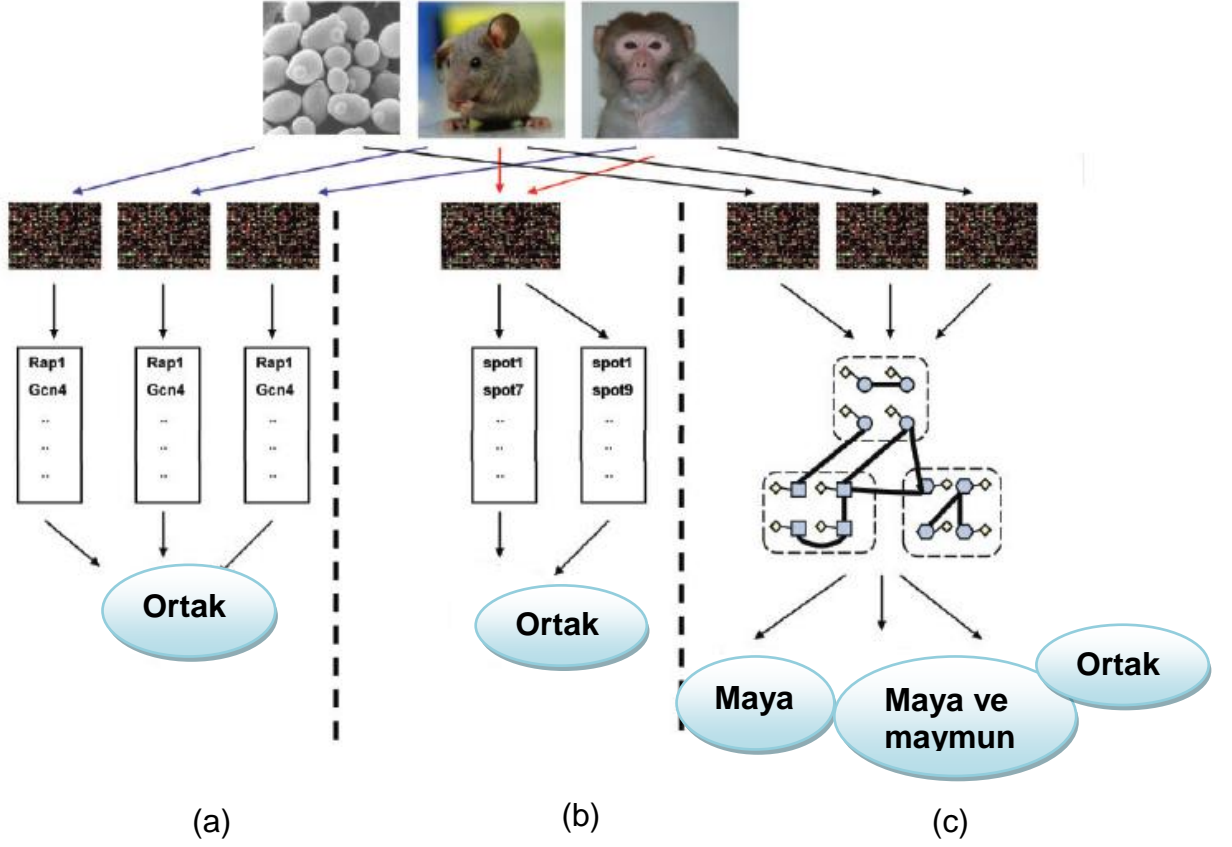
Analiz verilerinin mikroçip teknolojileri ile elde edilmesi birtakım sıkıntıları da beraberinde getirmektedir. Mikroçip verilerinin gürültü içeren (kayıp değerli) veriler olması, aynı organizmaların benzer işlevlerinin farklı laboratuarlardaki ölçümlerinin farklı olması bahsedilen sıkıntılardan bazılarıdır.

Az sayıda harfle (DNA) ifade edilebilen sekans veya etkileşimli verilerden farklı olarak, mikroçip ölçüm verileri türler arası analizde kıyaslanması zor sürekli değerleri içeren veri türleridir. Örneğin, insan ve mayaların hücre döngülerinde benzerlikler olsa da, döngü süreleri farklıdır (maya hücrelerinde 90 dk, insan hücrelerinde için 24 saat). Benzer şekilde mikroçip verilerindeki üstsel artışa rağmen, durum sayısının çok çeşitli olması türler arası analizdeki kıyaslamayı zorlaştıran bir diğer durumdur.

Türler arası meta-analiz temelde iki gruba ayrılmaktadır; eş ifade meta analizi (coexpression meta-analysis) ve ifade meta analizi (expression meta-analysis). Eş ifade meta analizi, bir türde eş ifade edilen genin diğer bir türde de eş ifade edilip edilmediğini araştırırken, ifade meta analizi farklı türlerdeki benzer genlerin ifade profillerindeki benzerlikleri araştırmaktadır. Ayrıca, eş ifade meta analizi farklı türlerle çalışılırken farklı durumların kullanılmasına izin verirken, ifade analizi tüm türlerde aynı durumlarda çalışılmasına izin vermektedir.

Şekil 2.5' de türler arası analiz yöntemlerinin farklı mikroçip veri kümelerini birleştirmeleri ve kıyaslamaları görüntülenmektedir. Şekil 2.5.a 'da ifade meta analizi verilmiştir. Farklı türlerden alınan gen ifade verileri farklı matrislerde tutulmakta ve diferansiyel olarak ifade edilen genleri tanımlamak için birbirlerinden bağımsız olarak analiz edilmektedir. Analiz sonucu elde edilen genler daha sonra türler arasında ortak olanları bulmak için kıyaslanmaktadır. Şekil 2.5.b' de ise farklı türlerden alınan gen ifade verilerinin aynı matriste toplandığı ve analiz için aynı

yöntemin kullanılarak ortak genlerin elde edildiği görülmektedir. Yalnız bu yöntemin birbirine çok benzer olan türlerin analizinde kullanılabileceğini söylemek mümkündür. Şekil 2.5.c' deki şekilde ise gen ifade verilerinin eş zamanlı olarak analiz edildiği görülmektedir. Her türden alınan örneklem farklı matrislere alınmakta fakat birlikte analiz edilmektedirler.



Şekil 2. 5 Türler arası analiz örneği [33]

3. YÖNTEMLER

Bu çalışmada tanıttığımız TriClustering yöntemi, yaygın kullanılan ikili kümeleme yöntemi olan CC algoritmasının üç yönlü kümelemeye uyarlanması ile geliştirilmiş yeni bir yöntemdir. CC algoritmasının tüm modülleri gen, durum (örneklem, deney) ve organizmadan oluşan üç boyutlu veri matrisinden anlamlı üçlü kümeler çıkarabilmek için uyarlanmıştır.

3.1 Üçlü Küme (Tricluster) Modeli

Üçlü küme ikili gruplar gibi gen ifade matrisinde tanımlanmaktadır. Gen ifade matrisi satırların genleri, sütunların örneklemeleri, derinliğin organizmayı temsil ettiği $m \times n \times k$ boyutlarında bir veri matrisidir. Matristeki (i, j, k) inci eleman, k inci organizmanın j örneklemindeki i geninin ifade değerinin reel sayı olarak değeridir.

İkili küme, belirli bir grup örnekleme veya durumda benzer davranışları gösteren genlerin veya belirli bir gen grubunun altında benzer davranışı gösteren örneklemelerin grubu olarak tanımlanmaktadır. Üçlü küme ise farklı organizmalarda tanımlanmış olan ikili gruplar kümesidir. Belirli bir örnekleme karşılık aynı grupta olan genler, gen ifade seviyeleri paralel bir artış veya azalış gösterdikleri için uyumlu (coherent) genler olarak isimlendirilmektedir. Aynı durum belirli bir gen grubuna karşılık aynı grupta olan örneklemeler için de geçerlidir.

Cheng ve Church ikilisine göre; bir ikili küme uyumluluk derecesi o gruptaki tüm elemanların varyansını temsil eden Orta Kare Değeri (Mean Squared Residue MSR, H) değeri ile ölçülmektedir (3.1). TriClustering yönteminde CC algoritmasında tanımlı tüm formüller üç boyutlu veri matrisine uyarlanması için revize edilmiştir. Herhangi bir üçlü kümedeki elemanın (a_{ijk}) kalıntı (residue) değeri, $a_{ijk} - a_{iJK} - a_{iJk} + a_{iJK} + a_{iJk}$ formülü ile hesaplanmaktadır. Burada a_{iJK} (3.2) değeri üçlü grubun i inci satırının ortalamasını, a_{iJk} (3.3) j inci sütununun ortalamasını, a_{iJK} (3.4) k inci organizmasının veri matrisinin ortalamasını temsil ederken, a_{iJK} (3.5) üçlü kümedeki tüm elemanların ortalamasını temsil etmektedir.

$$H(I, J, K) = \frac{1}{|I||J||K|} \sum_{i \in I, j \in J, k \in K} (a_{ijk} - a_{iJK} - a_{IJk} - a_{IJk} + a_{IJK})^2 \quad (3.1)$$

$$a_{iJK} = \frac{1}{|J||K|} \sum_{j \in J, k \in K} a_{ijk} \quad (3.2)$$

$$a_{IJk} = \frac{1}{|I||K|} \sum_{i \in I, k \in K} a_{ijk} \quad (3.3)$$

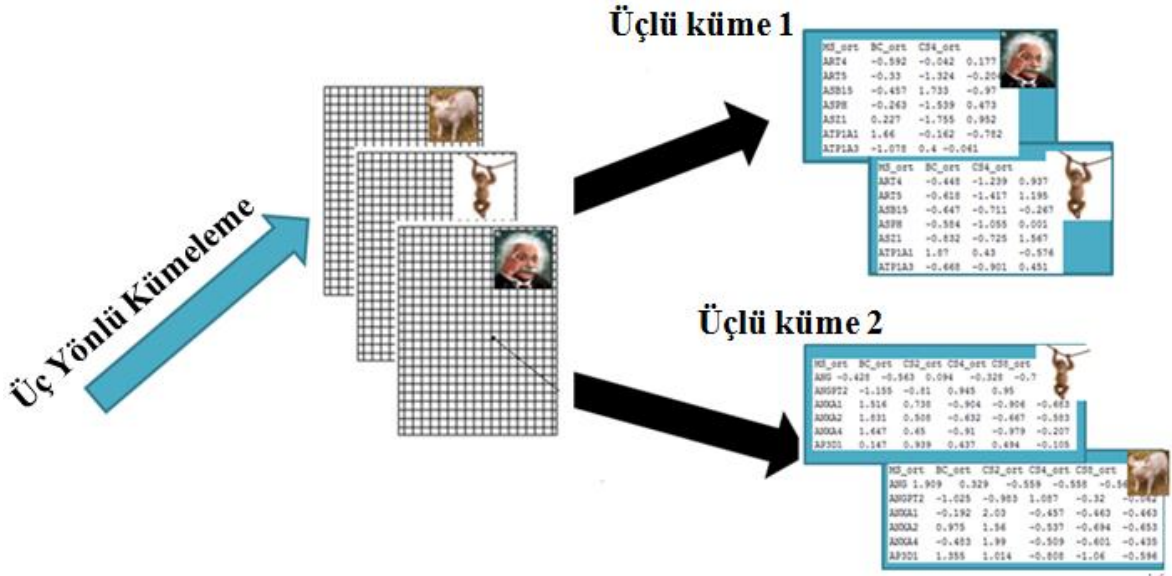
$$a_{IJk} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ijk} \quad (3.4)$$

$$a_{IJK} = \frac{1}{|I||J||K|} \sum_{i \in I, j \in J, k \in K} a_{ijk} \quad (3.5)$$

Bir ikili kümenin niteliği o gruptaki elemanların birbirine olan benzerliğini, uyumunu ölçen H değeri ile değerlendirilmektedir. Düşük H değeri grup elemanlarının benzerliğinin yüksek olduğunu gösterirken, yüksek H değeri elemanların uyumunun düşük olduğunu göstermektedir. TriClustering algoritması δ -üçlü küme (δ -tricluster) olarak isimlendirilen, H değeri delta (δ) değerinden küçük olan bir alt matris bulmayı hedeflemektedir. Algoritmanın temel amacı ise üçlü kümelerin boyutunu arttırmak ve MSR (H) değerini minimize etmektir.

3.2 TriClustering Algoritması

TriClustering algoritmasının temel amacı farklı organizmaların gen ifade matrisinden oluşan veri kümesini kullanarak n adet δ -üçlü grubun bulunmasıdır. Algoritma tüm veri kümesi (tek üçlü küme) ile işleme başlamakta, delta (δ) değeri kalıntı değerinden küçük veya eşit olana kadar tekrarlı olarak satırları, sütunları ve kaynakları silmektedir. Silme işleminden sonra, herhangi bir satırın, sütunun veya kaynağın eklenmesi H değerini yükseltene kadar silinen satır, sütun veya kaynaklardan eklemeler yapılır. İlk üçlü küme elde edildikten sonra, orijinal veri kümesinden üçlü kümede olan bileşenler çıkarılarak güncellemesi yapılır (Bileşen gen, örneklem veya organizmayı temsil etmektedir.). CC algoritmasının üç boyutlu veri matrisi için uyarlanan fonksiyonlarının tanımı bir sonraki kısımda detaylı olarak verilmiştir. Ayrıca algoritmanın üç organizmalı örnek bir veri kümesine uyarlanması ve sonrasında elde edilen örnek üçlü kümeler Şekil 3.1' de gösterilmiştir.



Şekil 3. 1 TriClustering Algoritmasının örnek bir veri kümesine uyarlanması

3.2.1 Algoritma-1: Tek bileşen silinmesi (Single node deletion)

Algoritma-1 en büyük MSR değerine sahip satır, sütun ve organizmanın tespit edilmesi ve bu bileşenin silinmesini içermektedir. Satır ve sütun silinmesinde matristen bir bileşen eksilirken, organizma silinmesinde veri matrisinden organizma matrisi silinmektedir.

Algoritma Adımları :

Girdi : Üç boyutlu gen ifade matrisi A, MSR değeri için belirlenen sınır değeri; delta (δ) ≥ 0

Çıktı : Alt matris, üçlü küme A_{IJ}

Atamalar : $A_{IJ} = A$

İterasyon :

1. Tüm satırların ($i \in I$) a_{iJK} değeri, sütunların ($j \in J$) a_{iJK} değeri, organizmaların ($k \in K$) a_{iJK} değeri ve $H(I, J, K)$ değeri hesaplanır.

Eğer ; $H(I, J, K) \leq \delta$ ise A_{IJ} döndürülerek işlem bitirilir.

2. En büyük d değerine sahip olan satır bulunur;

$$d(i, k) = \frac{1}{|J||K|} \sum_{j \in J, k \in K} (a_{ijk} - a_{iJK} - a_{iJK} - a_{iJK} + a_{iJK})^2$$

En büyük d değerine sahip sütun bulunur;

$$d(j, k) = \frac{1}{|I||K|} \sum_{i \in I, k \in K} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2$$

En büyük d değerine sahip organizma matrisi bulunur;

$$d(j, k) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2$$

3. En büyük d değerine sahip olan satır, sütun ve organizma matrisi silinir.

3.2.2 Algoritma-2: Çoklu bileşen silinmesi (Multiple node deletion)

Algoritma-2'de satır, sütun ve organizma grupları silinmektedir. İterasyona bağlı olarak elde edilen üçlü kümenin MSR değeri azaldıkça bileşen gruplarının silinmesine devam edilmektedir.

Algoritma Adımları :

Girdi : Üç boyutlu gen ifade matrisi A, MSR değeri için belirlenen sınır değeri; delta (δ) ≥ 0 , çoklu silme için sınır parametresi; alfa (α) > 0.1 .

Çıktı : Alt matris, üçlü küme A_{IJ}

Atamalar : $A_{IJ} = A$

İterasyon :

1. Tüm satırların ($i \in I$) a_{iJK} değeri, sütunların ($j \in J$) a_{IJk} değeri, organizmaların ($k \in K$) a_{IJK} değeri ve $H(I, J, K)$ değeri hesaplanır.

Eğer ; $H(I, J, K) \leq \delta$ ise A_{IJ} döndürülerek işlem bitirilir.

2. Aşağıdaki eşitsizliği sağlayan satırlar silinir;

$$\frac{1}{|J||K|} \sum_{j \in J, k \in K} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 > \alpha H(I, J, K)$$

3. a_{iJK} , a_{Ijk} , a_{IJk} , a_{IJK} ve $H(I, J, K)$ değerleri yeniden hesaplanır.

4. Aşağıdaki eşitsizliği sağlayan sütunlar silinir;

$$\frac{1}{|I||K|} \sum_{i \in I, k \in K} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 > \alpha H(I, J, K)$$

5. a_{iJK} , a_{Ijk} , a_{IJk} , a_{IJK} ve $H(I, J, K)$ değerleri yeniden hesaplanır.

6. Aşağıdaki eşitsizliği sağlayan organizma matrisleri silinir;

$$\frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 > \alpha H(I, J, K)$$

7. Eğer herhangi bir satır, sütun veya organizma matrisi silinmediyse Algoritma-1' e dönülür.

3.2.3 Algoritma-3: Bileşen eklenmesi (Node addition)

Algoritma-3'de önceki algoritmalarda silinen bileşenlerden, MSR değerini yükseltmeyenlerin tekrar eklenmesi yapılmaktadır.

Algoritma Adımları :

Girdi : Üç boyutlu gen ifade matrisi A, MSR değeri için belirlenen sınır değeri; delta (δ) ≥ 0 , çoklu silme için sınır parametresi; alfa (α) > 0.1 .

Çıktı : Alt matris, üçlü küme A_{IJ}

Atamalar : $A_{IJ} = A$

İterasyon :

1. Tüm satırların ($i \in I$) a_{iJK} değeri, sütunların ($j \in J$) a_{Ijk} değeri, organizmaların ($k \in K$) a_{IJK} değeri ve $H(I, J, K)$ değeri hesaplanır.

Eğer ; $H(I, J, K) \leq \delta$ ise A_{IJ} döndürülerek işlem bitirilir.

2. Üçlü kümede olmayan ve aşağıdaki eşitsizliği sağlayan sütunlar eklenir;

$$\frac{1}{|I||K|} \sum_{i \in J, k \in K} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 \leq H(I, J, K)$$

3. a_{iJK} , a_{Ijk} , a_{IJk} , a_{IJK} ve $H(I, J, K)$ değerleri yeniden hesaplanır.

4. Üçlü kümede olmayan ve aşağıdaki eşitsizliği sağlayan satırlar eklenir;

$$\frac{1}{|J||K|} \sum_{j \in J, k \in K} (a_{ijk} - a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 \leq H(I, J, K)$$

5. a_{iJK} , a_{Ijk} , a_{IJk} , a_{IJK} ve $H(I, J, K)$ değerleri yeniden hesaplanır.

6. Üçlü kümede olmayan ve aşağıdaki eşitsizliği sağlayan satırların tersleri eklenir;

$$\frac{1}{|J||K|} \sum_{j \in J, k \in K} (-a_{ijk} + a_{iJK} - a_{Ijk} - a_{IJk} + a_{IJK})^2 \leq H(I, J, K)$$

7. a_{iJK} , a_{Ijk} , a_{IJk} , a_{IJK} ve $H(I, J, K)$ değerleri yeniden hesaplanır.

8. Üçlü kümede olmayan ve aşağıdaki eşitsizliği sağlayan organizma matrisleri eklenir;

$$\frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ijk} - a_{iJK} - a_{IjK} - a_{IJk} + a_{IJK})^2 \leq H(I, J, K)$$

9. Eğer herhangi bir satır, sütun veya organizma matrisi eklenmedi ise son I, J değerleri geri döndürülür.

3.2.4 Algoritma-4: TriClustering algoritması

Algoritma-4'de Triclustering algoritmasının tüm fonksiyonlarının bir arada uygulanışı gösterilmektedir.

Algoritma Adımları :

Girdi : Üç boyutlu gen ifade matrisi A, MSR değeri için belirlenen sınır değeri; delta (δ) ≥ 0 , çoklu silme için sınır parametresi; alfa (α) > 0.1 , iterasyon sayısı n, kalan satır,sütun ve organizma sayısı için sınır parametreleri sırası ile n_r, n_c, n_s .

Çıktı : Alt matris,üçlü küme A_{IJ}

Atamalar : Orijinal A matrisinin kopyası alınır ; Kopya_A = A ;

İterasyon (n adet):

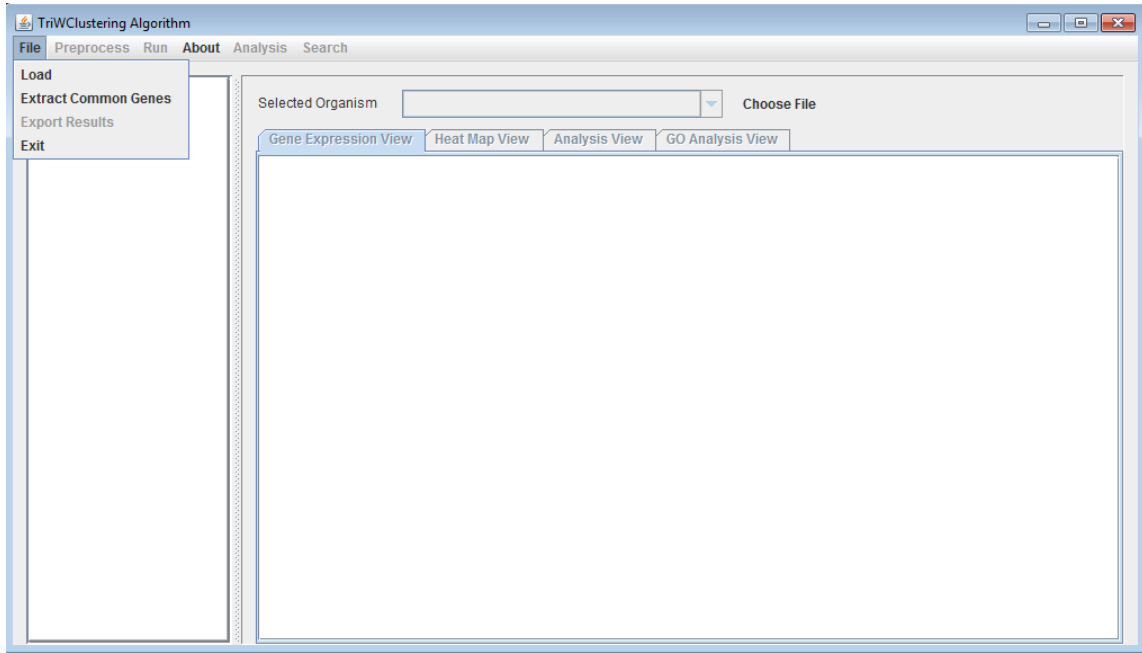
1. B matrisi Kopya_A matrisinin kopyasıdır.
2. Algoritma-2 B matrisine uygulanır. Eğer satır sayısı n_r 'den, sütun sayısı n_c 'den veya organizma sayısı n_s 'den küçük ise Çoklu Bileşen Silme Algoritması uygulanmaz.
3. Eğer Adım-2'de herhangi bir satır, sütun veya organizma matrisi silinmedi ise Algoritma-1 B matrisine uygulanır, sonuç matrisi C matrisidir.
4. Algoritma-3 C ve Kopya_A matrislerine uygulanır, sonuç matrisi üçlü küme olan D matrisidir.
5. D matrisi kaydedilir, D'de olan bileşenler Kopya_A matrisinden silinerek bir sonraki iterasyon için Kopya_A matrisi güncellenir.

4. GELİŞTİRİLEN ARAÇ

TriClust üç yönlü kümeleme yapan bir yazılım aracıdır. Kullanıcılar veri kümelerini araca yükleyebilir, ısı haritasını görüntüleyebilir, geliştirilen yöntemi veri kümesi üzerinde uygulayabilir, gen araması yapabilir ve analiz sonuçlarını elde edebilmektedirler.

4.1 İşlevler ve Kullanıcı Arayüzleri

TriClustering aracı çalıştırıldığında Şekil 4.1' de görünen arayüz kullanıcıyı karşılamaktadır. Bu arayüzde veri kümesi yükleme, ön işleme, algoritmayı çalıştırma, analiz ve arama işlemlerinin yapılabildiği görülmektedir. Fakat ilk çalıştırmada sadece veri kümesi yükleme işlemi ve ortak gen çıkarımı işlemleri aktif yapılarak kullanıcı veri kümesi yükleme işlemine yönlendirilmektedir. Kullanıcı farklı organizmalardan oluşan veri kümesi yüklendikten, ortak gen çıkarımı işlemini yaptıktan sonra Şekil 4-2'de görüldüğü üzere diğer işlemler de aktif hale getirilir.



Şekil 4. 1 Triclustering giriş arayüzü

Ortak gen çıkarımı (Extract common genes) işlemi yapılması zorunlu olan bir işlemdir, çünkü algoritma aynı gen (satur) ve durumlara (sütun) sahip farklı organizmaların gen ifade verilerini kullanarak analiz yapmaktadır. Dolayısıyla yüklenen veri kümesinin organizmalarının aynı sayıda ve sırada gen ve duruma

sahip olması beklenmektedir, aksi takdirde araç kullanıcıya ortak gen ve duruma sahip organizmaları yüklemesi konusunda uyarı vermektedir. Yüklenen organizmaların durum sayı ve sıralarının aynı olması beklenirken, genler için böyle bir şart aranmamaktadır. Araç, "Ortak Genleri Çıkart" (Extract Common Genes) komutu verildiğinde organizmaların ortak olan genlerini bulup aynı sırada olacağını şekilde çok boyutlu matrisi oluşturup, işleme hazır hale getirmektedir.

Kullanıcı veri kümesini yükledikten sonra, soldaki görünüm ağaç yapısından (Orijinal veri, Ön işlenmiş veri vb.) ve organizma listesinden yapacağı seçime göre veri kümesi ile ilgili görüntüleme (Gen ifade verisi görünümü, Isı haritası görünümü, Analiz görünümü, GO analizi görünümü) yapabilmektedir. Şekil 4.2' de insan (human) organizmasına ait gen ifade verisi görüntülenirken, yüklenen verinin ön işleme işlemi de yapılabilmektedir. Ön işleme işlemi her satırın ortalaması

gene_names	CS1_ort	CS2_ort	CS4_ort	CS8_ort	MS_ort	BC_ort
ABAT	1.0	1.44	1.73	3.0	1.0	1.43333333
ABAT	96.83666667	69.85666667	92.45333333	146.49	40.33	25.22666667
ABCA4	16.61	17.48666667	15.33333333	60.1	21.30333333	17.84666667
ABCC1	214.44	188.39	225.07	221.01	26.14	29.85333333
ACACA	73.67	79.72	138.40666667	297.52	174.86666667	473.54333333
ACIDL	107.80333333	137.99	160.16333333	939.64	72.04	248.35666667
ACE	7.306666667	6.61	7.863333333	35.72	14.28333333	9.463333333
ACO2	112.1066667	109.8166667	190.0466667	623.38	138.72333333	293.3466667
ACP1	8.213333333	12.21666667	8.486666667	35.01	6.986666667	7.406666667
ACR	170.8	153.2933333	146.0666667	262.44	52.99666667	44.14
ACTA1	9.443333333	16.29666667	11.69666667	226.79	14.79	13.16666667
ACTB	926.4766667	1019.18	1837.2933333	8570.26	2195.02	4255.966667
ACTR3	476.0066667	411.7466667	558.0266667	846.68	479.73	540.4966667
ACVR1	643.9033333	526.5366667	688.2166667	317.46	97.78666667	275.1433333
ACVR2B	6.306666667	8.923333333	5.773333333	14.26	3.673333333	1.596666667
ADA	1435.976667	1314.313333	1369.486667	1107.05	312.5033333	567.2133333
ADAM10	711.7566667	423.1666667	592.34	362.0	167.9933333	433.3233333
ADAM12	54.03	43.87333333	51.35333333	177.81	40.93666667	35.48666667
ADAM2	9.223333333	12.0	8.046666667	29.72	12.1	7.603333333
ADAM9	26.03666667	26.84333333	25.09666667	85.44	18.84666667	19.52666667
ADAMTS4	44.69333333	42.15666667	41.11333333	147.4	52.98333333	47.59333333
ADAMTS5	11.13333333	13.71	12.22	51.25	22.37333333	16.83333333
ADCY7	6.9	15.06333333	13.26	24.9	8.646666667	8.943333333
ADCYAP1R1	13.85	11.22666667	10.93	29.41	12.18	8.943333333
ADM	169.07	132.4233333	788.9233333	3875.38	1636.906667	1835.046667
ADORA1	38.39666667	14.3	13.83	30.02	12.24666667	8.336666667
ADORA1A	9.716666667	9.713333333	6.73	50.06	12.27666667	13.99333333
ADRA2A	17.65666667	10.87333333	15.36	71.8	16.88	26.18666667
ADRA2C	28.89666667	24.31	35.18333333	136.23	47.90333333	68.15333333
ADRB1	4.27	3.933333333	5.076666667	11.05	5.656666667	4.91
ADRB2	19.16	22.06333333	22.5	65.79	26.79333333	20.35333333
ADRB3	17.01666667	13.26666667	13.94333333	65.52	24.24666667	17.03333333
ADRBK1	51.64333333	62.38	68.52	101.83	68.73666667	43.81
AGER	7.96	4.563333333	2.9	29.47	8.306666667	9.626666667
ACPAT1	74.27	80.00333333	77.44333333	162.38	91.70666667	74.59666667
AGRP	29.35	28.16	31.71	166.76	71.24666667	59.78
AHR	8.096666667	4.98	6.426666667	17.54	5.323333333	9.293333333
AHSB	5.07	10.86666667	5.15	20.67	9.43	3.18
AIF1	25.47333333	22.25	21.12333333	70.28	20.28333333	20.13333333
AIP	46.64666667	40.79333333	47.81333333	114.33	31.84	29.89666667
AIP1	50.16666667	50.45333333	52.02	145.47	54.36	55.63333333
AK2	14.34333333	15.61666667	15.52333333	52.06	27.91333333	19.58666667
AK3	7.416666667	6.316666667	6.276666667	20.83	6.933333333	8.506666667

Şekil 4. 2 Orijinal verinin gen ifade verisi görünümü

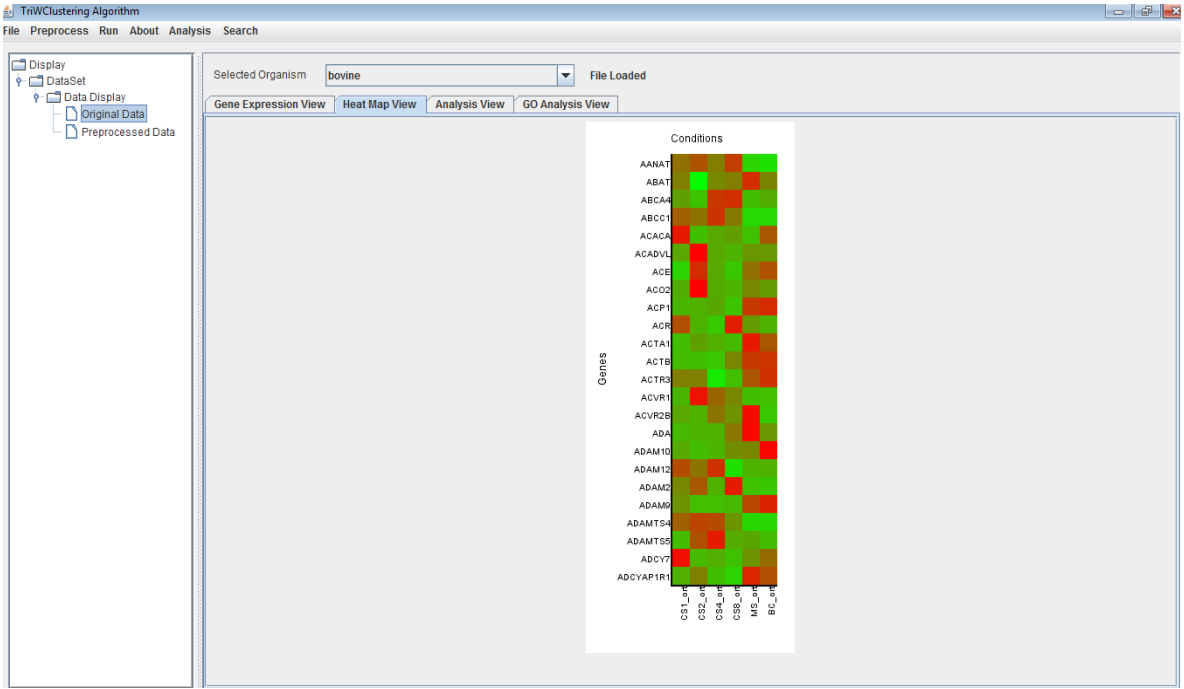
0, varyansı 1 olacak şekilde tüm sayılar 0-1 arasında bir sayıya ölçeklendirilmektedir. Kullanıcı soldaki ağaç yapısından ön işlenmiş veriyi seçerek veriyi görüntüleyebilmektedir (Şekil 4.3).

Gene Expression View	Heat Map View	Analysis View	GO Analysis View			
Selected Organism	human	File Loaded				
Gene Names	CS1_ort	CS2_ort	CS4_ort	CS8_ort	MS_ort	BC_ort
ANKK1	-0.81	-0.217	0.175	1.897	-0.81	-0.225
ABAT	0.418	-0.204	0.317	1.563	-0.885	-1.21
ABCA4	-0.47	-0.419	-0.543	2.028	-0.2	-0.399
ABCC1	0.663	0.392	0.774	0.732	-1.299	-1.261
ACACA	-0.86	-0.821	-0.441	0.592	-0.204	1.733
ACADVL	-0.516	-0.424	-0.357	2.008	-0.624	-0.089
ACE	-0.557	-0.619	-0.507	1.979	0.067	-0.364
ACO2	-0.67	-0.682	-0.276	1.915	-0.535	0.247
ACPF1	-0.444	-0.077	-0.419	2.012	-0.556	-0.516
ACR	0.399	0.187	0.097	1.532	-1.052	-1.161
ACTA1	-0.45	-0.372	-0.424	2.041	-0.389	-0.408
ACTB	-0.756	-0.724	-0.444	1.86	-0.322	0.384
ACTR3	-0.494	-0.916	0.038	1.921	-0.473	-0.077
ACVR1	0.946	0.439	1.137	-0.464	-1.422	-0.646
ACVR2B	-0.102	0.489	-0.222	1.693	-0.696	-1.164
ADA	0.894	0.634	0.752	0.191	-1.508	-0.963
ADAM10	1.4	-0.135	0.765	-0.46	-1.491	-0.081
ADAM12	-0.243	-0.429	-0.292	2.026	-0.483	-0.582
ADAM2	-0.466	-0.134	-0.607	1.987	-0.122	-0.66
ADAM9	-0.297	-0.266	-0.334	2.024	-0.578	-0.551
ADAMTS4	-0.431	-0.492	-0.517	2.031	-0.232	-0.361
ADAMTS5	-0.664	-0.495	-0.593	1.969	0.074	-0.29
ADCY7	-0.698	0.283	-0.005	1.848	-0.738	-0.691
ADCYAP1R1	-0.077	-0.426	-0.465	1.995	-0.299	-0.73
ADM	-0.878	-0.904	-0.451	1.758	0.167	0.308
ADORA1	1.591	-0.441	-0.48	0.865	-0.614	-0.943
ADRA1A	-0.45	-0.45	-0.633	2.019	-0.283	-0.194
ADRA2A	-0.387	-0.685	-0.488	1.992	-0.421	-0.012
ADRA2C	-0.664	-0.774	-0.515	1.892	-0.212	0.271
ADRB1	-0.587	-0.715	-0.281	1.967	-0.061	-0.344
ADRB2	-0.572	-0.411	-0.386	2.02	-0.148	-0.506
ADRB3	-0.405	-0.591	-0.558	2.003	-0.046	-0.404
ADRBK1	-0.724	-0.189	0.119	1.78	0.129	-1.115
AGER	-0.261	-0.613	-0.786	1.971	-0.225	-0.088
AGPAT1	-0.52	-0.396	-0.451	2.022	-0.144	-0.513
AGRP	-0.675	-0.679	-0.612	1.921	0.13	-0.086
AHR	-0.11	-0.777	-0.468	1.911	-0.704	0.147
AHSG	-0.625	0.283	-0.613	1.818	0.058	-0.921
AF1	-0.225	-0.387	-0.443	2.032	-0.486	-0.493
AP	-0.17	-0.352	-0.129	1.985	-0.637	-0.699
APL1	-0.49	-0.387	-0.441	2.039	-0.379	-0.345
AK2	-0.677	-0.589	-0.595	1.918	0.258	-0.316
AK3	-0.417	-0.254	-0.624	2.018	-0.505	-0.219

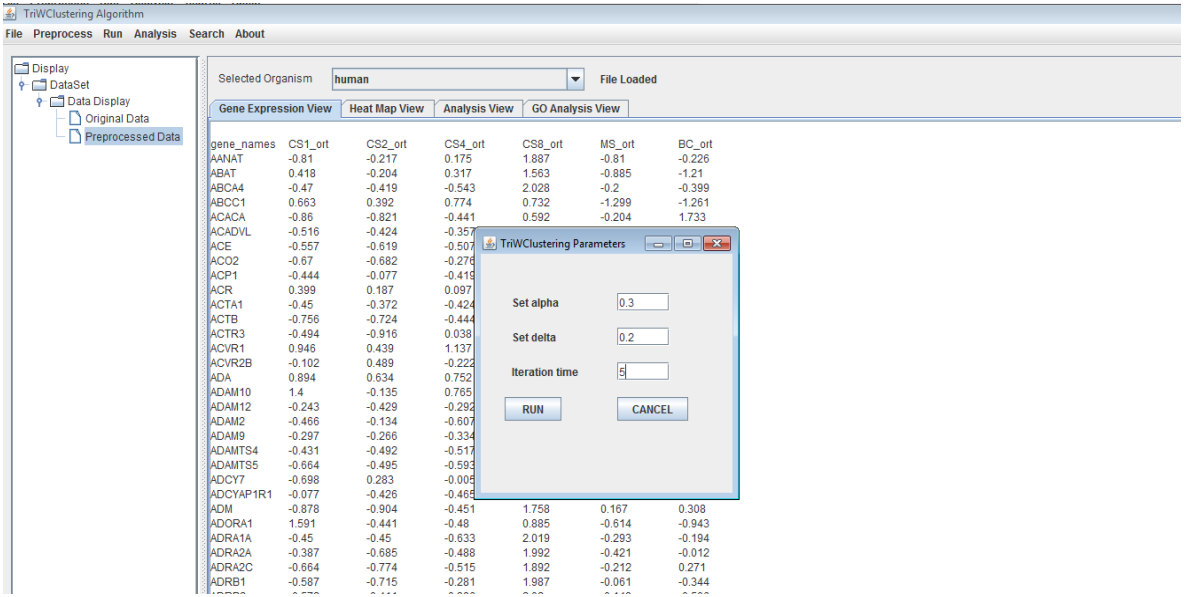
Şekil 4. 3 Ön işlenmiş verinin gen ifade verisinin görünümü

Şekil 4.4 'de ise sığır (bovine) organizmasına ait ısı haritası görüntülenmektedir. Isı haritası, matris verisindeki sayısal değerlerin renklerle kodlandığı grafiksel bir veri sunumudur. Büyük değerler koyu yeşil veya siyah ile gösterilirken, küçük değerler daha açık renk tonları ile gösterilmektedir.

Veri kümesinin yükleme ve ön işleme işlemleri yapıldıktan sonra TriClustering algoritmasını çalıştırmak için Şekil 4.5'te görüldüğü gibi kullanıcı gerekli parametreleri girmelidir. Eğer kullanıcı soldaki ağaç yapısından veri ile ilgili bir seçim yapmamış ise algoritma orijinal veri üzerinden, ağaçtan seçim yapılmış ise seçilen veriye göre çalıştırılacaktır.

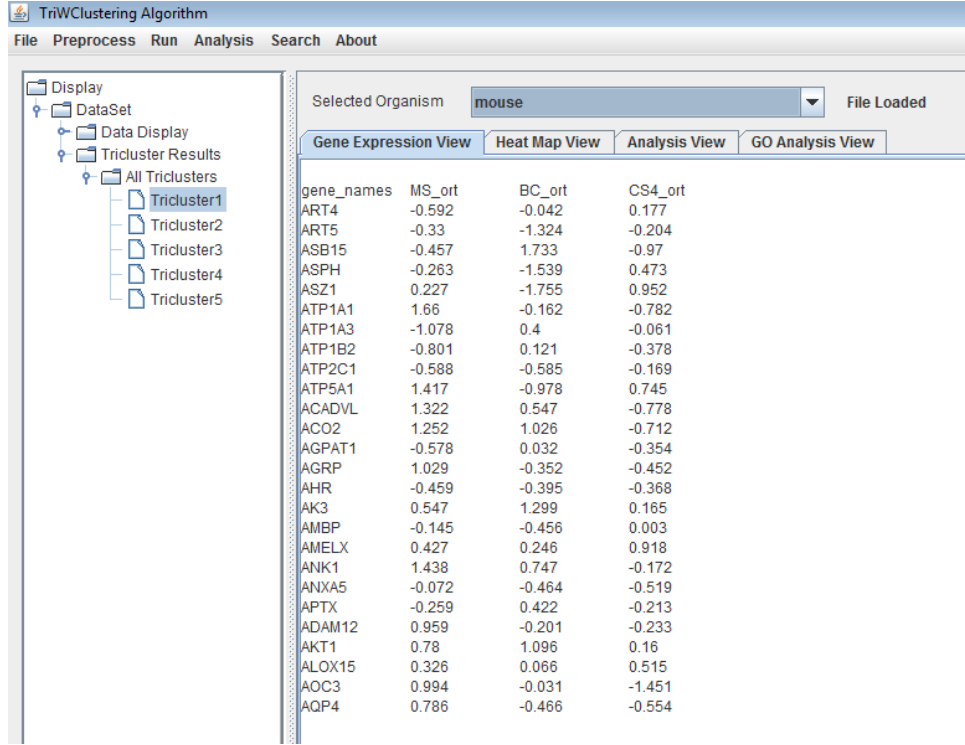


Şekil 4. 4 Orijinal verinin ısı haritası görünümü



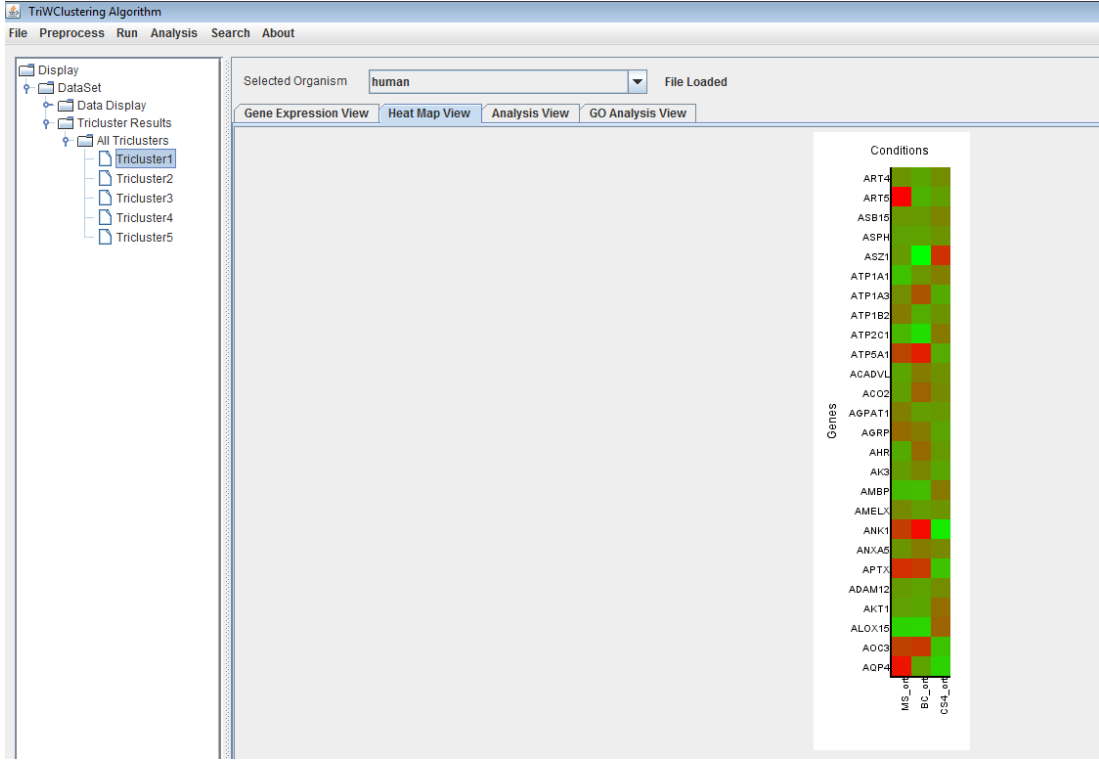
Şekil 4. 5 TriClustering algoritmasının çalıştırma arayüzü

Algoritmanın çalıştırılması sonucunda oluşan üçlü kümeler ağaç yapısına eklenmekte, seçilen üçlü kümenin Şekil 4-6'daki gibi gen ifade verisi, Şekil 4-7'deki gibi ısı haritası görüntülenebilmektedir.



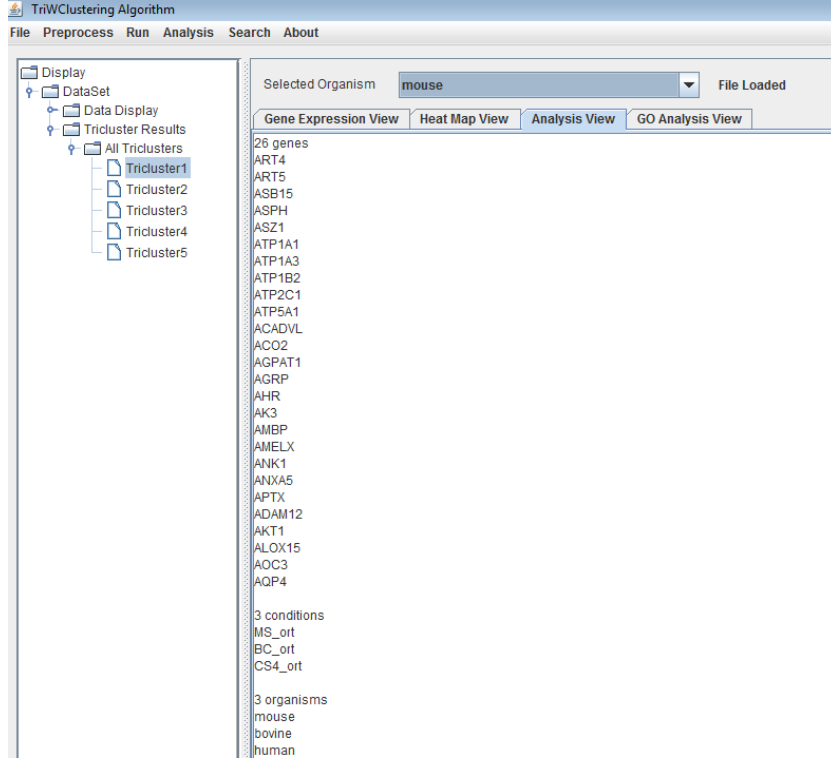
gene_names	MS_ort	BC_ort	CS4_ort
ART4	-0.592	-0.042	0.177
ART5	-0.33	-1.324	-0.204
ASB15	-0.457	1.733	-0.97
ASPH	-0.263	-1.539	0.473
ASZ1	0.227	-1.755	0.952
ATP1A1	1.66	-0.162	-0.782
ATP1A3	-1.078	0.4	-0.061
ATP1B2	-0.801	0.121	-0.378
ATP2C1	-0.588	-0.585	-0.169
ATP5A1	1.417	-0.978	0.745
ACADVL	1.322	0.547	-0.778
ACO2	1.252	1.026	-0.712
AGPAT1	-0.578	0.032	-0.354
AGRP	1.029	-0.352	-0.452
AHR	-0.459	-0.395	-0.368
AK3	0.547	1.299	0.165
AMBP	-0.145	-0.456	0.003
AMELX	0.427	0.246	0.918
ANK1	1.438	0.747	-0.172
ANXA5	-0.072	-0.464	-0.519
APTX	-0.259	0.422	-0.213
ADAM12	0.959	-0.201	-0.233
AKT1	0.78	1.096	0.16
ALOX15	0.326	0.066	0.515
AOC3	0.994	-0.031	-1.451
AQP4	0.786	-0.466	-0.554

Şekil 4. 6 Oluşan üçlü kümelerden birinin gen ifade verisi görünümü



Şekil 4. 7 Oluşan üçlü kümelerden birinin ısı haritası görünümü

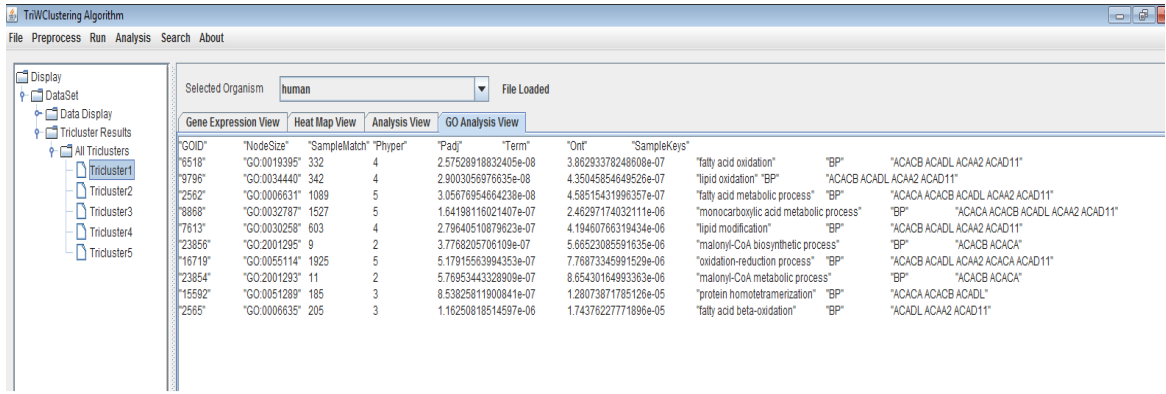
Oluşan üçlü kümede hangi genlerin, durumların ve organizmaların olduğunu görüntüleyebilmek için ise Şekil 4.8' de görülen analiz görüntülemesi yapılabilmektedir. Analiz görünümünde hangi genlerin hangi durumlarda aktif olduklarını görebilmek ve hangi organizmaların aynı üçlü kümede olduklarını tespit edebilmek biyolojik sonuç analizinde kullanılan önemli verilerdir.



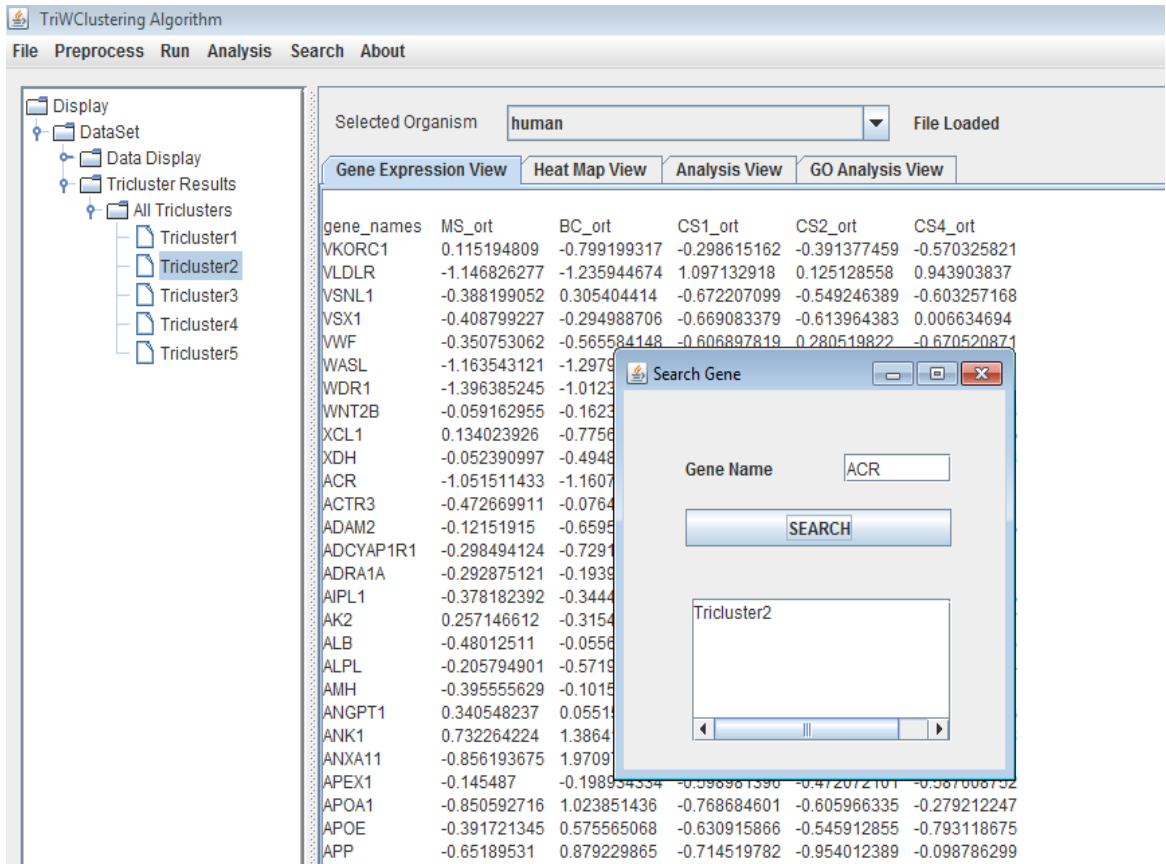
Şekil 4. 8 Oluşan üçlü kümelerden birinin analiz görünümü

Bu görünümlere ek olarak GO (Gene Ontology) analizi sonuçlarının görüntülenmesi de Şekil 4-9'da verilmiştir. TriClustering yöntemi çalıştırma sonucunda oluşan üçlü kümelerin GO analiz sonuçlarından ilk 10 tanesi araçta görüntülenmiş sonuçların tümü sonuç dosyasına kaydedilmiştir. Şekil 4-9'daki birinci üçlü grubun (Tricuster1) GO analiz sonucundaki GO terimleri, bu terimlerin GO ID leri, eşleşen örneklem sayısı, hesaplanan phyper değeri, ontoloji ve o GO teriminin kapsadığı genlerin isimleri yer almaktadır.

Elde edilen üçlü kümelerde gen aramasının yapılması ise Şekil 4.10'da gösterilmektedir. Bu arama sayesinde oluşan çok sayıdaki üçlü kümelerde hangi genlerin bulunduğu kolaylıkla tespit edilebilmektedir.



Şekil 4. 9 Oluşan üçlü kümelerden birinin (Tricluster1) GO analiz görünümü



Şekil 4. 10 Üçlü kümelerde gen araması yapılması

4.2 Teknik Altyapı

Yazılım geliştirilmesinde Java, R ve MATLAB dilleri kullanılmıştır. Algoritmanın uygulanması ve arayüzlerin geliştirilmesinde Java, GO analizinde ise R dili kullanılırken, Dunn indeks uygulamasında ise MATLAB dili kullanılmıştır. R' da yapılan GO analizi Java ile hazırlanan arayüze entegre edilmiş, MATLAB ile yapılan Dunn indeks uygulaması arayüzden bağımsız olarak yapılmıştır. Entegre işlemi için R komutları bir komut dizisi (script) haline getirilmiş, Java'dan çağırımı ve çalıştırımı gerçekleştirilmiştir.

Kod geliştirme ortamı olarak Netbeans IDE 7.1.2 editörü kullanılmıştır. Yöntemin geliştirildiği ve farklı parametreler ile test edilmesi, 2.10 GHz işlemci hızına sahip Intel(R) Core(TM)2 Duo işlemcili, 4,00 GB RAM'e sahip makinada yapılmıştır.

Geliştirilen yazılım aracı işletim sisteminden bağımsız, Java 1.7, R 2.15 veya üstü sürümlerinin yüklü olduğu her makinada çalışabilmektedir.

5. VERİ KÜMELERİ

Geliştirilen TriClustering yönteminin performansı üç örnek çalışma ile raporlanmıştır. Bu örnek çalışmalar NCBI GEO [44] veri merkezinden alınan 3 farklı gen ifade verisi üzerine yöntemin uygulanması ve sonuçların analizini kapsamaktadır. Farklı organizmaların gen ifade verilerini içeren veri kümelerinde organizmaların ortak genleri ve örneklemi tespit edilerek, veri kümesindeki her organizmanın eşit sayıda gen ve örneklem sahip olması sağlanmıştır.

Birinci veri kümesi (DS-1 olarak kısaltılacaktır) iki organizmadan oluşan; insan (human) ve fare (mus musculus), GSE19184 erişim numarasına sahip bir veri kümesidir. İnsana ait veri kümesinde 48702 gen, 46 örneklem (bu veri kümesinde örneklem hücre tipidir), fareye ait veri kümesinde 46628 gen, 52 örneklem bulunmaktadır. DS-1, tümör oluşumu esnasında insan ve fareye ait kanser hücrelerindeki gen ifade verilerinin değişiminin gözlenmesi için oluşturulmuş bir veri kümesidir. Örnek çalışmada kullanılmak üzere ortak genler ve örneklem çıkarılarak üç boyutlu veri kümesi oluşturulmuştur. Buna ek olarak aynı hücre tipine sahip örneklemelerin ortalama değerleri alınarak örneklem sayısında eleme yapılmıştır. Bu işlemler sonucunda DS-1 veri kümesi 9132 genden, 11 örneklemden ve 2 organizmadan oluşan, kayıp değer içermeyen bir veri kümesi olarak örnek çalışmada kullanılmıştır.

İkinci veri kümesi (DS-2) iki organizmadan oluşan; insan ve fare (*Rattus norvegicus*), GSE34251 erişim numarasına sahip bir veri kümesidir. İnsana ait veri kümesinde 41108 gen ve 60 örneklem, fareye ait veri kümesinde 41120 gen ve 60 örneklem bulunmaktadır. DS-2, insan ve fare karaciğer gözelerinin 2, 3, 7, 8 miktarında tetrachlorodibenzo-p-dioxin (TCDD)'e maruz kaldıklarında gen ifade verilerindeki değişimi gözlemlemek için oluşturulmuş bir veri kümesidir [45]. Deney ölçümleri, beş farklı insan ve fareye 24 saat süreyle miktarları 0.00001-100nM aralığında değişen, 11 farklı TCDD kimyasalı verilerek yapılmıştır. Organizmaların ortak gen ve örneklem çıkarımı yapıldığında ise DS-2, 12890 genden, 59 örneklemden ve iki organizmadan oluşan, kayıp değer içermeyen bir veri kümesi haline gelmiştir.

Üçüncü veri kümesi (DS-3) ise üç organizmadan oluşan; insan, sığır (*Bos Taurus*), fare (*Mus musculus*), GSE18290 erişim numarasına sahip bir veri kümesidir. İnsana ait veri kümesinde 54675 gen, 17 örneklem, sığır veri kümesinde 24128 gen, 17 örneklem, farenin veri kümesinde 22690 gen ve 17 örneklem bulunmaktadır. Memelilerin erken gelişimleri, embriyo hücrelerinin regulasyon (gen ifadesinin düzenlenmesi) ağları ile kontrol edilmektedir. DS-3 veri kümesi bu bilgiden yola çıkarak erken embriyo gelişim evrelerinde yapılan ölçümlerin dinamik gen ifade verilerinin örüntüleri ile ilgili önemli bilgi sağlayabileceği görüşü üzerine oluşturulmuş bir veri kümesidir [46]. Deney ölçümleri farklı embriyo gelişim aşamalarında yapılmıştır. Sığırın gen ifade verisi ölçümleri oosit (olgunlaşmamış dişi gamet), tek iki, dört, sekiz ve onaltı hücre aşamalarında, morula ve blastokist (embriyo gelişimi süresince embriyonik hücrelerin oluşturduğu ortası sıvı ile dolu hücresel yapı) evrelerinde yapılmıştır. İnsan ve fare ölçümleri de tek, iki, dört, sekiz hücre aşamaları, morula ve blastokist evrelerinde alınmıştır. Eleme sonrasında DS-3, 1182 gende, 6 örneklemden, 3 organizmadan oluşan, kayıp değer içermeyen bir veri kümesi haline almıştır.

6. SONUÇLAR

TriClustering algoritması farklı parametreler ile üç farklı veri kümesine uygulanmıştır. Deneysel deneme yöntemiyle belirlenen parametre değerleri α (çoklu silme işlemi için sınır parametresi) ve δ (MSR değeri için sınır parametresi) 0.3, 0.7 ve 1 olacak şekilde ayarlanmıştır. Her bir parametrenin değer değişikliğinin sonucu nasıl etkilediğini gözlemlemek amacıyla parametrelere değişik değerler atanarak işlem yapılmıştır. Parametre değişikliğinde α parametresi sabit iken, δ 'nın değerinin azaltılması (0.1 azaltma) ve δ parametresi sabit iken, α 'nın azaltılması (0.1 azaltma) yöntemi izlenmiştir. Ayrıca satır silme limiti (silme işleminden sonra geriye kalan minimum satır sayısı) olan n_r , sütun silme limiti olan n_c ve kaynak silme limiti olan n_s değerleri sırasıyla 10, 2, 2 olacak şekilde belirlenmiştir.

Daha öncede belirtildiği gibi üçlü kümelerin boyutları ve MSR değeri üçlü kümelerin uyumluluğunu ölçen değerlerdir. Bunun yanında bölüm 1.1.8'de anlatılan p değerinin bir hesaplama türü olan Phyper fonksiyonu, GO terimlerinin zenginleşme analizinde kullanılan istatistiksel bir testtir. Phyper parametresi ise elde edilen üçlü kümedeki genlerin uyumunu test etmede kullanılan bir parametredir. Her bir gen grubu için, maksimum değere sahip Phyper değeri tespit edilerek PhyperMax ismiyle kullanılmaktadır.

Algoritma parametreleri, MSR ve PhyperMax değeri arasında bir ilişkinin olup olmadığını anlamak için diğer parametreler sabit tutulurken α ve δ parametrelerinin değerleri arttırılmıştır. Üç veri kümesi üzerinde yapılan tüm deneylerin sonuçları Ekler bölümünde verilmiştir. Ayrıca α ve δ değerleri ile sonuçların ilişkisi Çizelge 6.1' de özetlenmektedir. Sonuçlara göre α 'nın artması MSR veya PhyperMax değerlerini etkilemezken, δ değerinin artması PhyperMax değerinin artmasına sebep olmaktadır. Buna ek olarak, gen gruplarının istatistiksel olarak anlamlandırılması ise p değerinin hesaplanması ile yapılmıştır. Gen gruplarının istatistiksel olarak doğrulanması, o gruptaki genlerin belirlenen gen açıklamaları ile ne kadar uyumlu olduğunun bir göstergesidir. İstatistiksel doğrulama için BioConductor'dan [47] ilgili yazılım paketlerinin fonksiyonları kullanılmıştır.

Çizelge 6. 1 Parametrelerin değişimi

Üç veri kümesinin MSR ve PhyperMax değerleri	DS-1		DS-2		DS-3	
	MSR	Phyper Max	MSR	Phyper Max	MSR	Phyper Max
α ↑	↓	↓	↑	↑	↓	↑
δ ↔						
α ↔	↓	↑	↑	↑	↑	↑
δ ↑						

Bu paketler aracılığı ile elde edilen gen gruplarında hangi GO terimlerinin zenginleştikleri tespit edilmiştir.

'GOHyperGALL' [48] fonksiyonu, tüm GO düğümleri için hipergeometrik dağılım testini kullanarak bir Bonferroni doğrulanmış p değeri hesaplamaktadır. Daha sonra Bonferroni doğrulanmış hipergeometrik testi kullanılarak ve p değeri için 0.01 limiti baz alınarak (0.01 den küçük olan p değerleri alınarak), elde edilen tüm üçlü kümelerin GO terimlerinin fonksiyonel zenginleşme durumları tespit edilmiştir. Tüm veri kümeleri için GO terimi zenginleşen grupların yüzdeleri Çizelge 6.2' de verilmiştir. Örneğin DS-1 için 15/21 oranının verilmesi, elde edilen 21 gen grubunun 15'inin biyolojik olarak anlamlı olduğunu, bu genlerin benzer biyolojik görevlerde yer aldıklarını göstermektedir. Çizelge 6.2' ye göre geliştirilen algoritma farklı veri kümelerinden biyolojik olarak anlamlı gen grupları elde etmede başarılı bir algoritmadır. Gen gruplarının biyolojik olarak anlamlı olduğunu göstermeye yarayan GO analizinin bir örneği de Çizelge 6.3' de verilmiştir. Bu çizelgede anlamlı olan bazı gen gruplarının birlikte yer aldıkları biyolojik işlemde GO id'si, gen grubunun istatistiksel olarak anlamını temsil eden p değeri, bu işlemin tanımı ve dahil olduğu ontoloji (Biological process-BP vb) verilmektedir. Örneğin çizelgenin ilk satırında çok hücreli organizmaların işlevinde (multicellular organismal process) 152 tane genin yer aldığı ve bu grubun istatistiksel anlamının 6.3E-23 olan Phyper

değeri ile verildiği görülmektedir. Bu değerlendirme sonuçlarını elde etmede daha öncede belirtildiği gibi p değeri için 0.01 limiti belirlenmiş, bu değerün üstünde p değerine sahip olan gen grupları analiz sonuçlarına dahil edilmemiştir. Sonuçlardan açıkça görüleceği üzere geliştirilen TriClustering algoritması biyolojik olarak anlamlı gen grupları bulmada başarılı bir yöntemdir.

Çizelge 6. 2 GO terimi zenginleşen gen gruplarının veri kümelerine göre yüzdesi

GO terimi zenginleşen gen gruplarının yüzdesi	DS-1	DS-2	DS-3
$\alpha = 0.3$	32/39	32/50	21/21
$\delta = 0.2$	%82.05	%64	%100

Elde edilen sonuçların istatistiksel doğrulanmasında kullanılan bir diğer yöntem Dunn indeksidir. Dunn indeksi yapılan kümelemenin başarılı olup olmadığını test etmede kullanılan bir ölçüm indeksidir. Bir başka deyişle, bu indeks kümelerin sıklığını (küme elemanlarının varyans değerinin küçük olması) ve ayrıklığını (kümeler arasındaki mesafenin minimum olması) test etmede kullanılmaktadır. Büyük Dunn indeks değeri daha iyi bir kümelemeyi temsil etmektedir. Geliştirilen üç-yönlü kümeleme (TriClustering) yönteminin iki-yönlü kümeleme yöntemlerinden daha iyi olduğunu gösterebilmek için, tek organizmalı veri kümesinden elde edilen kümelerin Dunn indeksi ile çok organizmalı veri kümesinden elde edilen kümelerin Dunn indeks değerleri kıyaslanmıştır. Çizelge 6.4' ten de görüleceği üzere karşılaştırılan 7 durumdan (kırmızı renk ile gösterilmiştir) 5 tanesinde çok organizmalı veri kümesinin Dunn indeks değeri daha yüksek çıkmıştır. Bu da çok organizmalı analiz yöntemlerinin tek organizmalı, tek yönlü analiz yöntemlerine kıyasla daha başarılı olduklarının bir göstergesidir.

Çizelge 6. 3 Örnek bir gen grubunun GO analiz sonuçları

GO ID	Phyper	Term	Ont.
GO:0032501	6.39735E-23	Multicellular organismal process	BP
GO:0065008	5.34575E-22	Regulation of biological quality	BP
GO:0010033	6.78266E-22	Response to organic substance	BP
GO:0003008	5.97546E-20	System process	BP
GO:0042221	2.80595E-19	Response to chemical stimulus	BP
GO:0009719	2.02422E-16	Response to endogenous stimulus	BP
GO:0048856	2.97324E-16	Anatomical structure development	BP
GO:0023052	3.34414E-16	Signaling	BP
GO:0050896	3.37298E-16	Response to stimulus	BP
GO:0009725	4.85337E-16	Response to hormone stimulus	BP

Çizelge 6. 4 Dunn indeks sonuçları

Veri kümesi	Organizma	Dunn İndeks değeri	
		Tek Org	Çoklu Org
DS-1	İnsan	0.0467	0.0371
DS-1	Fare	0.0432	0.0462
DS-2	İnsan	0.0059	0.0306
DS-2	Fare	0.0053	0.0268
DS-3	İnsan	0.0109	0.0104
DS-3	Fare	0.0054	0.0129
DS-3	Sığır	0.0038	0.0070

7. TARTIŞMA VE GELECEK ÇALIŞMALAR

Bu tezde üç yönlü yeni bir kümeleme yöntemi olan TriClustering yöntemi ve yazılım aracı geliştirilmiştir. Farklı parametrelere bağlı üç yönlü kümelemenin yapılabileceği bir yöntem olan TriClustering, türler arası benzerlikleri, hastalıkları tespit etmede kullanılabilecek bir analiz yöntemidir. TriClustering yöntemi organizmayı temsil eden üçüncü boyutun gen matrisine eklenmesiyle oluşan üç boyutlu veri kümesinin analizi için geliştirilmiş, iki yönlü kümeleme yönteminin uzantısı olan bir yöntemdir.

TriClustering algoritması, iki yönlü kümelemede sıkça kullanılan Cheng ve Church (CC)'un, iteratif olarak satır/sütun silmeye/eklemeye dayalı ikili kümeleme tekniğinin üç boyutlu gen matrisi için uyarlanması ile geliştirilmiştir. Geliştirilen yöntem birbirleriyle uyumlu, üçlü küme olarak adlandırılan grupların üç boyutlu (gen-durum-organizma) gen matrisinde analiz edilmesini amaçlamaktadır. Yazılım aracı ise TriClustering algoritmasının farklı veri kümelerine uygulanabilmesini, verilerin, sonuçların ısı haritaları ve GO terimi zenginleştirme istatistikleriyle görselleştirilmesini sağlamaktadır. Yazılım aracı ayrıca elde edilen üçlü kümeler içerisinde gen aramasının yapılmasına olanak sağlamaktadır.

Geliştirilen yöntem NCBI GEO veri merkezinden alınan üç farklı, çok organizmalı gen ifade verisine uygulanmıştır GO terimi zenginleştirme analizi ve Dunn indeksi kullanılarak sonuçların biyolojik ve istatistiksel anlamlandırılması gerçekleştirilmiştir. Deneysel sonuçlar, geliştirilen yöntemin biyolojik olarak anlamlı olan üçlü kümeler bulabildiğini, yöntemin ve yazılım aracının gen ifade verisi kullanılarak türler arası analizde kullanılabileceğini göstermektedir. İstatistiksel sonuçlar ise çok organizmalı veri kümesine üçlü yönlü kümeleme yönteminin uygulanmasının, tek organizmalı veri kümesine iki yönlü kümeleme yönteminin uygulamasına kıyasla daha iyi kümeleme sonuçları verdiğini göstermektedir.

Yapılan çalışma kullandığı yöntemler açısından makine öğrenme, istatistiksel veri analizi ve paralel veri işleme kapsamında sayılabilir. Projenin öne çıkan özelliği az sayıda deneye karşılık çok sayıda ölçüm içeren veri kümelerinde daha etkin analiz amacıyla farklı türdeki ilişkili verinin güçlü bir altyapı üzerinde bütünleştirilmesidir.

Geliştirilen yöntem ve yazılım aracı pek çok hücresel döngü ve mekanizmanın belirlenmesine yardımcı olacak ve bu yöntemlerin kullanımını sağlayacak görsel web araçları, kanser arařtırmaları ve ilaç tasarımı gibi biyotıp çalışmalarına katkı sağlayacaktır. Yazılım aracı özellikle biyologların genomik veri analizi, türler arası benzerliklerin arařtırılması ve çeřitli hastalıkların tespiti vb. çalışmalarında işlerini kolaylařtırıcı bir araç haline gelecektir.

Veri tabanlarında çok organizmalı deneylerin sayısı arttıkça daha fazla karmaşık hesaplamalara dayalı analiz yöntemlerine ihtiyaç duyulacaktır. Böylece bu ve benzeri çalışmaların daha fazla önem kazanacağı düşünölmektedir.

Gelecek çalışmalar için ise geliştirilen yöntem, benzer veri tipleri üzerinde tanımlı farklı problemler için de uygulanabilir olacaktır.

KAYNAKLAR LİSTESİ

- [1] SCHENA, Mark, SHALON, Dari, DAVIS, Ronald and BROWN, Patrick. Quantitative monitoring of gene expression patterns with a complementary DNA microarray, American Association for the Advancement of Science, Vol.270, s.467-470, 1995.
- [2] LOCKHART, D. et al. Expression monitoring by hybridization to high-density oligonucleotide arrays, Nat. Biotechnol, vol.14, s.1675–1680, 1996.
- [3] JIANG, Daxin, TANG, Chun, ZHANG, Aidong, Cluster Analysis for gene expression data: a survey, Knowledge and Data Engineering, IEEE Transaction on.16, no.11, s.1370-1386, 2004.
- [4] http://en.wikipedia.org/wiki/Cluster_analysis, 10.06.13.
- [5] <http://www.bilgisayarkavramlari.com/2008/12/15/k-ortalama-algoritmasi-k-means-algorithm/>
- [6] CHENG, Yizong and CHURCH, George M., Biclustering of expression data, Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology (ISMB), 8:93–103, 2000.
- [7] HARTIGAN, J.A, Direct Clustering of a Data Matrix, Journal of the American Statistical Association, Vol.67, s.123-129, 1972.
- [8] MADEIRA, Sara, OLIVEIRA, Arlindo L, Biclustering Algorithms for Biological Data Analysis: A Survey, IEEE Transactions On Computational Biology And Bioinformatics, Vol.1, No.1, s.24-45, 2004.
- [9] TANAY, Amos, SHARAN, Roded, SHAMIR, Ron, Biclustering Algorithms: A Survey, Handbook of computational molecular biology, 2005.
- [10] <http://www.geneontology.org/GO.doc.shtml>, 05.06.13.
- [11] Using the Gene Ontology (GO) for analysis of expression data, Jane Lomax EMBL-EBI, Slayt 30.
- [12] BAGGERLY, Keith A., COOMBES, Kevin R., MORRIS, Jeffrey S., An Introduction to High-Throughput Bioinformatics Data, Bayesian Inference in

Gene Expression and Proteomics. Ed. KA Do, P Mueller, M Vannucci. New York: Cambridge University Press, 1-33, 2006.

- [13] EISEN, Michael B., SPELLMAN, Paul T., BROWN, Patrick O., BOTSTEIN, David, Cluster Analysis and display of genome-wide expression patterns, *Genetics*, Vol.95, s.14863-14868, 1998.
- [14] MIRKIN, Boris, *Mathematical Classification and Clustering*, Kluwer Academic Publishers, s.144, 1996.
- [15] BERGMAN, Sven, IHMELS, Jan, BARKAI, Naama, Iterative Signature Algorithm for the analysis of large-scale gene expression data, *Physical review E*, vol. 67(3), 031902, s.28-39, 2009.
- [16] GREMALSCHI, Stefan, ALTUN, Gulsah, Mean Squared Residue Based Biclustering Algorithms, *Proceedings of International Symposium on Bioinformatics Research and Applications (ISBRA'08)*, Springer LNBI (Lecture Notes in Computer Science) 4983, s. 232-243, 2008.
- [17] GREMALSCHI, Stefan, ALTUN, Gulsah, ASTROVSKAYA, Irina, ZELIKOVSKY, Alexander, Mean Square Residue Biclustering with Missing Data and Row Inversions., *Bioinformatics Research and Applications Lecture Notes in Computer Science*, Vol.5542, s.28-39, 2009.
- [18] ANGIULLI, Fabrizio, PIZZUTI, Clara, Gene Expression Biclustering using Random Walk Strategies, *Data Warehousing and Knowledge Discovery Lecture Notes in Computer Science* Vol. 3589, s.509-519, 2005.
- [19] BEN-DOR, Amir, CHOR, Benny, KARP, Richard, YAKHINI, Zohar Discovering Local Structure in Gene Expression Data: The Order-Preserving Submatrix Problem, *Journal Of Computational Biology*, Vol.10, s-373-384, 2003.
- [20] LAZZERONI, Laura, OWEN, Art, Plaid Models For Gene Expression Data, *Statistica Sinica*, s.61-86, 12, 2002.

- [21] GETZ, Gad; LEVINE, Erel; DOMANY, Eytan. Coupled two-way clustering analysis of gene microarray data, *Proceedings of the National Academy of Sciences*, vol.97., s.12079-12084, 2000.
- [22] CALDAS, Jose, KASKI, Samuel, Hierarchical Generative Biclustering for MicroRNA Expression Analysis, *Journal of Computational Biology*, Vol.18, No.3, s.251-261, 2011.
- [23] BARKOW, S, BLEULER, S., PRELIC, A., ZIMMERMANN, P., and E. ZITZLER, E, BicAT: a biclustering analysis toolbox, *Bioinformatics*, vol.22(10), s.1282-1283, 2006.
- [24] SANTAMARIA, R., THERÓN, R. and QUINTALES, L. BicOverlapper: A tool for bicluster visualization, *Bioinformatics*, vol.24, s.1212-1213 , 2008.
- [25] ERNST, Jason and BAR-JOSEPH, Ziv, STEM: a tool for the analysis of short time series gene expression data, *BMC Bioinformatics*, Vol.7, 2006.
- [26] SHANNON, Paul, MARKIEL, Andrew, OZIER, Owen, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Research*, vol.13, s.2498-2504, 2003.
- [27] MAERE, Steven, HEYMANS, Karel and KUIPER, Martin, BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks, *Bioinformatics*, Vol.21, s.3448-3449, 2005.
- [28] ACAR, Evrim, YENER, Bülent, Unsupervised Multiway Data Analysis: A Literature Survey, *IEEE Transactions On Knowledge And Data Engineering*, Vol. 21, No. 1,s.6-20, 2009.
- [29] HUOPANIEMI, Ilkka, SUVITAIVAL, Tommi, NIKKILA, Janne, OREŠIĆ, Matej, KASKI Samuel, Multivariate multi-way analysis of multi-source data, *Bioinformatics*, Vol.26,s.391-398, 2010.
- [30] VERMUNT, J.K, A hierarchical mixture model for clustering three-way data sets, *Computational Statistics & Data Analysis*, Vol.5, s.5368-5376, 2007.
- [31] ZHAO, Lizhuang, ZAKI, Muhammed J.,TRICLUSTER: An Effective Algorithm for Mining Coherent Clusters in 3D Microarray Data, *Proceedings of the 2005*

ACM SIGMOD International Conference on Management of data, s.694-705, 2005.

- [32] TCHAGANG, A.B et al., Mining biological information from 3D short time-series gene expression data: the OPTricluster algorithm, *BMC Bioinformatics*, vol.13, s.54, 2012.
- [33] WILKINS, A.S. *The Evolution of Developmental Pathways*. Sinauer Associates, Sunderland, MA, USA, 2001.
- [34] LU, Yong, HUGGINGS, Peter and JOSEPH, Ziv-Bar, Cross-species analysis of microarray expression data, *Bioinformatics*, vol.25, s.1476-1483, 2009.
- [35] STARK, Alexander, BRENNECKE, Julius, RUSSELL, Robert B., and COHEN, Stephen M., Identification of *Drosophila* microRNA targets, *PLoS biology* 1, no. 3 : e60., 2003.
- [36] ODOM, Duncan. et al., Tissue-specific transcriptional regulation has diverged, significantly between human and mouse, *Nature Genetics*, vol.39, s.730–732, 2007.
- [37] SHARAN, R. et al. Cover article: conserved patterns of protein interaction in multiple species. *Proc. Natl Acad. Sci. USA*, vol.102(6), s.1974–1979, 2005.
- [38] HARBISON, C.T. et al., Transcriptional regulatory code of a eukaryotic genome, *Nature*, vol.431, s.99–104, 2004.
- [39] KROGAN, N.J. et al., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*, *Nature*, vol.440, s.637–643, 2006.
- [40] ALEXANDER, P.A. et al., The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proceedings of the National Academy of Sciences*, vol.104, s.11963–11968, 2007.
- [41] ALTSCHUL, S.F. et al. Basic local alignment search tool, *J. Mol. Biol.*, 215, s.403–410, 1990.

- [42] DURBIN,R. et al. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, UK, 1998.
- [43] KAMVYSSELIS, M. et al., Whole-genome comparative annotation and regulatory motif discovery in multiple yeast species, In Proceedings of the Seventh International Conference on Research in Computational Molecular Biology, ACM Press, s.157–166, 2003.
- [44] <http://www.ncbi.nlm.nih.gov/geo/>, 06.10.12.
- [45] BLACK, Michael B. et al., Cross-species Comparisons of Transcriptomic Alterations in Human and Rat Primary Hepatocytes Exposed to 2,3,7,8-Tetrachlorodibenzo-p-dioxin, Toxicological Sciences, vol.127, s.199-215, 2012.
- [46] XIE, Dan, CHEN, Chieh-Chun, PTASZEK, Leon M. et al., Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species, Genome Research, vol 20., s.804-815, 2010.
- [47] http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual#biocon_intro, 02.02.13
- [48] http://faculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts/GOHyperGAll.txt , 02.02.13.

EKLER LİSTESİ

Bu bölümde her bir veri kümesinin organizmalarına tek ve hepsi bir aradayken Triclustering algoritmasının uygulanmasının sonuçları verilmiştir. Her bir deneyde tüm üçlü kümeler için, algoritma parametreleri (alfa,delta) ve iterasyon sayıları değiştirilerek elde edilen GO terimi sayıları, pHyperMax değeri ve MSR değerleri listelenmiştir. Bu tablolar tek organizmalı ve çok organizmalı analiz sonuçlarının kıyaslanmasını kolaylaştırmaktadır.

ÜK = Üçlü Küme

DS-1 veri kümesinin insan organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-1 1 organizma İnsan	GO terimi zenginleşen ÜK sayısı	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK1 0	ÜK1 1	ÜK1 2	ÜK13	ÜK1 4	ÜK1 5
alfa = 0.3	20	GO terimi sayısı	381	205	110	0	84	17	140	2	13	1	3	0	65	3	3
delta = 0.2		pHyper max değeri	0.01	0.01	0.01		0.01	0.01	0.01	0.00	0.01	0.00	0.01		0.01	0.01	0.01
ite=23		gen X durum X org	3760 9 1	1389 5 1	583 3 1	48 8 1	1244 9 1	603 3 1	361 3 1	416 9 1	138 8 1	54 2 1	100 4 1	11 7 1	190 11 1	19 6 1	34 10 1
		MSR değeri	0.271 3	0.285 5	0.07 16	0.37 97	0.414 4	0.11 74	0.18 27	0.41 35	0.38 71	0.04 23	0.44 38	0.43 83	0.473 0	0.66 41	0.70 50

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ORT
0	9	9	3	2	4	8	3	46.30435
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
10 7 1	41 9 1	71 11 1	19 7 1	10 10 1	10 10 1	10 11 1	10 11 1	
0.5056	0.7306	0.7426	0.7427	0.7466	0.7794	0.8350	0.7901	0.4853

alfa = 0.7	20	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15
delta = 0.2		GO terimi sayısı	10	194	12	497	46	125	1	0	19	100	3	3	17	4	4
ite = 21		pHyper max değeri	8E-03	1E-02	7E-03	1E-02	9E-03	1E-02	8E-03		6E-03	1E-02	8E-03	5E-03	8E-03	8E-03	9E-03
		gen X durum X org	240 3 1	1466 6 1	257 4 1	2922 2 1	365 4 1	1706 3 1	56 5 1	106 4 1	412 2 1	866 3 1	128 4 1	53 4 1	250 2 1	91 2 1	13 4 1
		MSR değeri	0.2647	0.1346	0.1223	0.2479	0.0947	0.1452	0.145 1	0.1590	0.0499	0.2828	0.1083	0.188 0	0.2113	0.124 1	0.236 3

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
12	1	1	1	9	1	50.47619
9E-03	9E-03	6E-03	8E-03	1E-02	7E-03	8E-03
135 2 1	16 3 1	11 5 1	17 4 1	13 3 1	8 2 1	
0.2800	0.1853	0.2222	0.4531	0.3645	0.3717	0.2091

alfa = 1	7	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ORT
delta = 0.2		GO terimi sayısı	521	391	208	9	9	11	1	164.2857
ite = 7		pHyper max değeri	1E-02	1E-02	1E-02	8E-03	7E-03	7E-03	6E-03	8E-03
		gen X durum X org	5683 7 1	2191 7 1	1040 4 1	137 6 1	56 6 1	16 6 1	8 4 1	
		MSR değeri	4E-01	3E-01	4E-01	5E-01	5E-01	4E-01	3E-01	4E-01

alfa = 0.7	20	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12
delta = 0.3		GO terimi sayısı	10	194	12	497	46	125	1	0	19	100	3	3
ite = 21		pHyper max değeri	0.008	0.010	0.007	0.010	0.009	0.010	0.008		0.006	0.010	0.008	0.005
		gen X durum X org	240 3 1	1466 6 1	257 4 1	2922 2 1	365 4 1	1706 3 1	56 5 1	106 4 1	412 2 1	866 3 1	128 4 1	53 4 1
		MSR değeri	0.2647	0.1346	0.1223	0.2479	0.0947	0.1452	0.1451	0.1590	0.0499	0.2828	0.1083	0.1880

ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
17	4	4	12	1	1	1	9	1	50.47619
0.008	0.008	0.009	0.009	0.009	0.006	0.008	0.010	0.007	0.008
250 2 1	91 2 1	13 4 1	135 2 1	16 3 1	11 5 1	17 4 1	13 3 1	8 2 1	
0.2113	0.1241	0.2363	0.2800	0.1853	0.2222	0.4531	0.3645	0.3717	0.2091

alfa = 0.7	20	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12
delta = 0.7		GO terimi sayısı	10	194	12	497	46	125	1	0	19	100	3	3
ite = 21		pHyper max değeri	0.008	0.010	0.007	0.010	0.009	0.010	0.008		0.006	0.010	0.008	0.005
		gen X durum X org	240 3 1	1466 6 1	257 4 1	2922 2 1	365 4 1	1706 3 1	56 5 1	106 4 1	412 2 1	866 3 1	128 4 1	53 4 1
		MSR değeri	0.2647	0.1346	0.1223	0.2479	0.0947	0.1452	0.1451	0.1590	0.0499	0.2828	0.1083	0.1880

ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
17	4	4	12	1	1	1	9	1	50.47619
0.008	0.008	0.009	0.009	0.009	0.006	0.008	0.010	0.007	0.008
250 2 1	91 2 1	13 4 1	135 2 1	16 3 1	11 5 1	17 4 1	13 3 1	8 2 1	
0.2113	0.1241	0.2363	0.2800	0.1853	0.2222	0.4531	0.3645	0.3717	0.2091

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	917		
ite = 2		pHyper max değeri	9.95E-03		
		gen X durum X org	9131 11 1	0 11 1	
		MSR değeri	0.87445		

DS-1 veri kümesinin fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-1 1 organizma Fare	GO terimi zenginleşen ÜK sayısı	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11
alfa = 0.3	18	GO terimi sayısı	154	328	41	137	33	42	21	159	5	0	1
delta = 0.2		pHyper max değeri	0.010	0.010	0.010	0.010	0.009	0.005	0.010	0.009	0.009		0.003
ite = 21		gen X durum X org	2378 6 1	1990 8 1	602 5 1	1299 4 1	518 8 1	897 2 1	14 3 1	966 2 1	183 10 1	43 8 1	48 8 1
		MSR değeri	0.311	0.482	0.364	0.402	0.391	0.310	0.404	0.415	0.339	0.525	0.559

ÜK12	ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
3	19	2	0	9	3	1	0	1	5	45.90476
0.009	0.010	0.010		0.009	0.002	0.007		0.008	0.009	0.008
61 10 1	26 10 1	10 11 1	21 10 1	18 6 1	17 3 1	11 7 1	10 6 1	10 11 1	9 11 1	
0.487	0.672	0.837	0.759	0.607	0.407	0.698	0.605	0.851	0.810	0.535

DS-1 1 organizma Fare	GO terimi zenginleşen ÜK sayısı	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11
alfa = 0.3	18	GO terimi sayısı	154	328	41	137	33	42	21	159	5	0	1
delta = 0.2		pHyper max değeri	0.010	0.010	0.010	0.010	0.009	0.005	0.010	0.009	0.009		0.003
ite = 21		gen X durum X org	2378 6 1	1990 8 1	602 5 1	1299 4 1	518 8 1	897 2 1	14 3 1	966 2 1	183 10 1	43 8 1	48 8 1
		MSR değeri	0.311	0.482	0.364	0.402	0.391	0.310	0.404	0.415	0.339	0.525	0.559

ÜK12	ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
3	19	2	0	9	3	1	0	1	5	45.90476
0.009	0.010	0.010		0.009	0.002	0.007		0.008	0.009	0.008
61 10 1	26 10 1	10 11 1	21 10 1	18 6 1	17 3 1	11 7 1	10 6 1	10 11 1	9 11 1	
0.487	0.672	0.837	0.759	0.607	0.407	0.698	0.605	0.851	0.810	0.535

alfa = 1	4	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ORT
delta = 0.2		GO terimi sayısı	866	150	45	12	0	0	0	153.2857
ite = 7		pHyper max değeri	1E-02	1E-02	1E-02	8E-03				9E-03
		gen X durum X org	5681 7 1	2149 6 1	780 7 1	386 5 1	77 5 1	34 5 1	15 5 1	
		MSR değeri	0.434	0.112	0.395	0.450	0.401	0.490	0.401	0.383

alfa = 0.7	17	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12
delta = 0.3		GO terimi sayısı	123	313	44	152	33	99	17	98	3	5	5	7
ite = 21		pHyper max değeri	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		gen X durum X org	1633 4 1	2992 2 1	734 4 1	1395 2 1	330 2 1	556 3 1	51 10 1	408 3 1	25 4 1	165 4 1	106 4 1	365 2 1
		MSR değeri	0.173	0.221	0.214	0.244	0.097	0.182	0.059	0.226	0.280	0.473	0.209	0.559

ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
0	0	2	0	0	4	1	3	2	43.38095
		0.00			0.00	0.01	0.00	0.01	0.01
49 4 1	53 3 1	104 2 1	73 2 1	41 6 1	16 6 1	11 4 1	11 2 1	12 2 1	
0.064	0.334	0.231	0.268	0.110	0.096	0.412	0.145	0.222	0.229

alfa = 0.7	17	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12
delta = 0.3		GO terimi sayısı	123	313	44	152	33	99	17	98	3	5	5	7
ite = 21		pHyper max değeri	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		gen X durum X org	1633 4 1	2992 2 1	734 4 1	1395 2 1	330 2 1	556 3 1	51 10 1	408 3 1	25 4 1	165 4 1	106 4 1	365 2 1
		MSR değeri	0.173	0.221	0.214	0.244	0.097	0.182	0.059	0.226	0.280	0.473	0.209	0.559

ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
0	0	2	0	0	4	1	3	2	43.38095
		0.00			0.00	0.01	0.00	0.01	0.01
49 4 1	53 3 1	104 2 1	73 2 1	41 6 1	16 6 1	11 4 1	11 2 1	12 2 1	
0.064	0.334	0.231	0.268	0.110	0.096	0.412	0.145	0.222	0.229

DS-1 veri kümesinin insan ve fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-1 2 Organizma	GO terimi zenginleşen ÜK sayısı																
alfa = 0.3	32	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15
delta = 0.2		GO terimi sayısı	257	138	440	11	6	2	2	7	17	18	9	0	1	0	1
ite = 39		pHyper max değeri	0.00963	0.00984	0.00958	0.00948	0.00974	0.00941	0.00801	0.00764	0.00798	0.00901	0.00831		0.00976		0.00727
		gen X durum X org	225492	207092	2915112	10112	32292	10112	41522	39722	142112	10112	10112	154102	64102	28112	10112
	MSR değeri	0.55660	0.63574	0.69506	0.86825	0.73804	0.86462	0.25654	0.27647	0.88889	0.87746	0.89334	0.77482	0.68533	0.89247	0.88633	

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34
22	22	8	4	4	0	3	0	1	0	20	1	2	4	0	0	2	8	4
0.00986	0.00728	0.00937	0.00860	0.00885		0.00984		0.00394		0.00995	0.00989	0.00498	0.00569			0.00644	0.00940	0.00854
10112	10112	10112	10112	10112	10112	47102	10112	10112	10112	10112	43112	10112	10112	10112	10112	10112	10112	16102
0.81197	0.88977	0.86042	0.89122	0.87896	0.86945	0.73466	0.83532	0.88138	0.86376	0.87383	0.89544	0.89016	0.87954	0.89541	0.88202	0.89714	0.89084	0.82507

ÜK35	ÜK36	ÜK37	ÜK38	ÜK39	ORT
9	1	14	7	1	26.820513
0.00661	0.00369	0.00879	0.00986	0.00571	0.00822
10112	1482	10112	10112	10112	
0.88926	0.61487	0.90334	0.85951	0.86546	0.80433

alfa = 0.7	18	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14
delta = 0.2		GO terimi sayısı	145	338	107	11	55	16	19	29	1	10	0	27	27	1
ite = 22		pHyper max değeri	9.97E-03	9.94E-03	9.95E-03	9.01E-03	9.97E-03	0.0056725	0.0091565	0.0097692	0.0057583	0.0084223		0.009089	0.0097779	0.005518
		gen X durum X org	1764 3 2	3059 2 2	1124 2 2	756 2 2	833 2 2	221 2 2	303 2 2	249 2 2	133 2 2	145 2 2	65 2 2	111 2 2	136 2 2	53 3 2
		MSR değeri	0.253	0.275	0.213	0.230	0.206	0.372	0.462	0.193	0.239	0.244	0.216	0.357	0.382	0.496

ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ORT
0	2	1	1	0	0	8	4	36.454545
	0.0087772	0.0082487	0.0090059				0.0093346	0.0086688
49 7 2	29 3 2	37 8 2	20 3 2	18 2 2	12 9 2	10 2 2	4 2 2	
0.629	0.516	0.699	0.460	0.669	0.855	0.497	0.323	0.399

alfa = 1	7	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ORT
delta = 0.2		GO terimi sayısı	690	260	63	8	6	1	1	1	128.75
ite = 8		pHyper max değeri	9.98E-03	9.73E-03	9.98E-03	2.93E-03	9.60E-03	7.53E-03	5.09E-03	9.10E-03	7.99E-03
		gen X durum X org	5303 6 2	1725 6 2	1457 7 2	390 6 2	164 6 2	62 6 2	22 4 2	8 5 2	
		MSR değeri	0.525	0.559	0.654	0.526	0.682	0.649	0.630	0.626	0.607

alfa = 0.7	18	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13
delta = 0.3		GO terimi sayısı	145	338	107	11	55	16	19	29	1	10	0	27	27
ite = 22		pHyper max değeri	0.010	0.010	0.010	0.009	0.010	0.006	0.009	0.010	0.006	0.008		0.009	0.010
		gen X durum X org	1764 3 2	3059 2 2	1124 2 2	756 2 2	833 2 2	221 2 2	303 2 2	249 2 2	133 2 2	145 2 2	65 2 2	111 2 2	136 2 2
		MSR değeri	0.253	0.275	0.213	0.230	0.206	0.372	0.462	0.193	0.239	0.244	0.216	0.357	0.382

ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ORT
1	0	2	1	1	0	0	8	4	36.454545
0.006		0.009	0.008	0.009			0.008	0.009	0.009
53 3 2	49 7 2	29 3 2	37 8 2	20 3 2	18 2 2	12 9 2	10 2 2	4 2 2	
0.496	0.629	0.516	0.699	0.460	0.669	0.855	0.497	0.323	0.399

alfa = 0.7	18	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13
delta = 0.7		GO terimi sayısı	145	338	107	11	55	16	19	29	1	10	0	27	27
ite = 22		pHyper max değeri	0.010	0.010	0.010	0.009	0.010	0.006	0.009	0.010	0.006	0.008		0.009	0.010
		gen X durum X org	1764 3 2	3059 2 2	1124 2 2	756 2 2	833 2 2	221 2 2	303 2 2	249 2 2	133 2 2	145 2 2	65 2 2	111 2 2	136 2 2
	MSR değeri	0.253	0.275	0.213	0.230	0.206	0.372	0.462	0.193	0.239	0.244	0.216	0.357	0.382	

ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ORT
1	0	2	1	1	0	0	8	4	36.454545
0.006		0.009	0.008	0.009			0.008	0.009	0.009
53 3 2	49 7 2	29 3 2	37 8 2	20 3 2	18 2 2	12 9 2	10 2 2	4 2 2	
0.496	0.629	0.516	0.699	0.460	0.669	0.855	0.497	0.323	0.399

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	917		917
ite = 2		pHyper max değeri	9.95E-03		0.0099479
		gen X durum X org	9131 11 2	0 11 2	9131 11 2
	MSR değeri	0.898	NaN	0.898	

DS-2 veri kümesinin insan organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-2 1 Organizma İnsan	GO terimi zenginleşen ÜK sayısı	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14
alfa = 0.3	27	GO terimi sayısı	303	55	74	4	2	23	0	9	88	15	15	0	9	0
delta = 0.2		pHyper max değeri	0.009	0.009	0.010	0.003	0.006	0.010		0.007	0.010	0.009	0.010		0.010	
ite=38		gen X durum X org	4452 2 1	1079 2 1	1243 19 1	446 3 1	278 4 1	974 3 1	334 4 1	724 2 1	800 2 1	452 8 1	360 6 1	107 6 1	226 3 1	123 7 1
		MSR değeri	0.119	0.037	0.370	0.048	0.037	0.228	0.079	0.054	0.016	0.176	0.311	0.309	0.046	0.357

ÜK1 5	ÜK16	ÜK1 7	ÜK1 8	ÜK1 9	ÜK20	ÜK21	ÜK22	ÜK23	ÜK2 4	ÜK25	ÜK2 6	ÜK2 7	ÜK28	ÜK29	ÜK3 0	ÜK31	ÜK3 2	ÜK33	ÜK34	ÜK35	ÜK36	ÜK3 7	ÜK3 8	ORT
5	0	2	0	0	0	14	0	1	4	1	0	1	0	7	2	4	3	1	3	5	10	12	5	17.8157 89
0.00 1		0.00 8				0.005		0.005	0.00 7	0.007		0.00 9		0.009	0.00 8	0.007	0.00 7	0.006	0.007	0.010	0.009	0.00 9	0.00 7	0.008
41 2 1	126 3 1	10 2 1	22 6 1	36 3 1	120 4 1	164 4 1	131 13 1	102 2 1	70 3 1	105 3 1	50 3 1	47 7 1	41 19 1	70 25 1	16 4 1	18 16 1	45 7 1	14 40 1	17 34 1	13 27 1	13 23 1	14 2 1	5 4 1	
0.12 0	0.324	0.04 1	0.12 7	0.31 8	0.406	0.097	0.413	0.048	0.13 1	0.107	0.48 3	0.45 2	0.453	0.426	0.57 2	0.468	0.61 7	0.527	0.723	0.578	0.676	0.10 5	0.15 3	0.278

alfa = 0.7	2 8	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14
delta = 0.2		GO terimi sayısı	11	207	0	116	3	84	11	19	0	107	59	0	0	2
ite = 50		pHyper max değeri	1E-02	9E-03		1E-02	7E-03	9E-03	9E-03	9E-03		1E-02	1E-02			4E-03
		gen X durum X org	188 21 1	1821 16 1	23 22 1	1789 16 1	274 15 1	518 17 1	289 16 1	309 17 1	92 16 1	1172 20 1	1058 14 1	361 20 1	139 24 1	175 19 1
		MSR değeri	0.156	0.237	0.101	0.291	0.261	0.243	0.247	0.271	0.318	0.302	0.287	0.292	0.331	0.309

ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK2 4	ÜK2 5	ÜK2 6	ÜK27	ÜK28	ÜK29	ÜK3 0	ÜK3 1	ÜK32	ÜK3 3	ÜK3 4	ÜK3 5	ÜK3 6
5	0	1	12	3	0	1	0	0	14	8	3	10	10	0	0	0	7	2	0	4	1
5E-03		7E-03	5E-03	1E-02		4E-03			8E-03	9E-03	2E-03	8E-03	1E-02				8E-03	2E-03		7E-03	3E-03
197 15 1	100 19 1	380 17 1	105 15 1	177 16 1	211 14 1	135 23 1	182 17 1	117 20 1	21 19 1	40 21 1	34 16 1	233 12 1	198 13 1	266 15 1	29 19 1	65 16 1	305 12 1	25 20 1	49 15 1	55 16 1	213 8 1
0.288	0.367	0.369	0.331	0.360	0.270	0.330	0.329	0.337	0.335	0.318	0.315	0.281	0.295	0.390	0.321	0.336	0.360	0.365	0.322	0.351	0.293

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
0	0	1	0	0	0	0	0	0	5	0	0	10	4	14.4
		6E-05							8E-03			1E-02	6E-03	7E-03
62 15 1	27 15 1	88 12 1	188 11 1	51 11 1	46 14 1	68 14 1	23 15 1	62 11 1	39 17 1	29 16 1	68 18 1	20 18 1	44 13 1	
0.339	0.309	0.311	0.375	0.277	0.386	0.376	0.385	0.341	0.412	0.288	0.319	0.382	0.356	0.315

alfa = 1	8	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ORT
delta = 0.2		GO terimi sayısı	391	476	172	5	1	3	0	6	0	3	105.7
ite = 10		pHyper max değeri	8.98E-03	9.98E-03	9.49E-03	3.23E-03	3.02E-03	2.44E-03		5.95E-03		5.37E-03	0.0060572
		gen X durum X org	8060 29 1	2563 30 1	1310 32 1	537 32 1	224 34 1	115 31 1	39 33 1	27 33 1	10 36 1	3 37 1	
	MSR değeri	0.497	0.531	0.671	0.554	0.404	0.673	0.629	0.586	0.574	0.384	0.550	

alfa = 0.7	28	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13
delta = 0.3		GO terimi sayısı	11	207	0	116	3	84	11	19	0	107	59	0	0
ite = 50		pHyper max değeri	1E-02	9E-03		1E-02	7E-03	9E-03	9E-03	9E-03		1E-02	1E-02		
		gen X durum X org	188 21 1	1821 16 1	23 22 1	1789 16 1	274 15 1	518 17 1	289 16 1	309 17 1	92 16 1	1172 20 1	1058 14 1	361 20 1	139 24 1
	MSR değeri	0.156	0.237	0.101	0.291	0.261	0.243	0.247	0.271	0.318	0.302	0.287	0.292	0.331	

ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36
2	5	0	1	12	3	0	1	0	0	14	8	3	10	10	0	0	0	7	2	0	4	1
4E-03	5E-03		7E-03	5E-03	1E-02		4E-03			8E-03	9E-03	2E-03	8E-03	1E-02				8E-03	2E-03		7E-03	3E-03
175 19 1	197 15 1	100 19 1	380 17 1	105 15 1	177 16 1	211 14 1	135 23 1	182 17 1	117 20 1	21 19 1	40 21 1	34 16 1	233 12 1	198 13 1	266 15 1	29 19 1	65 16 1	305 12 1	25 20 1	49 15 1	55 16 1	213 8 1
0.309	0.288	0.367	0.369	0.331	0.360	0.270	0.330	0.329	0.337	0.335	0.318	0.315	0.281	0.295	0.390	0.321	0.336	0.360	0.365	0.322	0.351	0.293

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
0	0	1	0	0	0	0	0	0	5	0	0	10	4	14.4
		6E-05							8E-03			1E-02	6E-03	7E-03
62 15 1	27 15 1	88 12 1	188 11 1	51 11 1	46 14 1	68 14 1	23 15 1	62 11 1	39 17 1	29 16 1	68 18 1	20 18 1	44 13 1	
0.339	0.309	0.311	0.375	0.277	0.386	0.376	0.385	0.341	0.412	0.288	0.319	0.382	0.356	0.315

alfa = 0.7	delta = 0.7	ite = 50	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15
	28		GO terimi sayısı	11	207	0	116	3	84	11	19	0	107	59	0	0	2	5
			pHyper max değeri	1E-02	9E-03		1E-02	7E-03	9E-03	9E-03	9E-03		1E-02	1E-02			4E-03	5E-03
			gen X durum X org	188 21 1	1821 16 1	23 22 1	1789 16 1	274 15 1	518 17 1	289 16 1	309 17 1	92 16 1	1172 20 1	1058 14 1	361 20 1	139 24 1	175 19 1	197 15 1
			MSR değeri	0.156	0.237	0.101	0.291	0.261	0.243	0.247	0.271	0.318	0.302	0.287	0.292	0.331	0.309	0.288

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36
0	1	12	3	0	1	0	0	14	8	3	10	10	0	0	0	7	2	0	4	1
	7E-03	5E-03	1E-02		4E-03			8E-03	9E-03	2E-03	8E-03	1E-02				8E-03	2E-03		7E-03	3E-03
100 19 1	380 17 1	105 15 1	177 16 1	211 14 1	135 23 1	182 17 1	117 20 1	21 19 1	40 21 1	34 16 1	233 12 1	198 13 1	266 15 1	29 19 1	65 16 1	305 12 1	25 20 1	49 15 1	55 16 1	213 8 1
0.367	0.369	0.331	0.360	0.270	0.330	0.329	0.337	0.335	0.318	0.315	0.281	0.295	0.390	0.321	0.336	0.360	0.365	0.322	0.351	0.293

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
0	0	1	0	0	0	0	0	0	5	0	0	10	4	14.4
		6E-05							8E-03			1E-02	6E-03	7E-03
62 15 1	27 15 1	88 12 1	188 11 1	51 11 1	46 14 1	68 14 1	23 15 1	62 11 1	39 17 1	29 16 1	68 18 1	20 18 1	44 13 1	
0.339	0.309	0.311	0.375	0.277	0.386	0.376	0.385	0.341	0.412	0.288	0.319	0.382	0.356	0.315

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	705		705
ite = 2		pHyper max değeri	9.48E-03		9.48E-03
		gen X durum X org	12888 59 1	0 59 1	12888 59 1
		MSR değeri	0.927	NaN	0.927

DS-2 veri kümesinin fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-2 1 Organizma Fare	GO terimi zenginleşen ÜK sayısı	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK1 0	ÜK1 1	ÜK1 2	ÜK1 3	ÜK1 4	ÜK1 5
alfa = 0.3	15	GO terimi sayısı	252	205	7	182	2	0	4	5	0	0	2	3	2	5	2
delta = 0.2		pHyper max değeri	0.010	0.010	0.009	0.009	0.007		0.008	0.010			0.008	0.003	0.002	0.005	0.009
ite = 21		gen X durum X org	3724 3 1	4186 2 1	823 5 1	2741 2 1	128 4 1	162 2 1	275 2 1	273 2 1	119 4 1	73 14 1	89 5 1	51 9 1	38 17 1	49 20 1	38 3 1
		MSR değeri	0.180	0.077	0.194	0.157	0.093	0.119	0.443	0.127	0.293	0.448	0.299	0.388	0.490	0.631	0.117

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
2	0	12	0	0	4	32.809524
0.001		0.007			0.009	0.007
15 5 1	32 2 1	15 4 1	25 2 1	11 2 1	10 2 1	
0.231	0.110	0.309	0.113	0.196	0.220	0.249

alfa = 0.7	3 1	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15
delta = 0.2		GO terimi sayısı	0	83	11	9	46	25	25	4	24	33	0	7	0	0	0
ite = 50		pHyper max değeri		9E-03	8E-03	9E-03	9E-03	7E-03	7E-03	8E-03	9E-03	7E-03		9E-03			
		gen X durum X org	162 24 1	938 16 1	307 19 1	479 20 1	560 24 1	704 19 1	698 21 1	319 18 1	509 16 1	91 21 1	459 19 1	202 18 1	113 27 1	136 18 1	208 16 1
		MSR değeri	0.196	0.220	0.284	0.302	0.257	0.172	0.255	0.335	0.271	0.245	0.326	0.299	0.331	0.317	0.241

ÜK1 6	ÜK1 7	ÜK1 8	ÜK1 9	ÜK2 0	ÜK2 1	ÜK2 22	ÜK2 3	ÜK2 4	ÜK2 5	ÜK2 26	ÜK2 27	ÜK2 8	ÜK2 9	ÜK2 30	ÜK3 1	ÜK3 2	ÜK3 3	ÜK3 34	ÜK3 5	ÜK3 6	ÜK3 37	ÜK3 38	ÜK3 9	ÜK4 40	ÜK4 41	ÜK4 42
0	0	29	3	0	2	2	49	50	43	8	11	9	10	0	0	4	3	0	0	0	3	1	1	0	3	8
		1E-02	8E-03		1E-02	3E-03	3E-03	1E-02	9E-03	7E-03	9E-03	6E-03	8E-03			8E-03					5E-03	1E-03	8E-03		6E-03	8E-03
368 13 1	210 18 1	317 14 1	353 17 1	468 15 1	270 17 1	84 22 1	207 16 1	307 16 1	256 17 1	238 9 1	49 16 1	164 18 1	252 12 1	69 17 1	174 19 1	335 10 1	275 11 1	81 16 1	139 13 1	121 16 1	24 19 1	33 17 1	178 12 1	81 14 1	37 14 1	257 9 1
0.28 1	0.35 4	0.22 9	0.34 2	0.35 1	0.34 2	0.3 64	0.36 2	0.37 4	0.34 7	0.1 70	0.3 35	0.32 1	0.34 7	0.3 41	0.39 2	0.32 8	0.29 8	0.3 47	0.38 1	0.33 9	0.3 23	0.3 84	0.40 6	0.2 96	0.3 16	0.3 45

ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
1	0	5	1	0	0	20	0	10.6
3E-03		8E-03	1E-02			9E-03		7E-03
117 14 1	134 11 1	23 19 1	52 17 1	99 12 1	21 17 1	56 14 1	54 19 1	
0.322	0.332	0.377	0.456	0.314	0.355	0.374	0.399	0.320

alfa = 1	8	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ORT
delta = 0.2		GO terimi sayısı	340	369	290	19	8	26	0	2	1	0	105.5
ite = 10		pHyper max değeri	1E-02	1E-02	1E-02	9E-03	6E-03	9E-03		9E-03	7E-03		9E-03
		gen X durum X org	7121 39 1	3053 30 1	1671 31 1	578 33 1	234 34 1	139 36 1	58 33 1	14 40 1	10 34 1	10 32 1	
		MSR değeri	5E-01	6E-01	7E-01	7E-01	6E-01	7E-01	7E-01	6E-01	6E-01	6E-01	6E-01

alfa = 0.7	3 1	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	ÜK16
delta = 0.3		GO terimi sayısı	0	83	11	9	46	25	25	4	24	33	0	7	0	0	0	0
ite = 50		pHyper max değeri		9E-03	8E-03	9E-03	9E-03	7E-03	7E-03	8E-03	9E-03	7E-03		9E-03				
		gen X durum X org	162 24 1	938 16 1	307 19 1	479 20 1	560 24 1	704 19 1	698 21 1	319 18 1	509 16 1	91 21 1	459 19 1	202 18 1	113 27 1	136 18 1	208 16 1	368 13 1
		MSR değeri	2E-01	2E-01	3E-01	3E-01	3E-01	2E-01	3E-01	3E-01	3E-01	2E-01	3E-01	3E-01	3E-01	3E-01	2E-01	3E-01

ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34
0	29	3	0	2	2	49	50	43	8	11	9	10	0	0	4	0	0
	1E-02	8E-03		1E-02	3E-03	3E-03	1E-02	0.009	0.007	0.009	0.006	0.008			0.008		
210 18 1	317 14 1	353 17 1	468 15 1	270 17 1	84 22 1	207 16 1	307 16 1	256 17 1	238 9 1	49 16 1	164 18 1	252 12 1	69 17 1	174 19 1	335 10 1	275 11 1	81 16 1
4E-01	2E-01	3E-01	4E-01	3E-01	4E-01	4E-01	4E-01	0.347	0.170	0.335	0.321	0.347	0.341	0.392	0.328	0.298	0.347

ÜK35	ÜK36	ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
0	0	3	1	1	0	3	8	1	0	5	1	0	0	20	0	10.6
		0.005	0.001	0.008		0.006	0.008	0.003		0.008	0.010			0.009		0.007
139 13 1	121 16 1	24 19 1	33 17 1	178 12 1	81 14 1	37 14 1	257 9 1	117 14 1	134 11 1	23 19 1	52 17 1	99 12 1	21 17 1	56 14 1	54 19 1	
0.381	0.339	0.323	0.384	0.406	0.296	0.316	0.345	0.322	0.332	0.377	0.456	0.314	0.355	0.374	0.399	0.320

alfa = 0.7	3 1	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	ÜK16
delta = 0.7		GO terimi sayısı	0	83	11	9	46	25	25	4	24	33	0	7	0	0	0	0
ite = 50		pHyper max değeri		9E-03	8E-03	9E-03	9E-03	7E-03	7E-03	8E-03	9E-03	7E-03		9E-03				
		gen X durum X org	162 24 1	938 16 1	307 19 1	479 20 1	560 24 1	704 19 1	698 21 1	319 18 1	509 16 1	91 21 1	459 19 1	202 18 1	113 27 1	136 18 1	208 16 1	368 13 1
		MSR değeri	2E-01	2E-01	3E-01	3E-01	3E-01	2E-01	3E-01	3E-01	3E-01	2E-01	3E-01	3E-01	3E-01	3E-01	2E-01	3E-01

ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34
0	29	3	0	2	2	49	50	43	8	11	9	10	0	0	4	0	0
	1E-02	8E-03		1E-02	3E-03	3E-03	1E-02	0.009	0.007	0.009	0.006	0.008			0.008		
210 18 1	317 14 1	353 17 1	468 15 1	270 17 1	84 22 1	207 16 1	307 16 1	256 17 1	238 9 1	49 16 1	164 18 1	252 12 1	69 17 1	174 19 1	335 10 1	275 11 1	81 16 1
4E-01	2E-01	3E-01	4E-01	3E-01	4E-01	4E-01	4E-01	0.347	0.170	0.335	0.321	0.347	0.341	0.392	0.328	0.298	0.347

ÜK35	ÜK36	ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
0	0	3	1	1	0	3	8	1	0	5	1	0	0	20	0	10.6
		0.005	0.001	0.008		0.006	0.008	0.003		0.008	0.010			0.009		0.007
139 13 1	121 16 1	24 19 1	33 17 1	178 12 1	81 14 1	37 14 1	257 9 1	117 14 1	134 11 1	23 19 1	52 17 1	99 12 1	21 17 1	56 14 1	54 19 1	
0.381	0.339	0.323	0.384	0.406	0.296	0.316	0.345	0.322	0.332	0.377	0.456	0.314	0.355	0.374	0.399	0.320

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	705		705
ite = 2		pHyper max değeri	9E-03		9E-03
		gen X durum X org	12888 59 1	0 59 1	12888 59 1
		MSR değeri	9E-01	NaN	9E-01

DS-2 veri kümesinin insan ve fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-2 2 Organizma	GO terimi zenginleşen ÜK sayısı		ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	ÜK16
alfa = 0.3	32	Deney 1											ÜK1 1					
delta = 0.2		GO terimi sayısı	167	51	23	8	25	25	13	26	15	16	22	34	29	4	0	0
ite = 50		pHyper max değeri	0.008 3	0.009 2	0.009 2	0.007 3	0.009 0	0.00 95	0.004 8	0.010 0	0.00 89	0.009 7	0.00 97	0.006 0	0.009 8	0.001 4		
		gen X durum X org	1304 7 2	1001 7 2	1656 7 2	777 33 2	861 11 2	934 6 2	577 23 2	1062 15 2	374 2 2	259 23 2	689 7 2	368 53 2	149 17 2	132 10 2	168 47 2	103 31 2
		MSR değeri	0.333	0.374	0.385	0.658	0.493	0.35 1	0.587	0.577	0.21 6	0.577	0.49 8	0.715	0.597	0.531	0.751	0.703

ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36	ÜK37
0	9	2	0	0	0	0	0	15	0	2	2	1	0	0	2	1	0	1	9	1
	0.0079	0.0072						0.0086		0.0092	0.0092	0.0000			0.0088	0.0070		0.0031	0.0098	0.0065
86 24 2	234 30 2	131 31 2	100 46 2	250 3 2	89 37 2	28 2 2	92 43 2	109 24 2	71 16 2	127 2 2	133 52 2	23 48 2	71 32 2	47 21 2	60 9 2	51 45 2	48 27 2	18 22 2	21 59 2	41 48 2
0.621	0.679	0.658	0.744	0.463	0.664	0.49 2	0.742	0.706	0.604	0.136	0.821	0.778	0.722	0.708	0.580	0.787	0.657	0.638	0.889	0.824

ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
8	1	3	0	0	0	0	0	5	6	0	1	51	11.56
0.0025	0.0072	0.0095						0.0054	0.0099		0.0055	0.0098	0.0075
22 52 2	48 52 2	12 58 2	33 49 2	40 20 2	24 39 2	36 52 2	46 13 2	34 29 2	24 48 2	24 35 2	21 41 2	26 47 2	
0.817	0.766	0.862	0.845	0.658	0.799	0.857	0.567	0.702	0.840	0.787	0.806	0.832	0.648

alfa = 0.7	37	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	ÜK16
delta = 0.2		GO terimi sayısı	147	24	17	61	8	22	23	8	14	26	3	73	1	18	0	26
ite = 50		pHyper max değeri	9E-03	9E-03	9E-03	9E-03	5E-03	9E-03	8E-03	5E-03	8E-03	8E-03	8E-03	1E-02	5E-03	8E-03		9E-03
		gen X durum X org	535 12 2	1049 15 2	807 16 2	1100 9 2	128 13 2	581 10 2	858 12 2	918 8 2	614 9 2	753 8 2	295 10 2	908 7 2	114 12 2	388 9 2	81 14 2	363 10 2
	MSR değeri	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	5E-01

ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36
10	2	0	0	0	1	5	0	4	0	0	1	0	1	0	4	0	1	7	3
9E-03	8E-03				1E-02	6E-04		9E-03			8E-03		2E-03		6E-03		1E-03	1E-02	5E-03
290 10 2	410 11 2	134 11 2	101 14 2	138 12 2	103 7 2	343 5 2	143 7 2	116 9 2	118 11 2	148 6 2	161 5 2	147 7 2	43 9 2	53 9 2	74 8 2	106 7 2	90 10 2	44 8 2	24 9 2
5E-01	5E-01	5E-01	5E-01	5E-01	3E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	5E-01	4E-01	4E-01	5E-01	4E-01	4E-01	4E-01	5E-01

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
3	1	0	8	1	6	6	2	0	3	1	5	2	0	10.96
8E-03	6E-03		9E-03	1E-03	6E-03	9E-03	9E-03		9E-03	2E-03	6E-03	9E-03		7E-03
107 6 2	46 7 2	38 9 2	15 15 2	37 10 2	36 9 2	26 9 2	75 3 2	45 3 2	20 6 2	22 10 2	17 10 2	16 9 2	21 5 2	
4E-01	4E-01	5E-01	5E-01	4E-01	5E-01	5E-01	4E-01	4E-01	4E-01	5E-01	5E-01	5E-01	4E-01	4E-01

alfa = 1	7	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ORT
delta = 0.2		GO terimi sayısı	387	573	72	22	8	4	0	0	0	11	107.7
ite = 10		pHyper max değeri	9E-03	1E-02	9E-03	1E-02	1E-02	5E-03				1E-02	9E-03
		gen X durum X org	7389 36 2	3146 34 2	1533 28 2	534 29 2	161 30 2	67 33 2	31 30 2	14 33 2	10 35 2	3 32 2	
		MSR değeri	7E-01	8E-01	8E-01	8E-01	8E-01	8E-01	8E-01	6E-01	7E-01	7E-01	5E-01

alfa = 0.7	3 7	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15
delta = 0.3		GO terimi sayısı	147	24	17	61	8	22	23	8	14	26	3	73	1	18	0
ite = 50		pHyper max değeri	9E-03	9E-03	9E-03	9E-03	5E-03	9E-03	8E-03	5E-03	8E-03	8E-03	8E-03	1E-02	5E-03	8E-03	
		gen X durum X org	535 12 2	1049 15 2	807 16 2	1100 9 2	128 13 2	581 10 2	858 12 2	918 8 2	614 9 2	753 8 2	295 10 2	908 7 2	114 12 2	388 9 2	81 14 2
		MSR değeri	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36
26	10	2	0	0	0	1	5	0	4	0	0	1	0	1	0	4	0	1	7	3
9E-03	9E-03	8E-03				1E-02	6E-04		9E-03			8E-03		2E-03		6E-03		1E-03	1E-02	5E-03
363 10 2	290 10 2	410 11 2	134 11 2	101 14 2	138 12 2	103 7 2	343 5 2	143 7 2	116 9 2	118 11 2	148 6 2	161 5 2	147 7 2	43 9 2	53 9 2	74 8 2	106 7 2	90 10 2	44 8 2	24 9 2
5E-01	5E-01	5E-01	5E-01	5E-01	5E-01	3E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	5E-01	4E-01	4E-01	5E-01	4E-01	4E-01	4E-01	5E-01

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
3	1	0	8	1	6	6	2	0	3	1	5	2	0	10.96
8E-03	6E-03		9E-03	1E-03	6E-03	9E-03	9E-03		9E-03	2E-03	6E-03	9E-03		7E-03
107 6 2	46 7 2	38 9 2	15 15 2	37 10 2	36 9 2	26 9 2	75 3 2	45 3 2	20 6 2	22 10 2	17 10 2	16 9 2	21 5 2	
4E-01	4E-01	5E-01	5E-01	4E-01	5E-01	5E-01	4E-01	4E-01	4E-01	5E-01	5E-01	5E-01	4E-01	4E-01

alfa = 0.7	3 7	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	
delta = 0.7		GO terimi sayısı	147	24	17	61	8	22	23	8	14	26	3	73	1	18	0	
ite = 50		pHyper max değeri	9E-03	9E-03	9E-03	9E-03	5E-03	9E-03	8E-03	5E-03	8E-03	8E-03	8E-03	1E-02	5E-03	8E-03		
		gen X durum X org	535 12 2	1049 15 2	807 16 2	1100 9 2	128 13 2	581 10 2	858 12 2	918 8 2	614 9 2	753 8 2	295 10 2	908 7 2	114 12 2	388 9 2	81 14 2	
		MSR değeri	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01

ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ÜK22	ÜK23	ÜK24	ÜK25	ÜK26	ÜK27	ÜK28	ÜK29	ÜK30	ÜK31	ÜK32	ÜK33	ÜK34	ÜK35	ÜK36
26	10	2	0	0	0	1	5	0	4	0	0	1	0	1	0	4	0	1	7	3
9E-03	9E-03	8E-03				1E-02	6E-04		9E-03			8E-03		2E-03		6E-03		1E-03	1E-02	5E-03
363 10 2	290 10 2	410 11 2	134 11 2	101 14 2	138 12 2	103 7 2	343 5 2	143 7 2	116 9 2	118 11 2	148 6 2	161 5 2	147 7 2	43 9 2	53 9 2	74 8 2	106 7 2	90 10 2	44 8 2	24 9 2
5E-01	5E-01	5E-01	5E-01	5E-01	5E-01	3E-01	4E-01	4E-01	4E-01	4E-01	4E-01	4E-01	5E-01	4E-01	4E-01	5E-01	4E-01	4E-01	4E-01	5E-01

ÜK37	ÜK38	ÜK39	ÜK40	ÜK41	ÜK42	ÜK43	ÜK44	ÜK45	ÜK46	ÜK47	ÜK48	ÜK49	ÜK50	ORT
3	1	0	8	1	6	6	2	0	3	1	5	2	0	10.96
8E-03	6E-03		9E-03	1E-03	6E-03	9E-03	9E-03		9E-03	2E-03	6E-03	9E-03		7E-03
107 6 2	46 7 2	38 9 2	15 15 2	37 10 2	36 9 2	26 9 2	75 3 2	45 3 2	20 6 2	22 10 2	17 10 2	16 9 2	21 5 2	
4E-01	4E-01	5E-01	5E-01	4E-01	5E-01	5E-01	4E-01	4E-01	4E-01	5E-01	5E-01	5E-01	4E-01	4E-01

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	705		705
ite = 2		pHyper max değeri	9E-03		9E-03
		gen X durum X org	12888 59 2	0 59 2	
		MSR değeri	1E+00	NaN	1E+00

DS-3 veri kümesinin insan organizmasına TriClustering algoritmasının uygulanması sonuçları

DS3-1 organizma İnsan	GO terimi zenginleşen ÜK sayısı																				
alfa = 0.3	16	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ORT	
delta = 0.2		GO terimi sayısı	532	310	137	129	23	2	50	30	0	10	9	14	17	18	7	5	34	78.0588	
ite = 17		pHyper max değeri	0.010	0.010	0.010	0.010	0.009	0.005	0.010	0.009		0.004	0.008	0.010	0.008	0.008	0.009	0.009	0.010	0.009	0.009
		gen X cond X org	50651	28951	13531	3531	6431	1021	1431	1441	1021	1021	1021	1741	1431	1021	1251	1361	1021		
		MSR değeri	0.029	0.326	0.306	0.338	0.614	0.338	0.589	0.507	0.269	0.283	0.152	0.671	0.613	0.217	0.572	0.677	0.480		0.411

alfa = 0.7	7	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ORT
delta = 0.2		GO terimi sayısı	739	197	151	137	8	8	8	178.2857
ite = 7		pHyper max değeri	9.7E-03	9.7E-03	9.0E-03	9.9E-03	9.6E-03	8.7E-03	1.0E-02	9.5E-03
		gen X cond X org	75551	11221	16621	10151	2941	1021	721	
		MSR değeri	0.059	0.116	0.149	0.090	0.080	0.091	0.166	0.107

alfa = 1	5	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ORT
delta = 0.2		GO terimi sayısı	813	366	17	8	2	241.2
ite = 5		pHyper max değeri	9.35E-03	9.84E-03	9.85E-03	8.92E-03	8.03E-03	0.009197153
		gen X cond X org	85341	19441	9941	2731	731	
		MSR değeri	0.067	0.332	0.293	0.103	0.447	0.248

alfa = 0.7	7	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ORT
delta = 0.3		GO terimi sayısı	739	197	151	137	8	8	8	178.2857
ite = 7		pHyper max değeri	0.00965	0.00966	0.00901	0.00993	0.00958	0.00872	0.00995	0.0095
		gen X cond X org	75551	11221	16621	10151	2941	1021	721	
		MSR değeri	0.059	0.116	0.149	0.090	0.080	0.091	0.166	0.107

alfa = 0.7	1	Deney 5	ÜK1	ÜK2	ORT
delta = 0.7		GO terimi sayısı	1101		1101
ite = 2		pHyper max değeri	9.84E-03		9.84E-03
		gen X cond X org	1180 6 1	0 6 1	
		MSR değeri	0.365	NaN	0.365

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	1101		1101
ite = 2		pHyper max değeri	9.84E-03		9.84E-03
		gen X cond X org	1180 6 1	0 6 1	
		MSR değeri	0.365		0.365

DS-3 veri kümesinin sığır organizmasına TriClustering algoritmasının uygulanması sonuçları

DS3-1 Organizma Sığır	GO terimi zenginleşen ÜK sayısı	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ORT
alfa = 0.3	11	GO terimi sayısı	213	457	177	62	22	33	17	358	11	3	13	124.1818 2
delta = 0.2		pHyper max değeri	0.0096	0.0097	0.0096	0.008 6	0.007 5	0.007 8	0.009 0	0.0095	0.009 4	0.004 0	0.008 8	0.0085
ite = 11		gen X cond X org	217 2 1	327 2 1	159 4 1	53 3 1	54 4 1	26 4 1	20 5 1	293 2 1	10 5 1	12 5 1	9 5 1	
		MSR değeri	0.128	0.510	0.275	0.547	0.517	0.610	0.443	0.328	0.553	0.746	0.663	0.484

alfa = 0.7	12	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ORT
delta = 0.2		GO terimi sayısı	597	29	131	158	129	4	5	25	0	16	28	2	4	86.769231
ite = 13		pHyper max değeri	0.0099	0.0100	0.0091	0.0100	0.0099	0.0054	0.0090	0.0097		0.0078	0.0097	0.0049	0.0062	0.0085
		gen X cond X org	574 2 1	43 3 1	149 3 1	169 2 1	99 2 1	34 2 1	10 6 1	38 2 1	14 2 1	19 2 1	14 2 1	10 2 1	7 2 1	
		MSR değeri	0.238	0.025	0.126	0.145	0.166	0.095	0.349	0.313	0.130	0.073	0.413	0.126	0.318	0.194

alfa = 1	5	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ORT
delta = 0.2		GO terimi sayısı	728	500	49	1	16	258.8
ite = 5		pHyper max değeri	9.9E-03	9.7E-03	1.0E-02	5.1E-03	1.0E-02	8.9E-03
		gen X cond X org	798 3 1	292 3 1	65 3 1	18 3 1	7 2 1	
		MSR değeri	3.1E-01	2.7E-01	4.1E-01	2.7E-01	5.7E-01	3.7E-01

alfa = 0.7	12	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ORT
delta = 0.3		GO terimi sayısı	597	29	131	158	129	4	5	25	0	16	28	2	4	86.769231
ite = 13		pHyper max değeri	0.010	0.010	0.009	0.010	0.010	0.005	0.009	0.010		0.008	0.010	0.005	0.006	0.008
		gen X cond X org	574 2 1	43 3 1	149 3 1	169 2 1	99 2 1	34 2 1	10 6 1	38 2 1	14 2 1	19 2 1	14 2 1	10 2 1	7 2 1	
		MSR değeri	0.238	0.025	0.126	0.145	0.166	0.095	0.349	0.313	0.130	0.073	0.413	0.126	0.318	0.194

alfa = 0.7	12	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ÜK13	ORT
delta = 0.7		GO terimi sayısı	597	29	131	158	129	4	5	25	0	16	28	2	4	86.769231
ite = 13		pHyper max değeri	0.0099	0.0100	0.0091	0.0100	0.0099	0.0054	0.0090	0.0097		0.0078	0.0097	0.0049	0.0062	0.0085
		gen X cond X org	574 2 1	43 3 1	149 3 1	169 2 1	99 2 1	34 2 1	10 6 1	38 2 1	14 2 1	19 2 1	14 2 1	10 2 1	7 2 1	
		MSR değeri	0.238	0.025	0.126	0.145	0.166	0.095	0.349	0.313	0.130	0.073	0.413	0.126	0.318	0.194

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	1101		1101
ite = 2		pHyper max değeri	9.8E-03		9.8E-03
		gen X cond X org	1180 6 1	0 6 1	1180 6 1
		MSR değeri	8.3E-01	NaN	8.3E-01

DS-3 veri kümesinin fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS3-1 Organizma-Fare	GO terimi zenginleşen ÜK sayısı														
alfa = 0.3	9	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ORT			
delta = 0.2		GO terimi sayısı	353	326	92	67	29	341	39	14	62	147			
ite = 9		pHyper max değeri	1.0E-02	9.9E-03	9.5E-03	9.8E-03	1.0E-02	9.2E-03	8.4E-03	8.2E-03	9.9E-03	9.4E-03			
		gen X cond X org	358 2 1	361 3 1	53 6 1	109 5 1	10 5 1	243 5 1	25 2 1	18 6 1	3 3 1				
		MSR değeri	2.6E-01	2.2E-01	9.1E-03	3.2E-01	1.6E-01	2.6E-01	1.5E-01	7.7E-01	5.4E-01	3.0E-01			

alfa = 0.7	12	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ORT
delta = 0.2		GO terimi sayısı	327	305	56	295	55	21	35	68	30	59	1	20	106
ite = 12		pHyper max değeri	1.0E-02	1.0E-02	8.9E-03	9.9E-03	9.7E-03	6.8E-03	9.1E-03	9.0E-03	9.7E-03	1.0E-02	5.4E-03	9.7E-03	9.0E-03
		gen X cond X org	243 3 1	338 2 1	89 4 1	228 2 1	96 2 1	12 2 1	96 2 1	32 2 1	19 3 1	12 2 1	11 2 1	4 2 1	
		MSR değeri	1.8E-01	2.7E-01	7.6E-02	1.2E-01	1.3E-01	2.7E-02	2.3E-01	1.7E-01	2.8E-01	2.7E-01	3.2E-01	1.4E-01	1.8E-01

alfa = 1	6	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ORT
delta = 0.2		GO terimi sayısı	764	283	37	51	7	7	191.5
ite = 6		pHyper max değeri	9.8E-03	9.9E-03	9.9E-03	8.7E-03	9.5E-03	9.8E-03	9.6E-03
		gen X cond X org	710 4 1	338 3 1	73 3 1	36 4 1	16 3 1	7 4 1	
		MSR değeri	2.7E-01	3.2E-01	3.0E-01	1.5E-01	1.3E-01	3.7E-01	2.5E-01

alfa = 0.7	12	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ORT
delta = 0.3		GO terimi sayısı	327	305	56	295	55	21	35	68	30	59	1	20	106
ite = 12		pHyper max değeri	1.0E-02	1.0E-02	8.9E-03	9.9E-03	9.7E-03	6.8E-03	9.1E-03	9.0E-03	9.7E-03	1.0E-02	5.4E-03	9.7E-03	9.0E-03
		gen X cond X org	243 3 1	338 2 1	89 4 1	228 2 1	96 2 1	12 2 1	96 2 1	32 2 1	19 3 1	12 2 1	11 2 1	4 2 1	
		MSR değeri	1.8E-01	2.7E-01	7.6E-02	1.2E-01	1.3E-01	2.7E-02	2.3E-01	1.7E-01	2.8E-01	2.7E-01	3.2E-01	1.4E-01	1.8E-01

alfa = 0.7	12	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ÜK12	ORT
delta = 0.7		GO terimi sayısı	327	305	56	295	55	21	35	68	30	59	1	20	106
ite = 12		pHyper max değeri	1.0E-02	1.0E-02	8.9E-03	9.9E-03	9.7E-03	6.8E-03	9.1E-03	9.0E-03	9.7E-03	1.0E-02	5.4E-03	9.7E-03	9.0E-03
		gen X cond X org	243 3 1	338 2 1	89 4 1	228 2 1	96 2 1	12 2 1	96 2 1	32 2 1	19 3 1	12 2 1	11 2 1	4 2 1	
		MSR değeri	1.8E-01	2.7E-01	7.6E-02	1.2E-01	1.3E-01	2.7E-02	2.3E-01	1.7E-01	2.8E-01	2.7E-01	3.2E-01	1.4E-01	1.8E-01

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	1101		
ite = 2		pHyper max değeri	9.84E-03		
		gen X cond X org	1180 6 1	0 6 1	
		MSR değeri	0.7924576		

DS-3 veri kümesinin insan, sığır ve fare organizmasına TriClustering algoritmasının uygulanması sonuçları

DS-3 3 organizma	GO terimi zenginleşen ÜK sayısı													
alfa = 0.3	21	Deney 1	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	
delta = 0.2		GO terimi sayısı	264	371	113	16	109	6	41	29	56	21	39	
ite = 21		pHyper max değeri	0.0099	0.0098	0.0097	0.0096	0.0096	0.0037	0.0090	0.0090	0.0100	0.0092	0.0095	
		gen X durum X org	242 2 2	381 5 3	112 5 3	59 2 2	71 2 2	10 6 3	54 5 3	10 6 3	39 5 3	24 5 3	85 4 2	
		MSR değeri	0.2200	0.5727	0.6076	0.3112	0.1880	0.7138	0.6860	0.7118	0.6998	0.7392	0.4305	
	olmayan org	insan			insan	fare						fare		

ÜK12	ÜK13	ÜK14	ÜK15	ÜK16	ÜK17	ÜK18	ÜK19	ÜK20	ÜK21	ORT
17	2	52	23	44	2	37	62	12	22	63.714286
0.0081	0.0086	0.0089	0.0072	0.0096	0.0090	0.0084	0.0089	0.0058	0.0097	0.0087
11 6 3	10 6 2	10 6 2	10 6 2	10 6 3	10 6 3	10 6 3	10 6 2	10 6 2	2 6 2	
0.7809	0.6565	0.5956	0.5594	0.7493	0.7905	0.6970	0.4859	0.5665	0.4774	0.5828
	insan	insan	insan				human	human	fare	

alfa = 0.7	11	Deney 2	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ort
delta = 0.2		GO terimi sayısı	382	290	79	306	14	16	87	4	3	29	1	110.09091
ite = 11		pHyper max değeri	1E-02	1E-02	1E-02	1E-02	1E-02	7E-03	1E-02	3E-03	3E-03	9E-03	9E-03	8E-03
		gen X durum X org	357 2 2	281 2 3	144 2 3	259 2 2	45 2 2	15 2 2	30 2 2	14 3 2	11 2 2	14 5 3	10 6 2	
		MSR değeri	0.2260	0.4087	0.4248	0.2271	0.2847	0.4875	0.4132	0.3742	0.4194	0.7443	0.8041	0.4377
		olmayan org	fare			fare	fare	insan	fare	insan	fare		insan	

alfa = 1	5	Deney 3	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ORT
delta = 0.2		GO terimi sayısı	732	485	28	12	4	252.2
ite = 5		pHyper max değeri	1E-02	1E-02	8E-03	1E-02	6E-03	9E-03
		gen X durum X org	789 3 2	267 3 2	86 2 2	17 4 2	12 3 2	
		MSR değeri	0.2912457	0.4432414	0.3339121	0.4808815	0.3990297	0.3896621
		olmayan org	fare	fare	siğır	siğır	insan	

alfa = 0.7	11	Deney 4	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ORT
delta = 0.3		GO terimi sayısı	382	290	79	306	14	16	87	4	3	29	1	110.09091
ite = 11		pHyper max değeri	1E-02	1E-02	1E-02	1E-02	1E-02	7E-03	1E-02	3E-03	3E-03	9E-03	9E-03	8E-03
		gen X durum X org	357 2 2	281 2 3	144 2 3	259 2 2	45 2 2	15 2 2	30 2 2	14 3 2	11 2 2	14 5 3	10 6 2	
		MSR değeri	0.2260	0.4087	0.4248	0.2271	0.2847	0.4875	0.4132	0.3742	0.4194	0.7443	0.8041	0.4377
		olmayan org	fare			fare	fare	insan	fare	insan	fare		insan	

alfa = 0.7	11	Deney 5	ÜK1	ÜK2	ÜK3	ÜK4	ÜK5	ÜK6	ÜK7	ÜK8	ÜK9	ÜK10	ÜK11	ORT
delta = 0.7		GO terimi sayısı	382	290	79	306	14	16	87	4	3	29	1	110.09091
ite = 11		pHyper max değeri	1E-02	1E-02	1E-02	1E-02	1E-02	7E-03	1E-02	3E-03	3E-03	9E-03	9E-03	8E-03
		gen X durum X org	357 2 2	281 2 3	144 2 3	259 2 2	45 2 2	15 2 2	30 2 2	14 3 2	11 2 2	14 5 3	10 6 2	
		MSR değeri	0.2260	0.4087	0.4248	0.2271	0.2847	0.4875	0.4132	0.3742	0.4194	0.7443	0.8041	0.4377
		olmayan org	fare			fare	fare	insan	fare	insan	fare		insan	

alfa = 0.7	1	Deney 6	ÜK1	ÜK2	ORT
delta = 1		GO terimi sayısı	1101		1101
ite = 2		pHyper max değeri	0.0098		0.0098
		gen X durum X org	1180 6 3	0 6 3	1180 6 3
		MSR değeri	0.7798		0.7798
		olmayan org			