Bilim Akademisi - Bilkent Üniversitesi BYYO 2020
2 Temmuz 2020

Fonksiyonel Çizge Çekirdekleri ile Kanser Hastalarının Gruplanması

Öznur Taştan

otastan@sabanciuniv.edu

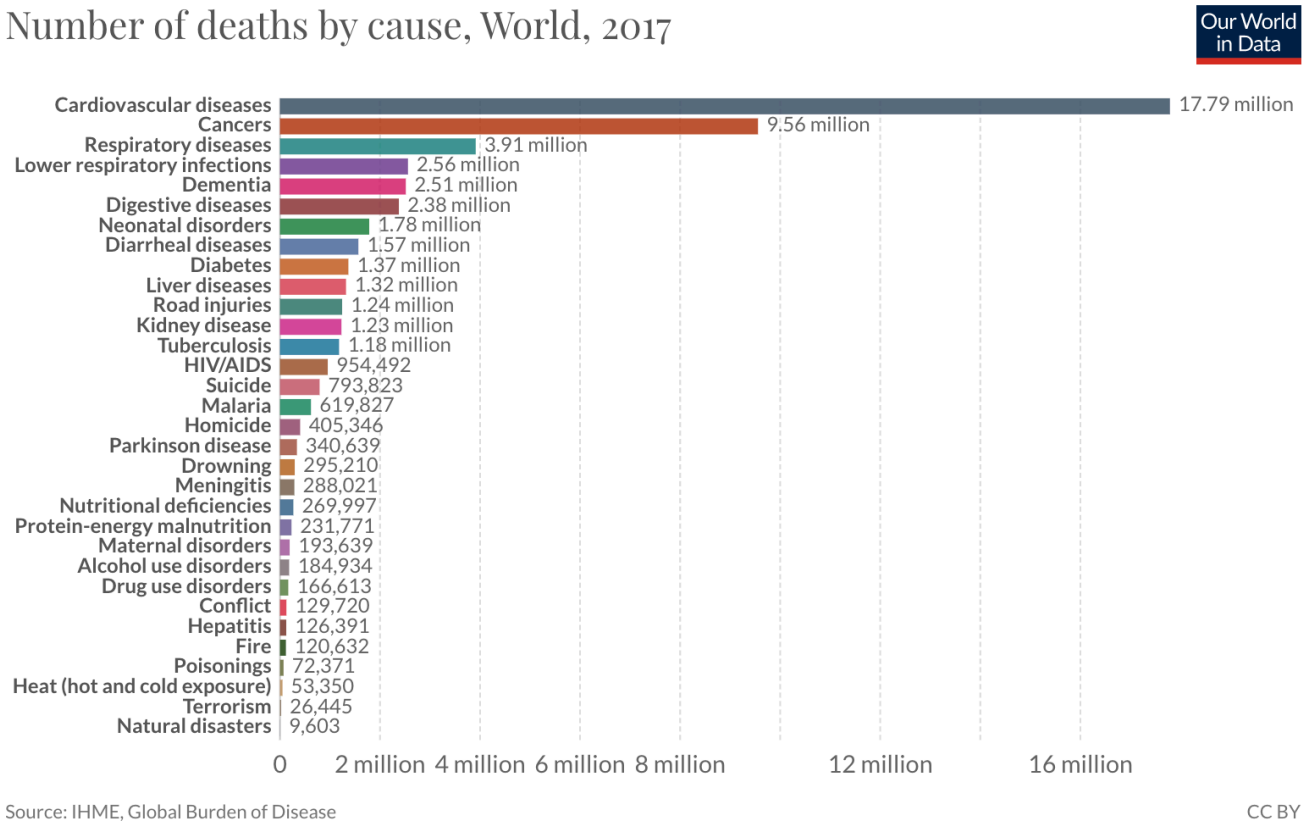
Sabancı Üniversitesi

Mühendislik ve Doğa Bilimleri Fakültesi
Bilgisayar Bilimi ve Mühendisliği
&
Moleküler Biyoloji, Genetik ve Biyomühendislik

Ana Hat

- Kanser hastalarını öbekleme
- Çoklu-omik verileri
- Çok bakış açılı öğrenme
- Çekirdeklerle öğrenme
- Çizge çekirdekleri
- PAMOGK – Yolak Bazlı Çizge Çekirdeğine dayalı Çoklu-Omik Öbekleme Algoritması

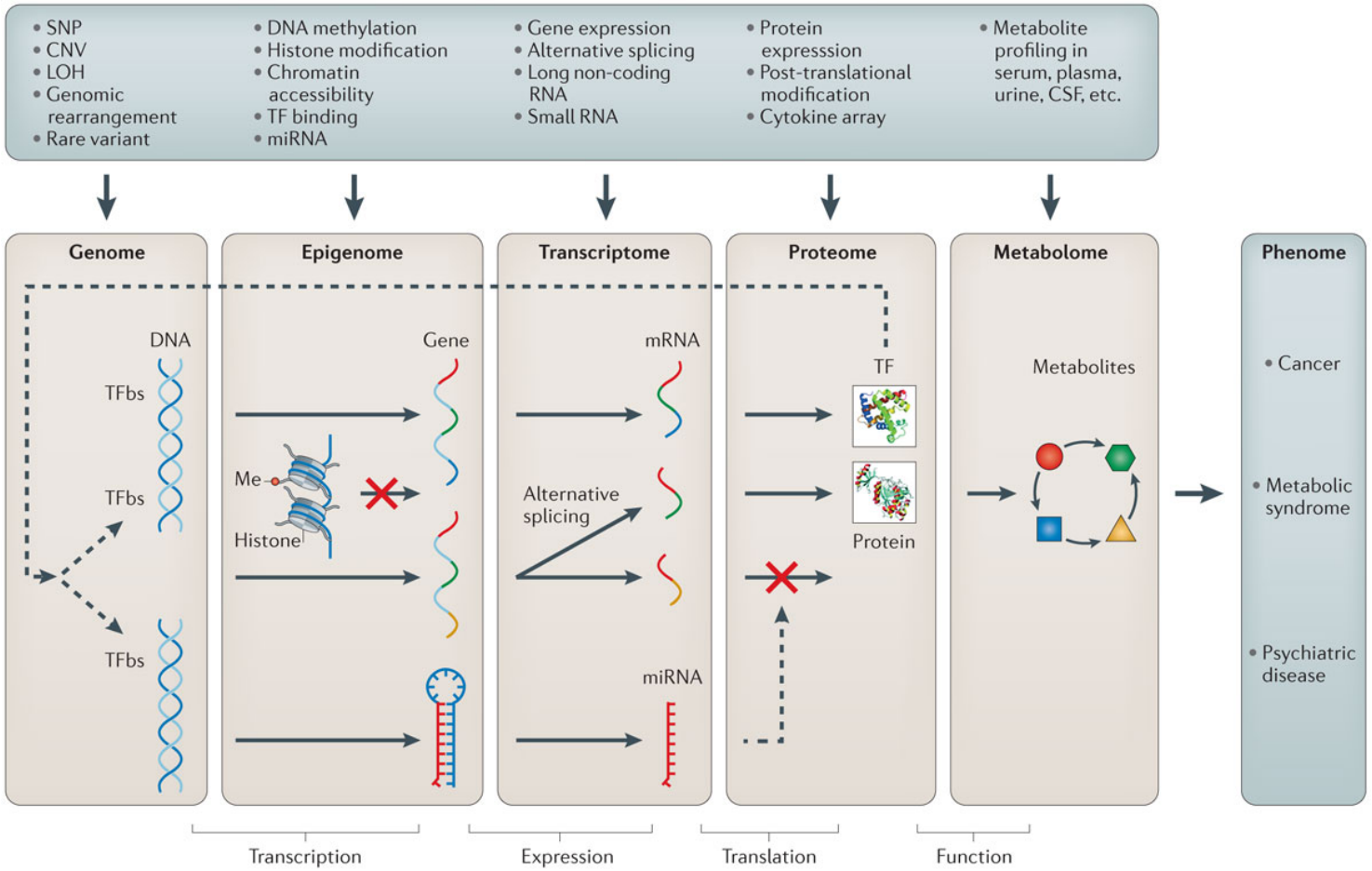
Kanser Hala bir Küresel bir Sağlık Problemi



Kanser gibi Karmaşık Hastalıklarla Müceadele

- Hücre biyolojisini anlamak
  - Veri toplayacak deney ve ölçüm enstrümanlarında teknolojik ve bilimsel ilerleme
  - Veriyi anlamlandıracak hesapsal yöntemlerde ilerleme

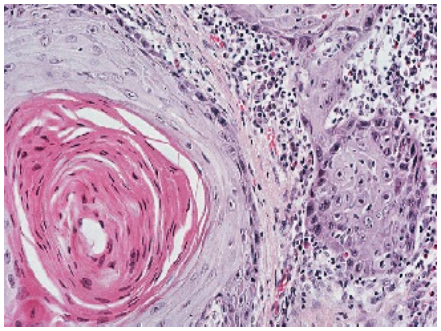
Omiks Cağı



Ritchie *ve* ark. Nature Reviews Genetics. 16, 85–97, 2015.

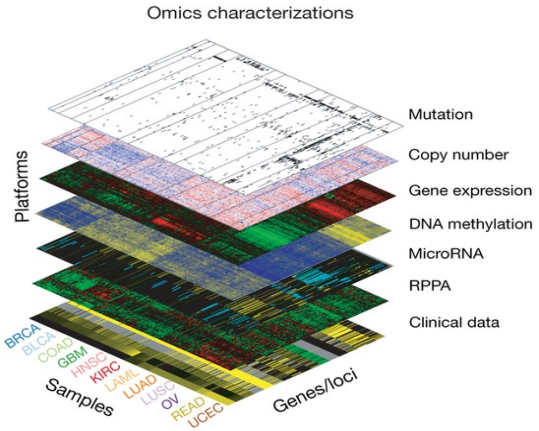
Kanser Genom Projesi

Tümör örnekleri



Deneyler

Hastaların moleküler değişim katalogları



Veri Analizi

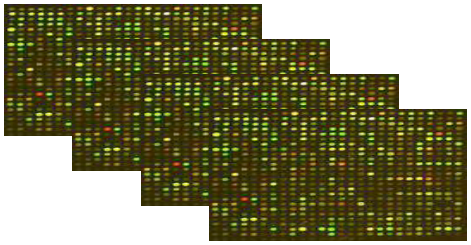
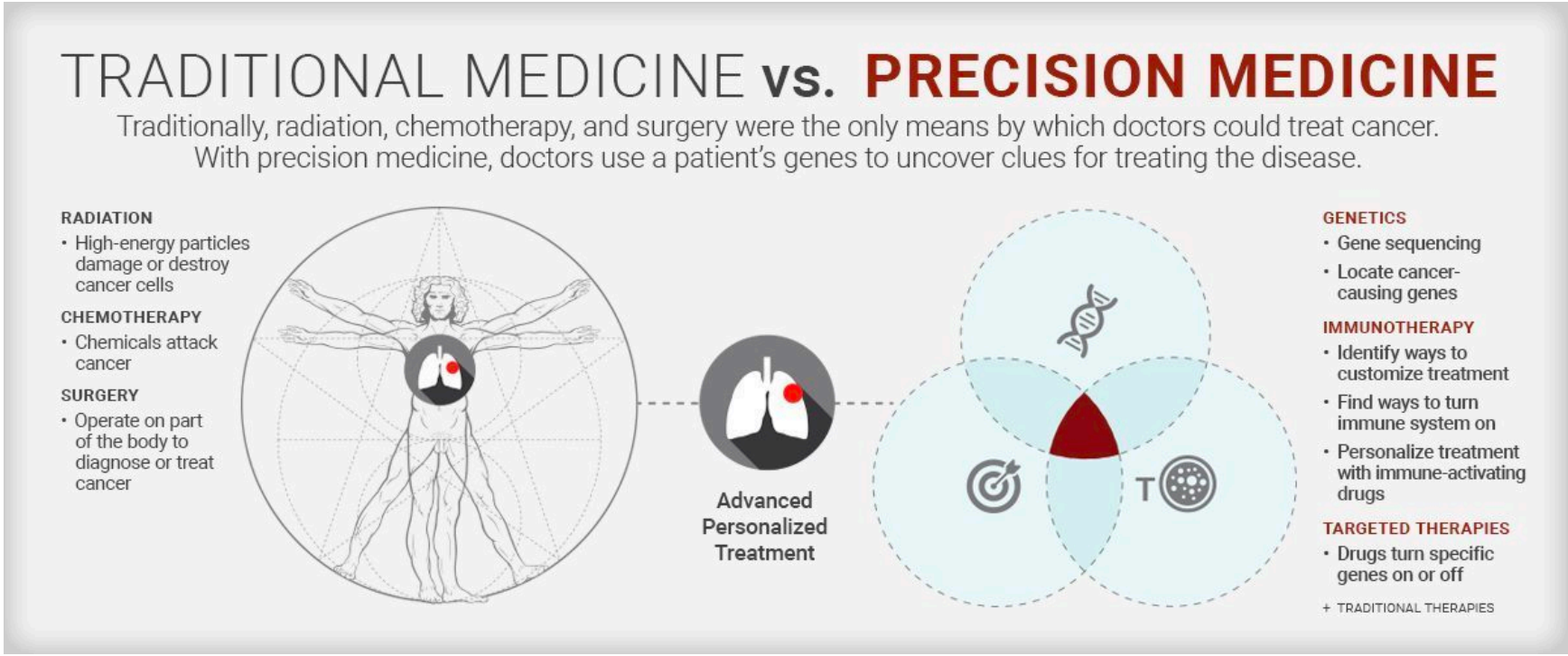


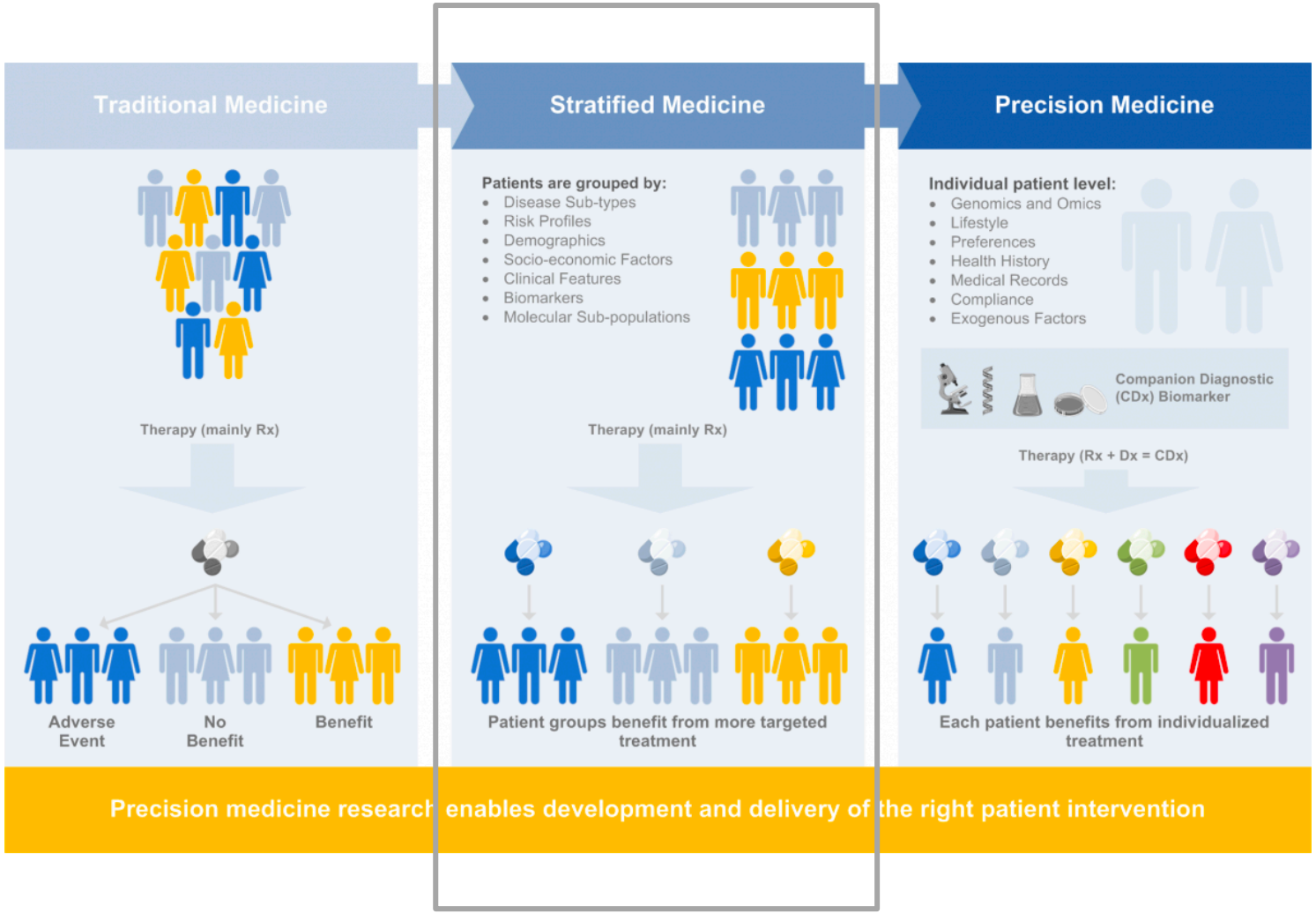
Figure kaynağı: Weinstein et al., Nature Genetics, 2013.

Kişiselleştirilmiş Tıp



Figur kaynagi: https://healthmatters.nyp.org/precision-medicine/

Farklı Ölçekte Tedavi Yaklaşımları



Figüre kaynağı: https://www.linguamatics.com/solutions/precision-medicine

Aynı Kanser Tipi Farklı Kilinik Özellikler

- Aynı kanser, farklı klinik yörüngeler
- Benzer moleküler değişim örüntüleri olan hasta alt gruplarının bulunması hedefli tedavilerin geliştirilmesine yardımcı olur

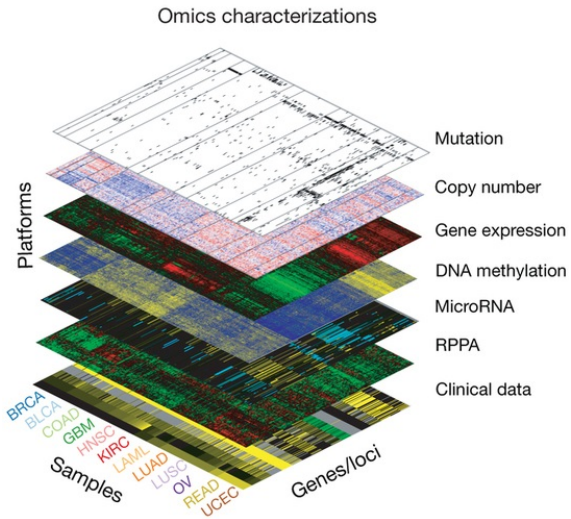


Figure kaynağı: Weinstein et al., Nature Genetics, 2013.

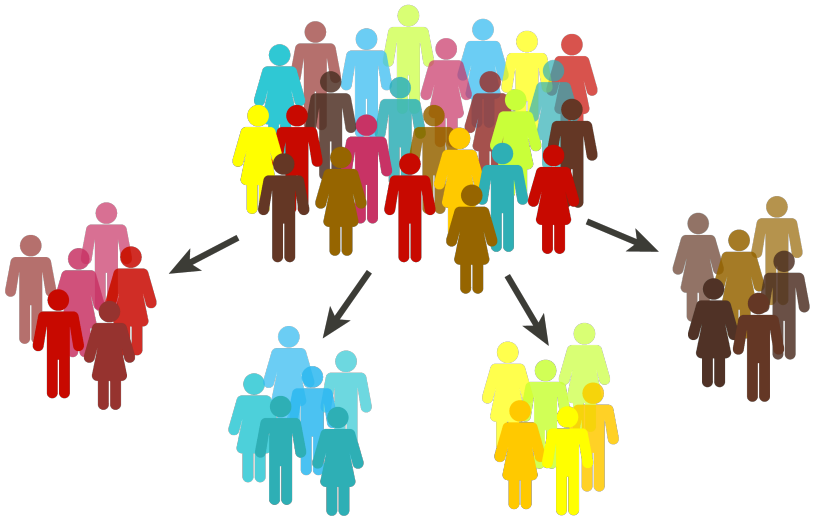
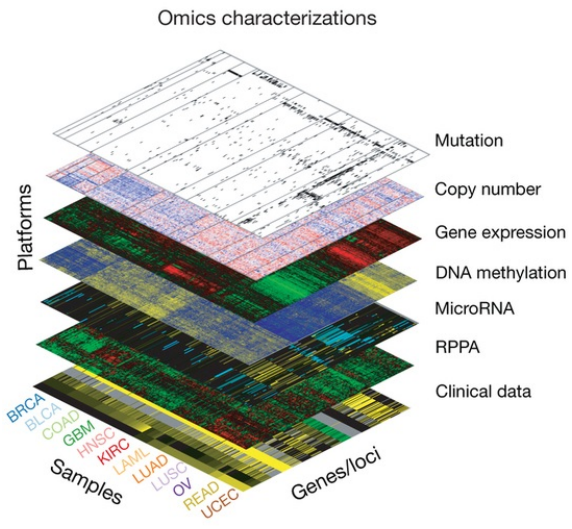
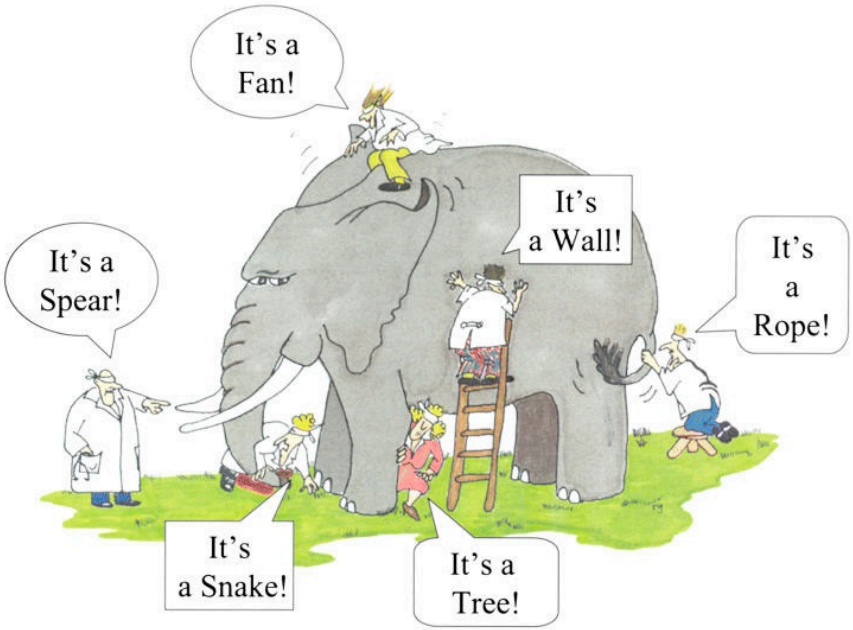


Figure kaynağı
http://www.pancreaticcancer.net.au/ground-breaking-nature/

Çoklu Bakış Açılı Öbekleme



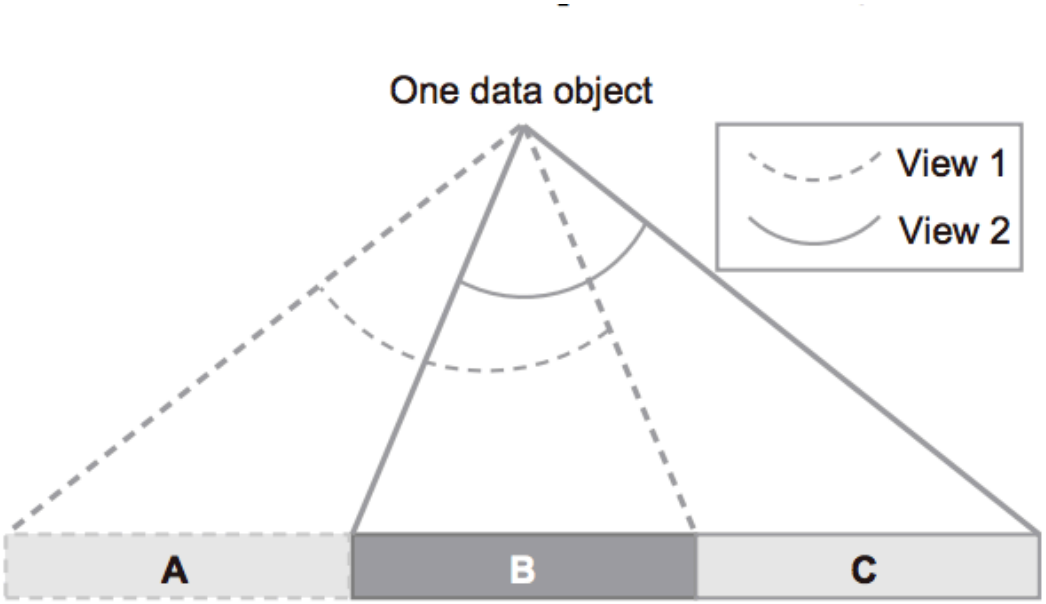
Figür kaynağı: Weinstein et al., Nature Genetics, 2013.



Figür kaynağı: https://towardsdatascience.com

Çoklu Bakış Açılı Öbekleme

- Uzlaşma prensibi
- Tamamlayıcılık prensibi



Figür: Yang and Wang, Big Data Mining and Analytics, 2018.

10546–10562 *Nucleic Acids Research, 2018, Vol. 46, No. 20*
doi: 10.1093/nar/gky889
Published online 8 October 2018

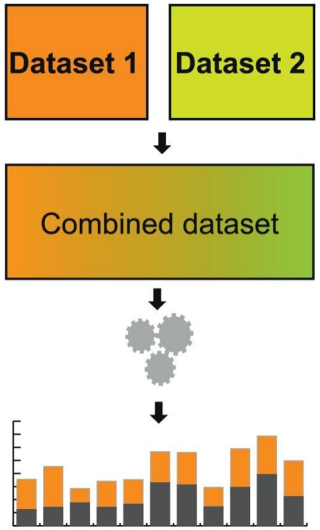
SURVEY AND SUMMARY

Multi-omic and multi-view clustering algorithms: review and cancer benchmark

Nimrod Rappoport and Ron Shamir*

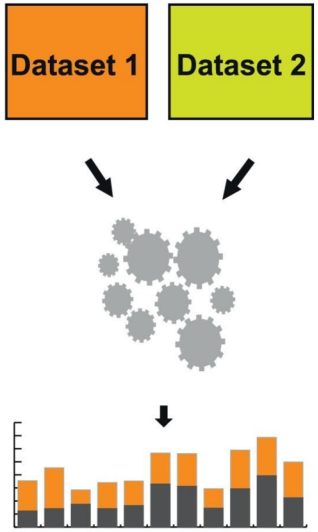
Multi-Omics Entegrasyonu

Early Integration



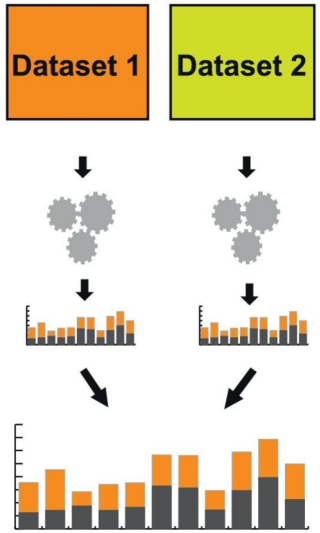
iCluster
(Shen et al. 2009, Bioinformatics)
LRACluster
(Wu et al. 2015, BMC Genomic)

Intermediate Integration



MCCA (Witten et al. 2009, Stat Appl Genet Mol Biol.)
SNF (Wang et al. 2014, Nature)
rMKL-LPP (Speicher et al. 2015, Bioinformatics)

Late Integration



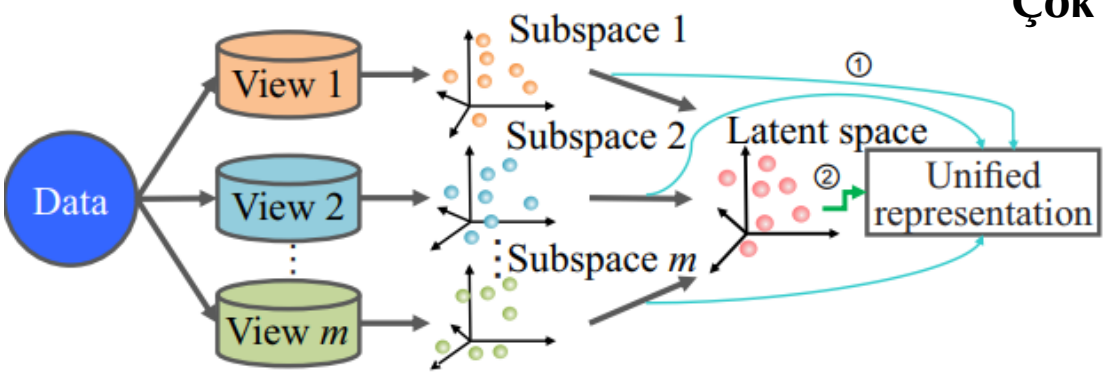
Consensus (Strehl et al. 2002, Journal of Machine Learning Research)
PINS (Nguyen et al. 2017, Genome Res)

Review article: Rappaport, Shamir. 2018, Nucleic Acid Research

Figür kaynağı
Zitnik et al. 2019, Information Fusion

Çoklu Bakış Açılı Kümeleme

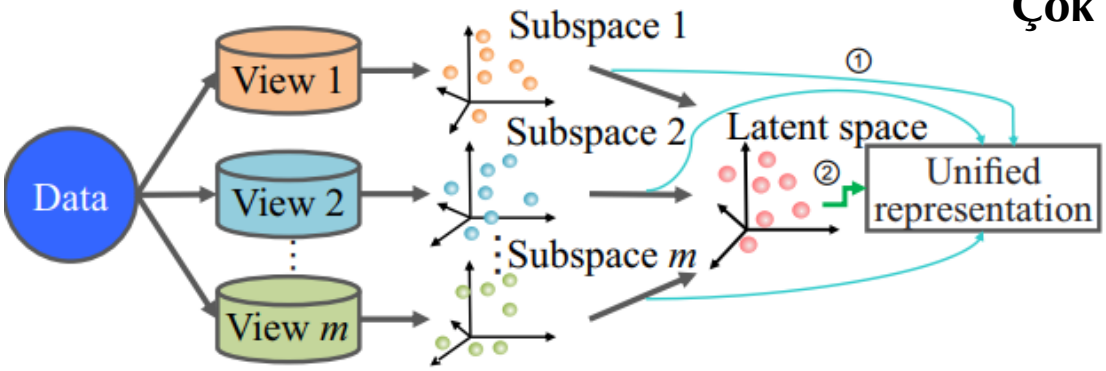
Çok bakış açılı altuzay kümeleme



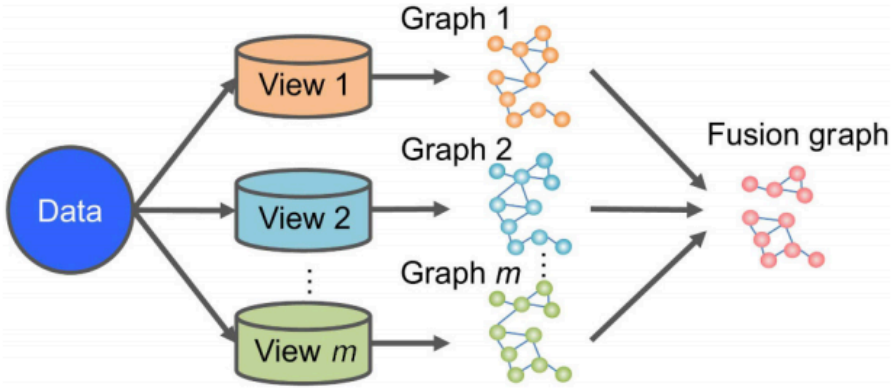
Figür kaynağı: Yang and Wang, Big Data Mining and Analytics, 2018.

Çoklu Bakış Açılı Kümeleme

Çok bakış açılı altuzay kümeleme

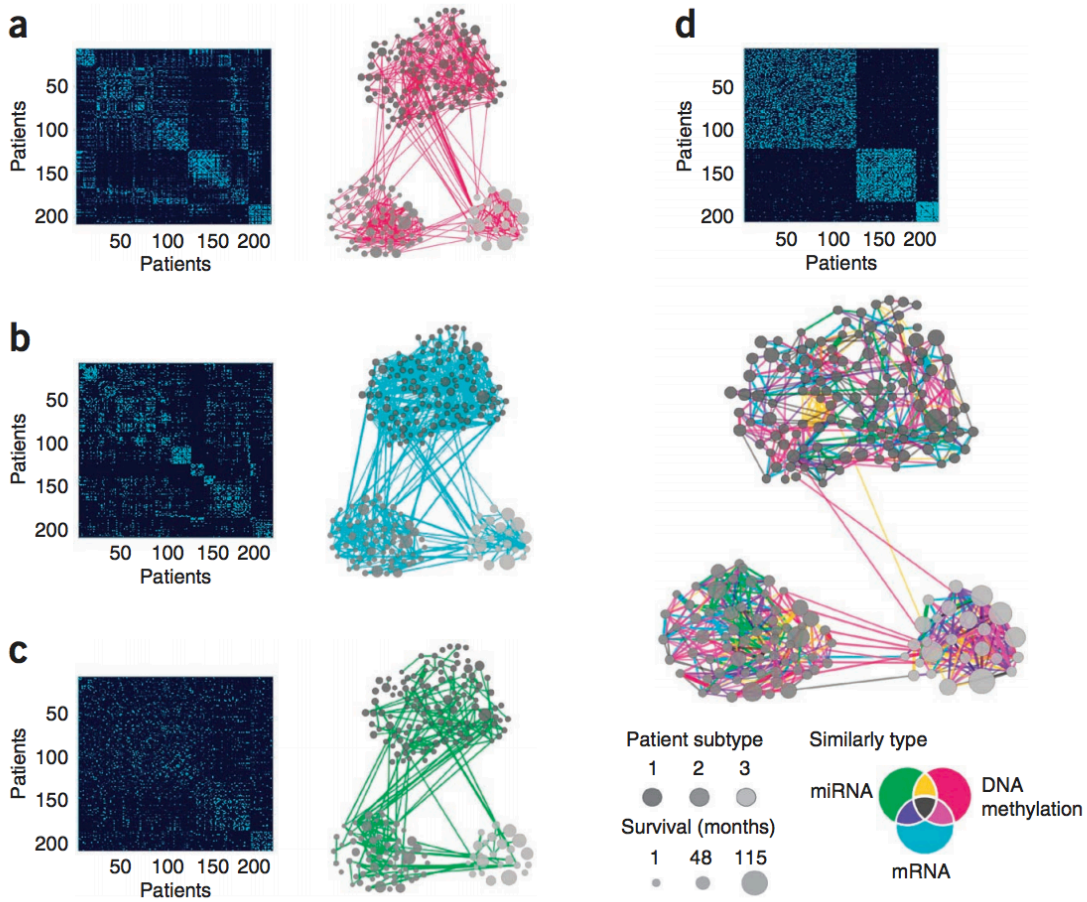


Çok bakış açılı füzyon tabanlı kümeleme



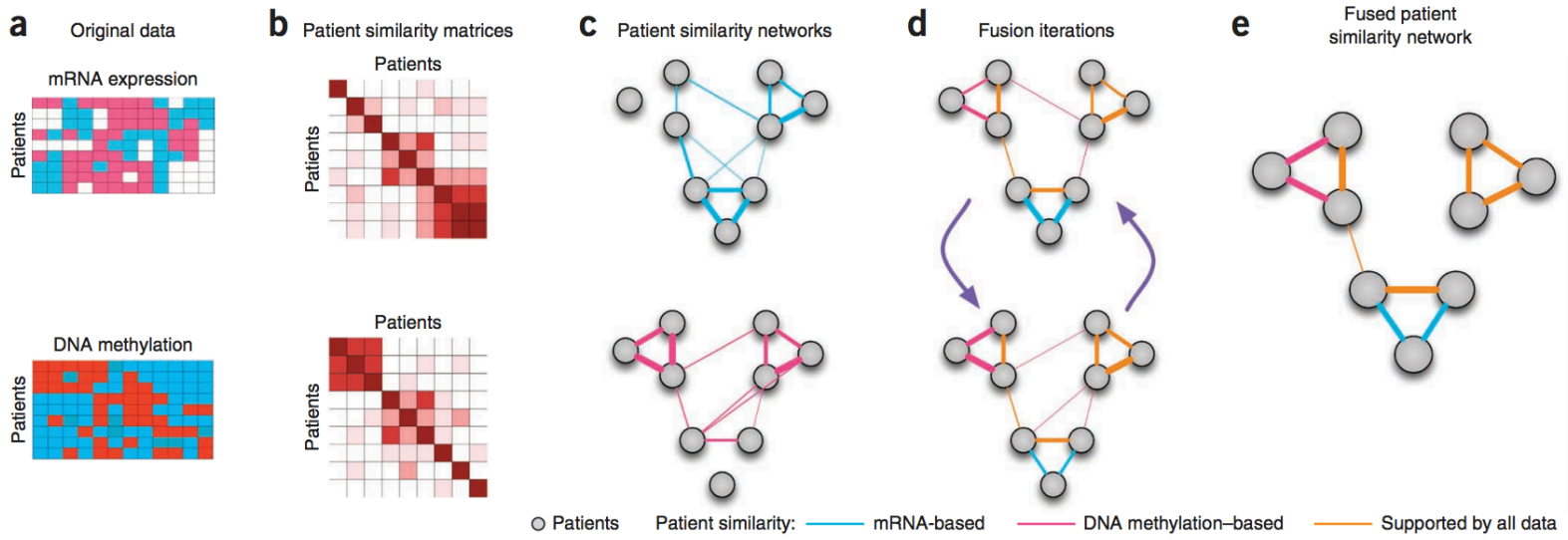
Figür kaynağı: Yang and Wang, Big Data Mining and Analytics, 2018.

SNF (Benzerlik Ağı Füzyon Algoritması)



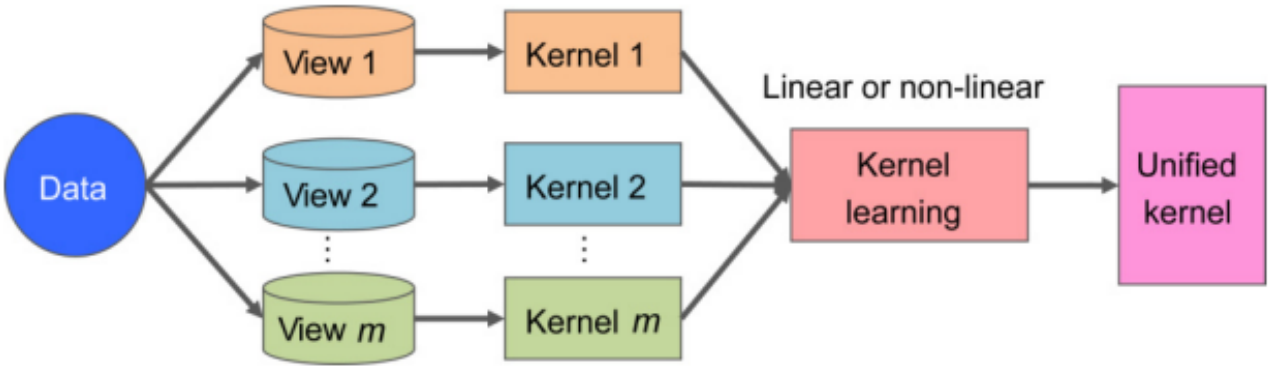
Wang et al. Nature Methods, 2014.

SNF (Benzerlik Ağı Füzyon Algoritması)



Wang et al. Nature Methods, 2014.

Çok Çekirdekli Öğrenme



Figür kaynağı: Yang and Wang, Big Data Mining and Analytics, 2018.

Öznitelik Uzayı Transformasyonu

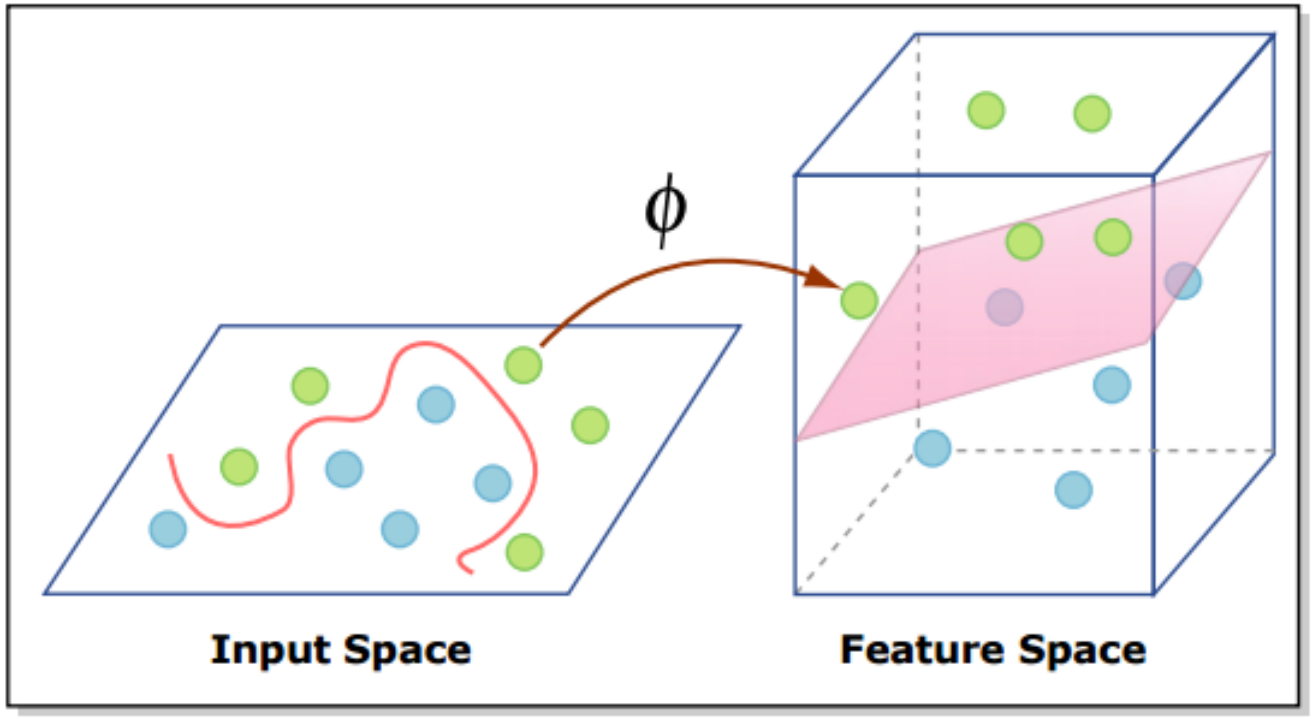
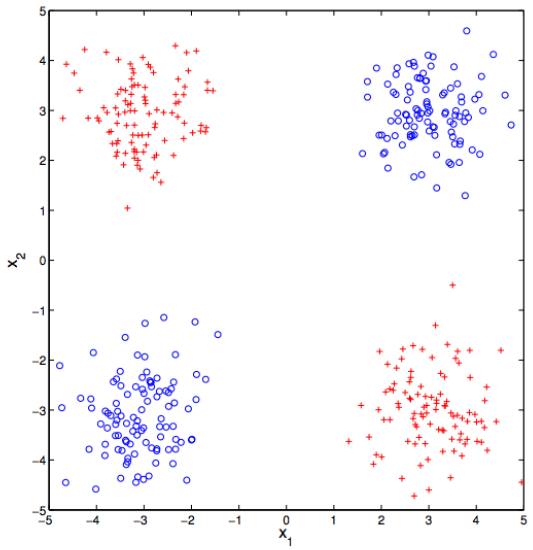
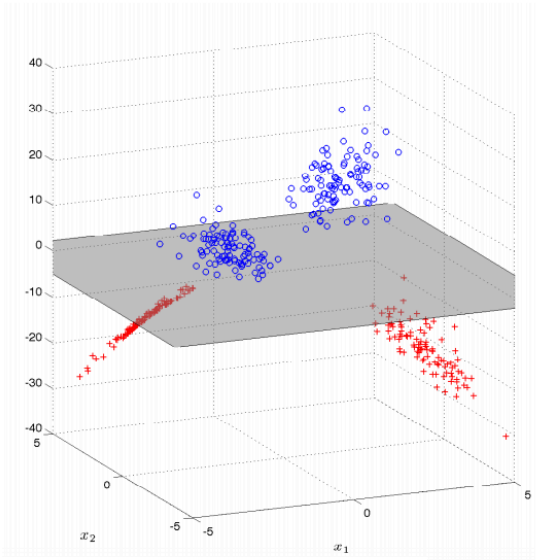


Image by MIT OpenCourseWare.

Çekirdekler ve öznitelik uzayı





$$\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \phi(x) = \begin{bmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{bmatrix}$$

Figür kaynağı: Aurthur Gretton

Çekirdek Fonksiyonları

$x_i, x_j \in \mathcal{X}$

$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_H$

$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

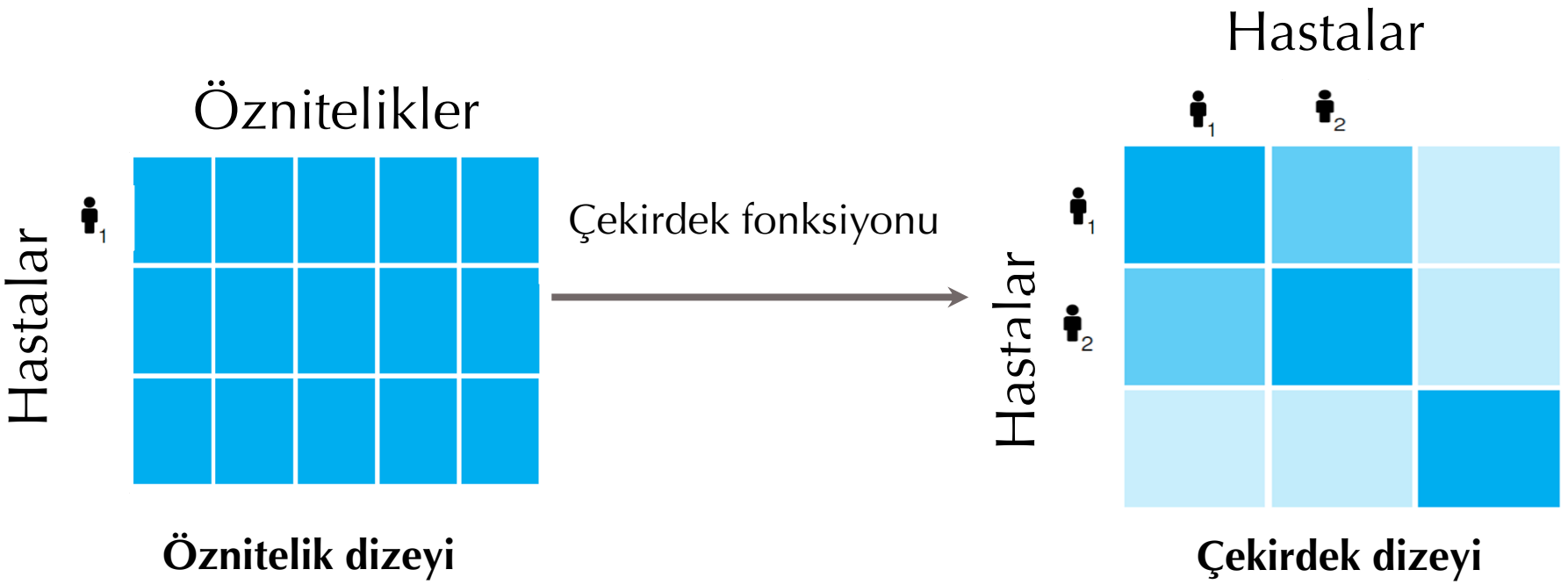
$\phi : \mathcal{X} \to H$

Doğrusal çekirdek fonksiyonu: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\mathsf{T} \mathbf{x}_j$

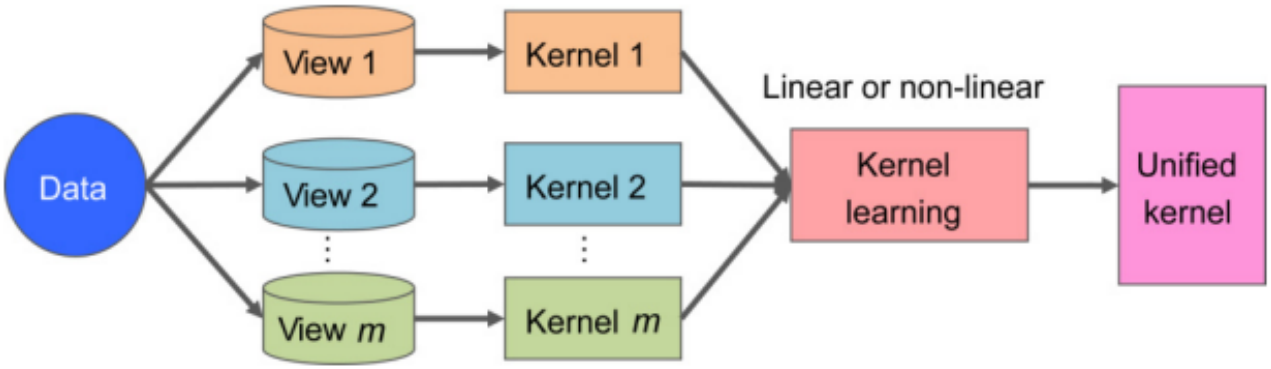
Gauss çekirdek fonksiyonu: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}$

Polinom çekirdek fonksiyonu: $K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i^\mathsf{T} \mathbf{x}_j)^d$

Çekirdek Dizeyi



Çok Çekirdekli Öğrenme



Figür kaynağı: Yang and Wang, Big Data Mining and Analytics, 2018.

Çoklu bakış açılı çekirdek k ortalama metodu

Amaç fonksiyonu:

v. bakış açısı üzerinde hesaplanmış çekirdek dizeyi

v. bakış açısının ağırlığı

$$\underset{\boldsymbol{Y},\boldsymbol{W}}{\text{minimize}} \quad \sum_{v=1}^{V} w_v^p (\text{Tr}(\boldsymbol{K}^{(v)}) - \text{Tr}(\boldsymbol{Y}^T \boldsymbol{K}^{(v)} \boldsymbol{Y}))$$

$$\text{subject to} \quad w_v \geq 0, \quad v = 1, 2, \ldots, V$$

$$\sum_{v=1}^{V} w_v = 1$$

$$p \geq 1$$

Öbekleme sonucunu içeren indikatör dizeyi

Aynı Kanser Tipi Farklı Kilinik Özellikler

- Aynı kanser, farklı klinik yörüngeler
- Benzer moleküler değişim örüntüleri olan hasta alt gruplarının bulunması hedefli tedavilerin geliştirilmesine yardımcı olur

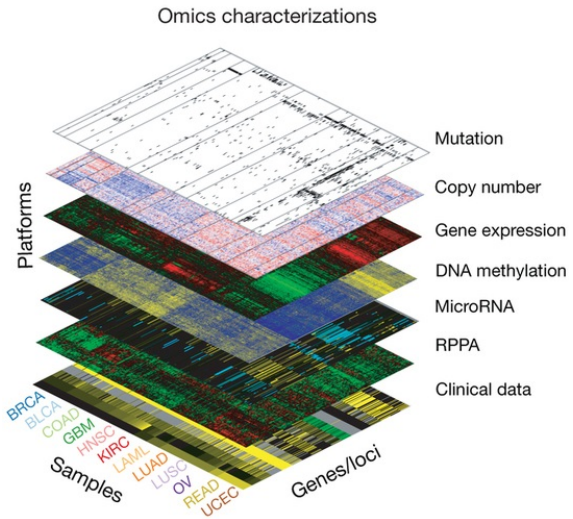


Figure kaynağı: Weinstein et al., Nature Genetics, 2013.

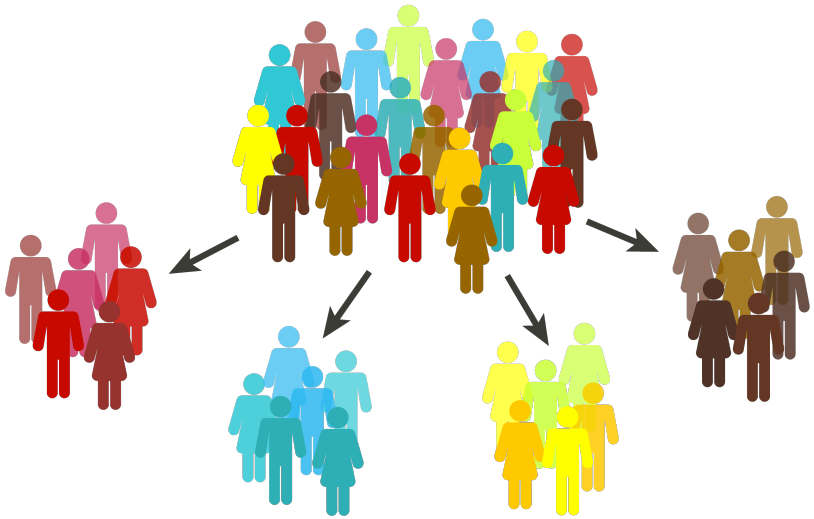
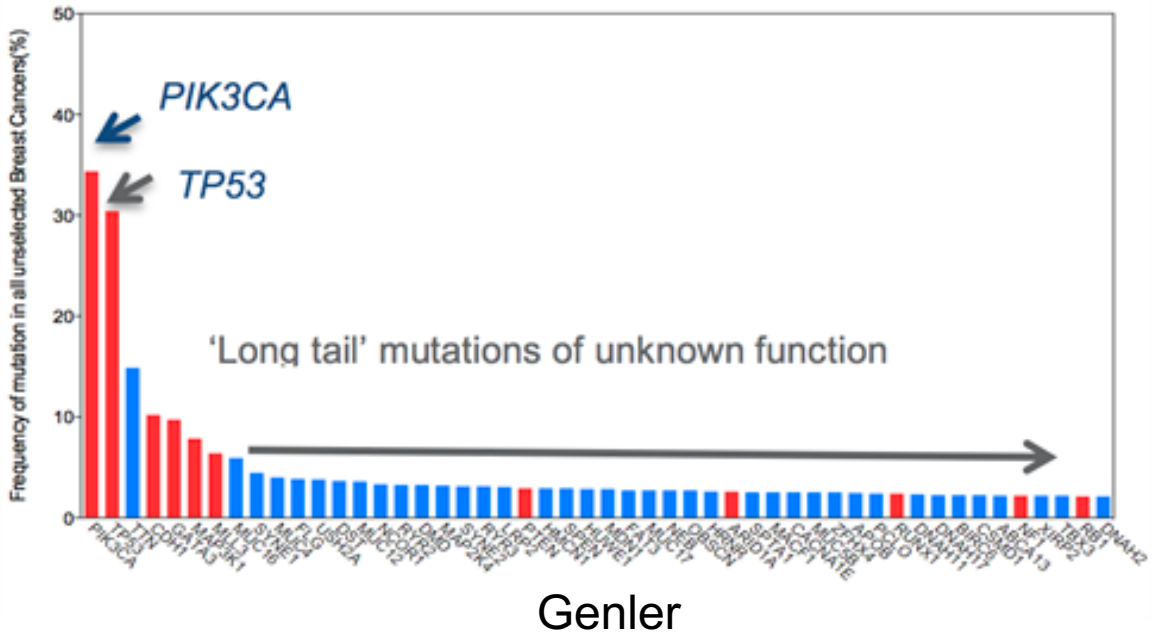


Figure kaynağı
http://www.pancreaticcancer.net.au/ground-breaking-nature/

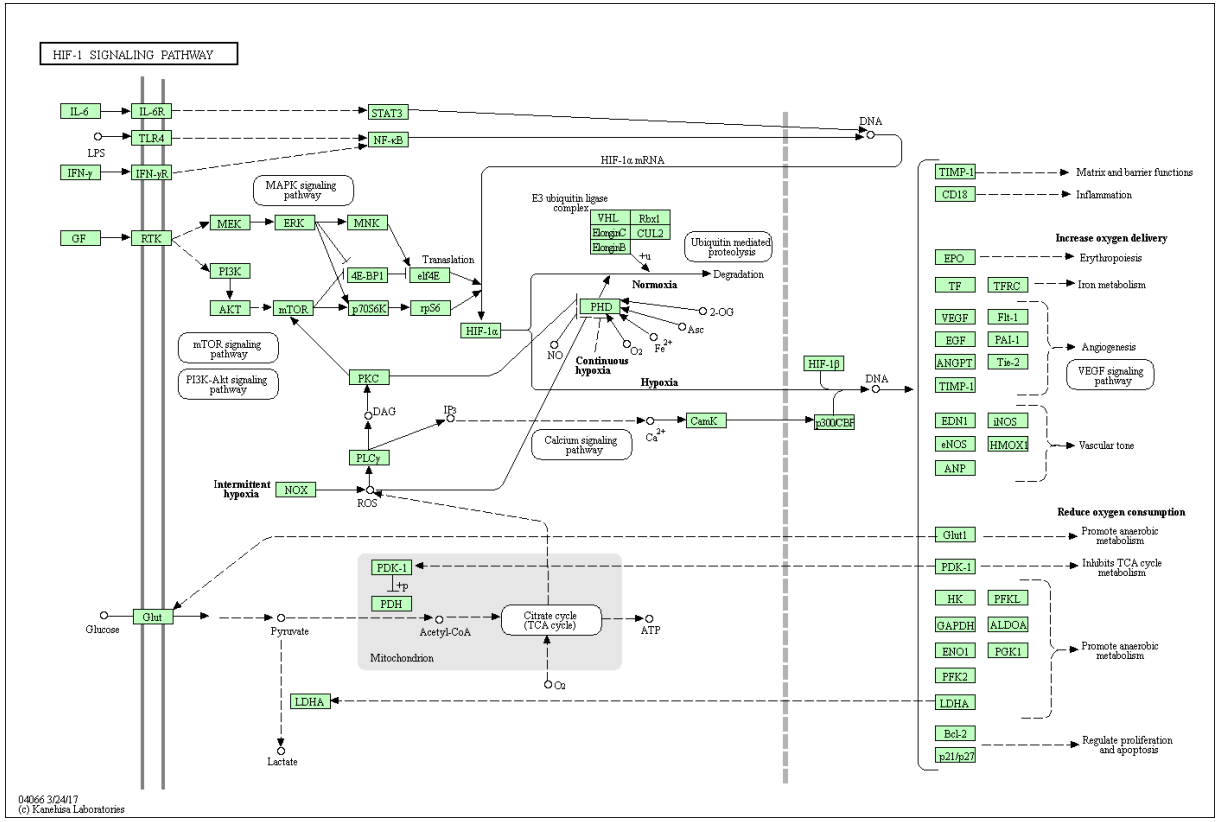
Ortak Değişimler Az



Figür kaynağı https://www.icr.ac.uk/

Biyolojik Yolaklar

- Molekül etkileşim ağları
- Çeşitli hücresel mekanizmalar hakkında biyolojik bilgileri özetler



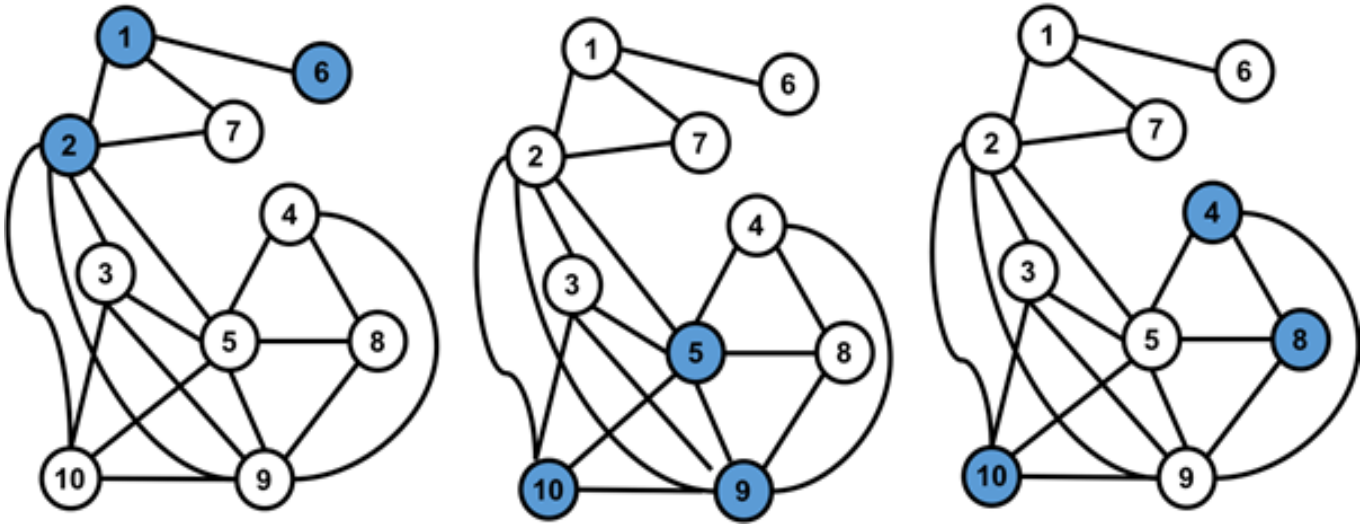
KEGG -- HIF-1 Sinyal Yolağı

Yolaklar

- Moleküler değişimlere hücresel ağlar bağlamında bakmak hasta benzerliklerini ortaya çıkarabilir.

Aynı Yolak, Farklı Hastalar

Hasta 1 **Hasta 2** **Hasta 3**



Çizge Çekirdekleri

İki çizge birbirine ne kadar benzerdir?

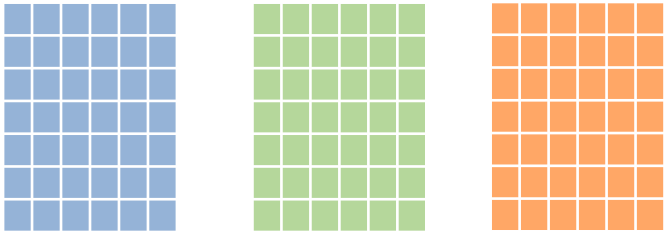
- Çizgeleri kümeleme ya da sınıflandırma gibi çizgeler üzerinde yapay öğrenme problemleri için temel bir sorudur.

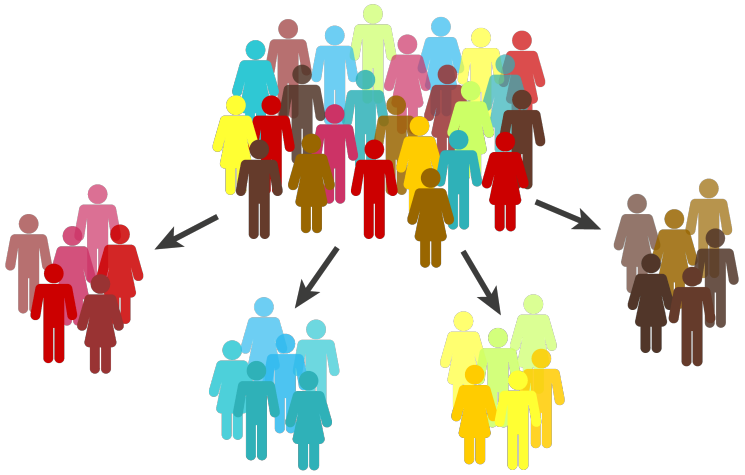
- **Çizge üzerindeki yolların benzerliği**
  - Shortest Path Graph Kernel (Borgwardt et al. 2005, ICDM)
  - Graph Hopper Kernel (Feragen et al. 2013, NIPS)

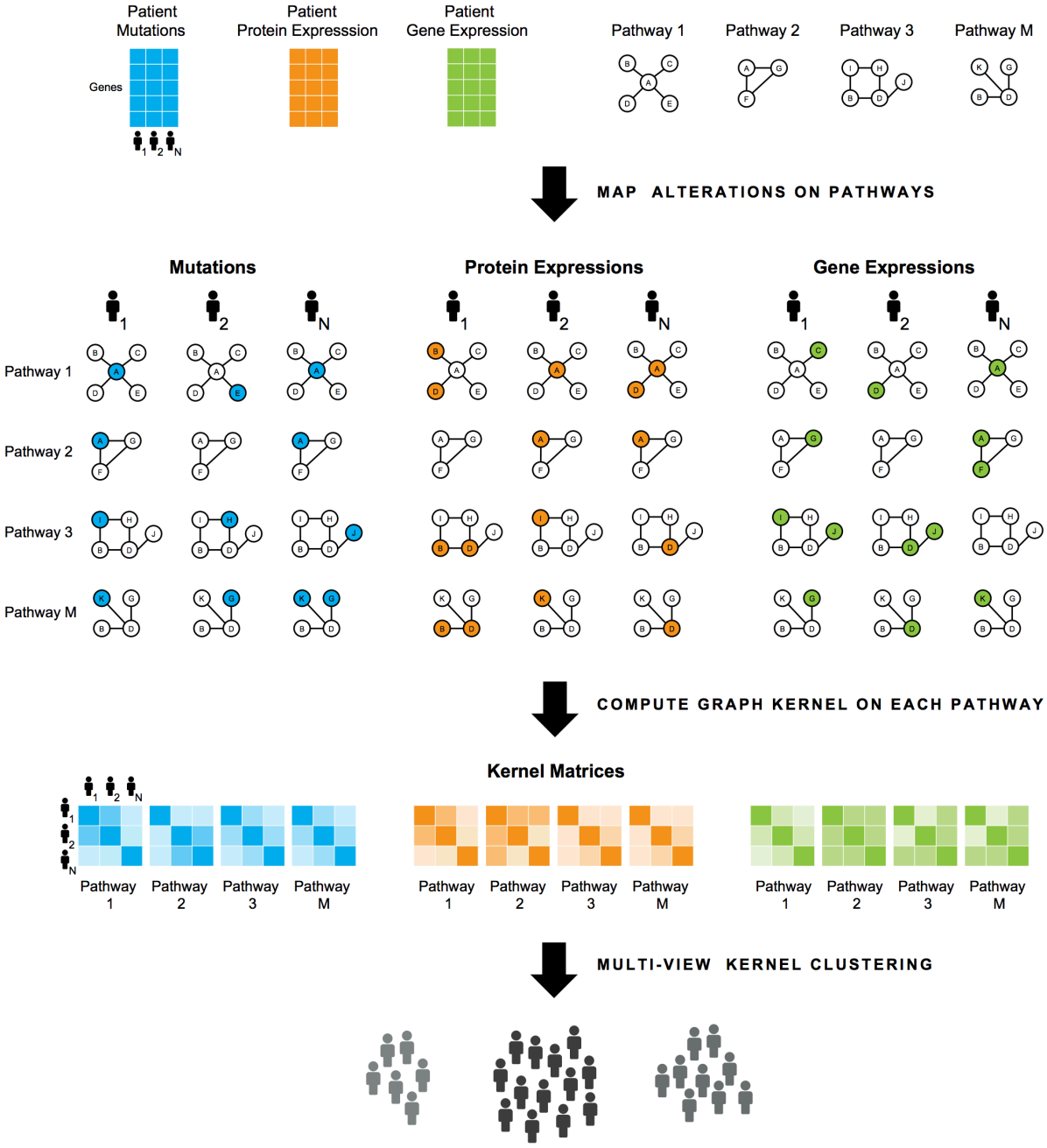
- **Düğüm etiket dağılım benzerliği**:
  - Propagation Graph Kernel (Neumann et al. 2015, Machine Learning)
  - Wasserstein Weisfeiler Lehman Graph Kernel (Togninalli et al. 2019, NeurIPS)

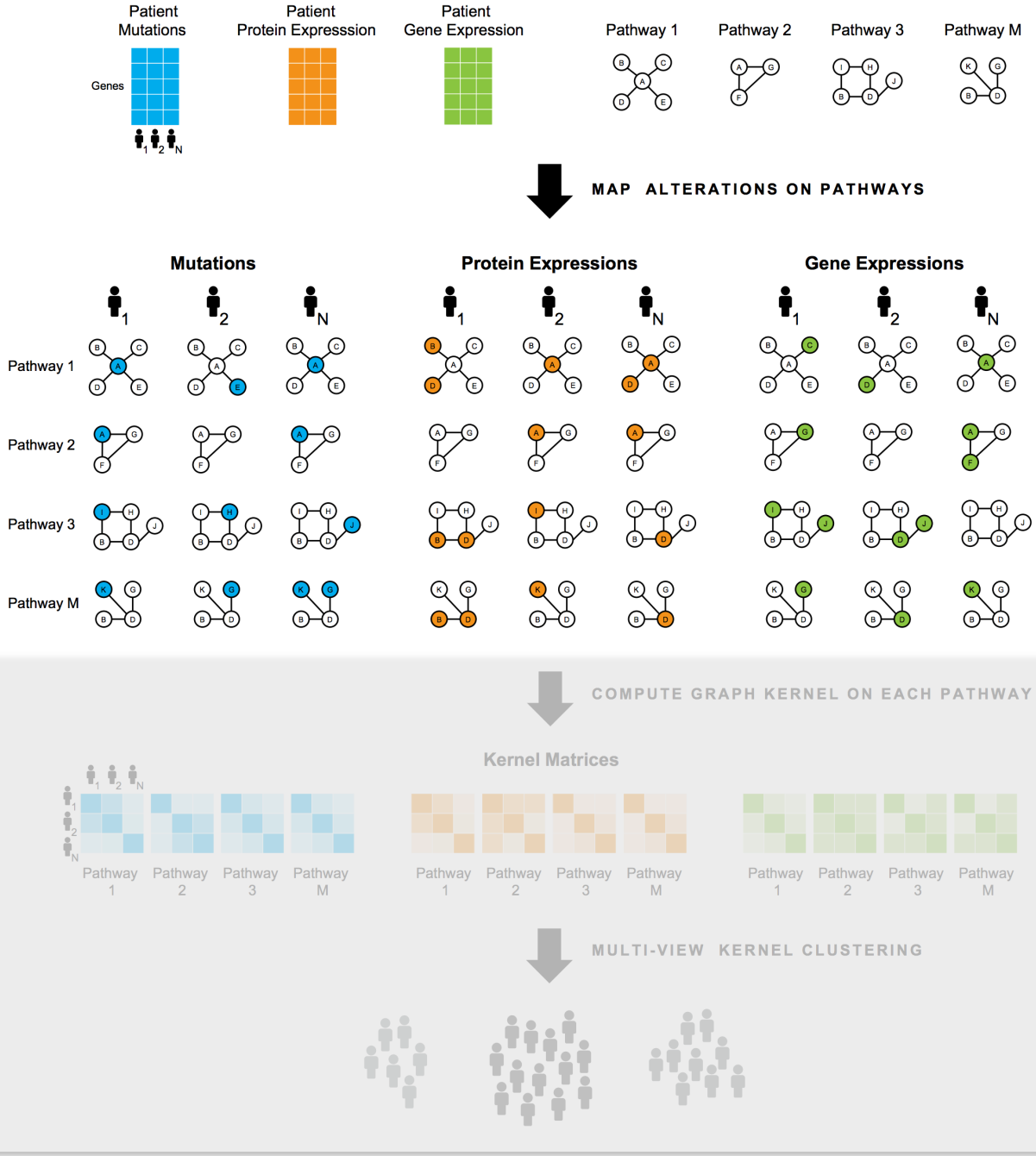
PAMOGK Ana Fikir

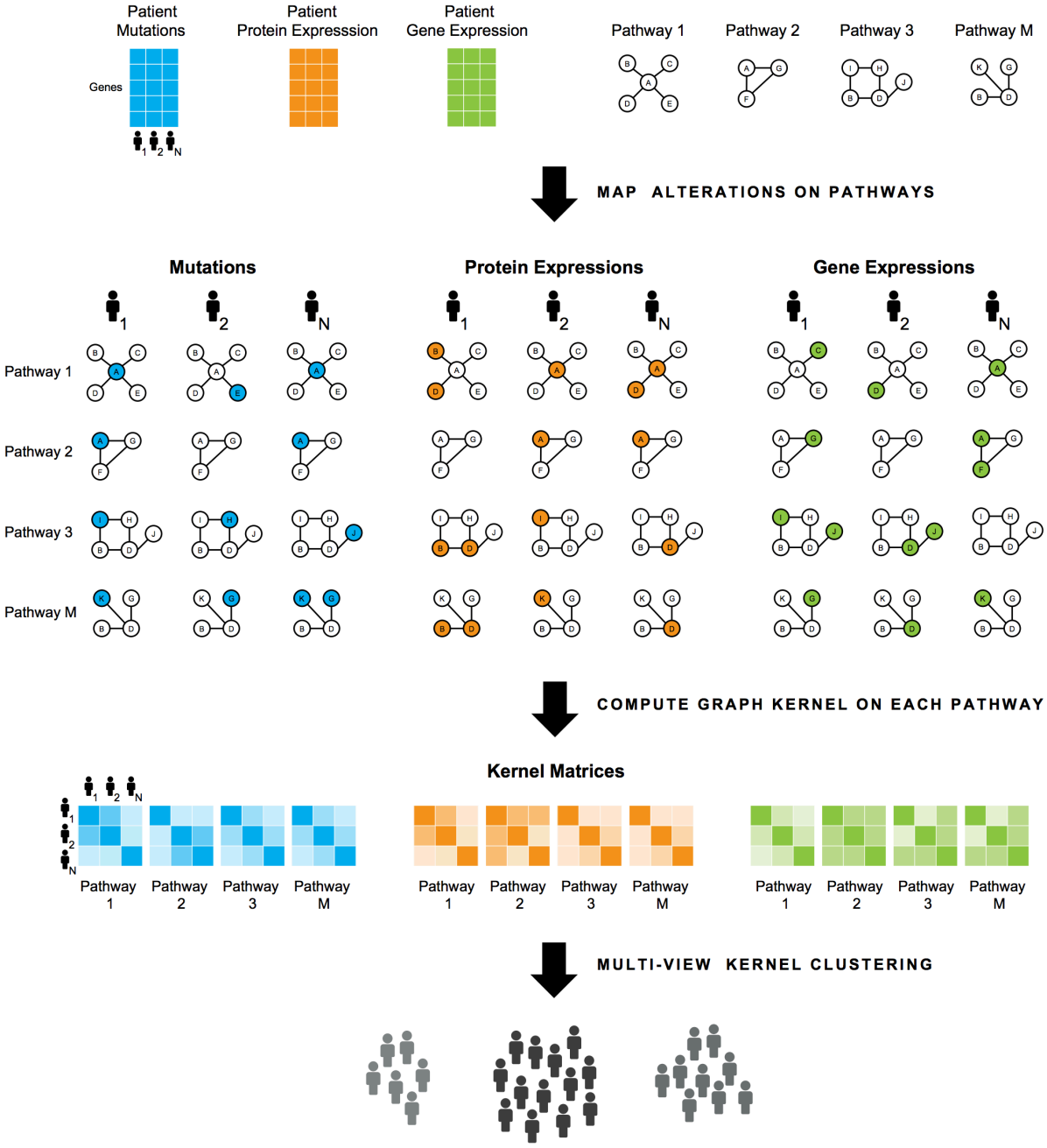
- Hem farklı omik verilerinden hem farklı yolaklardan gelen bakış açılarını entegre ederek hasta alt gruplarını bulmak

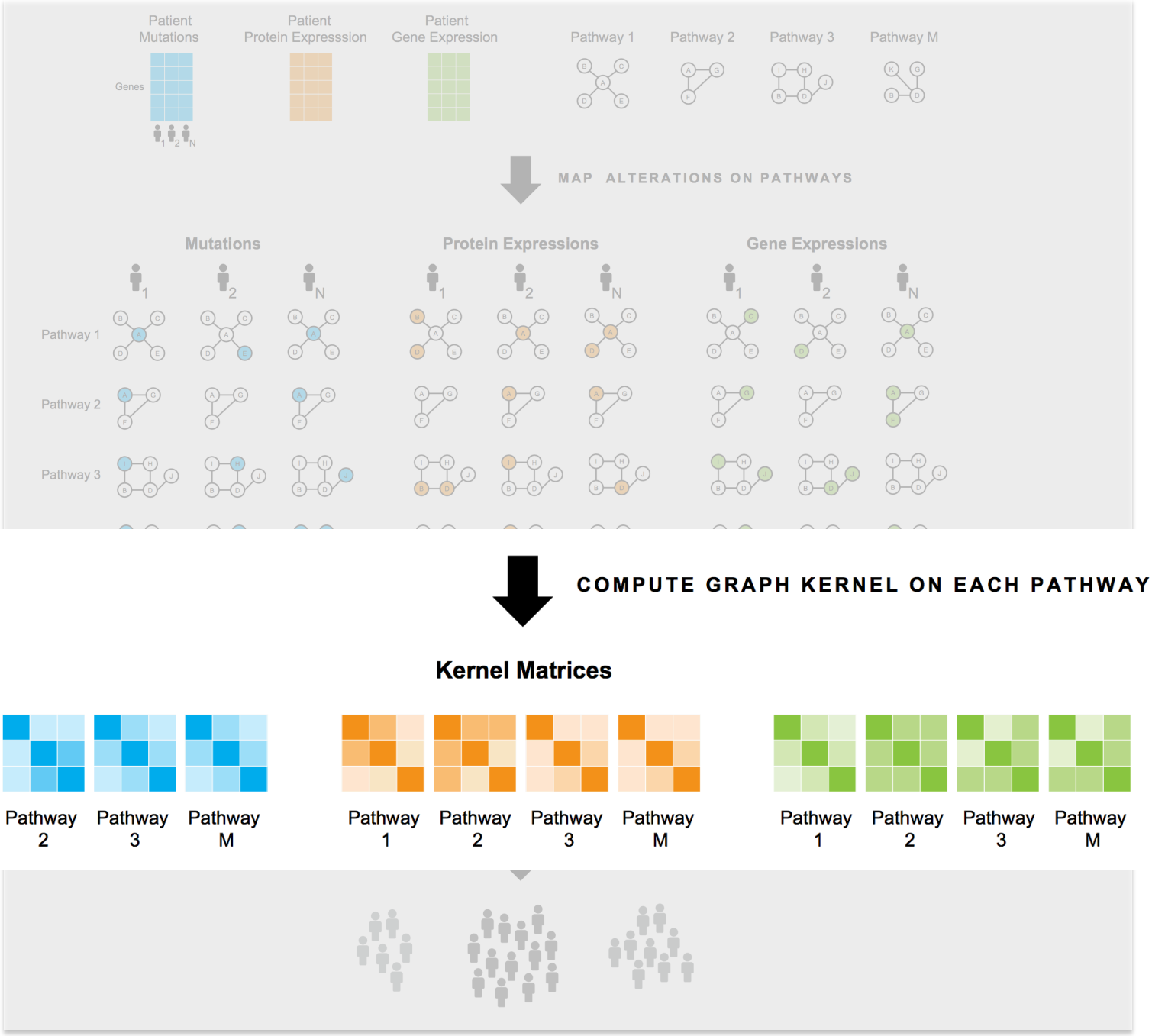


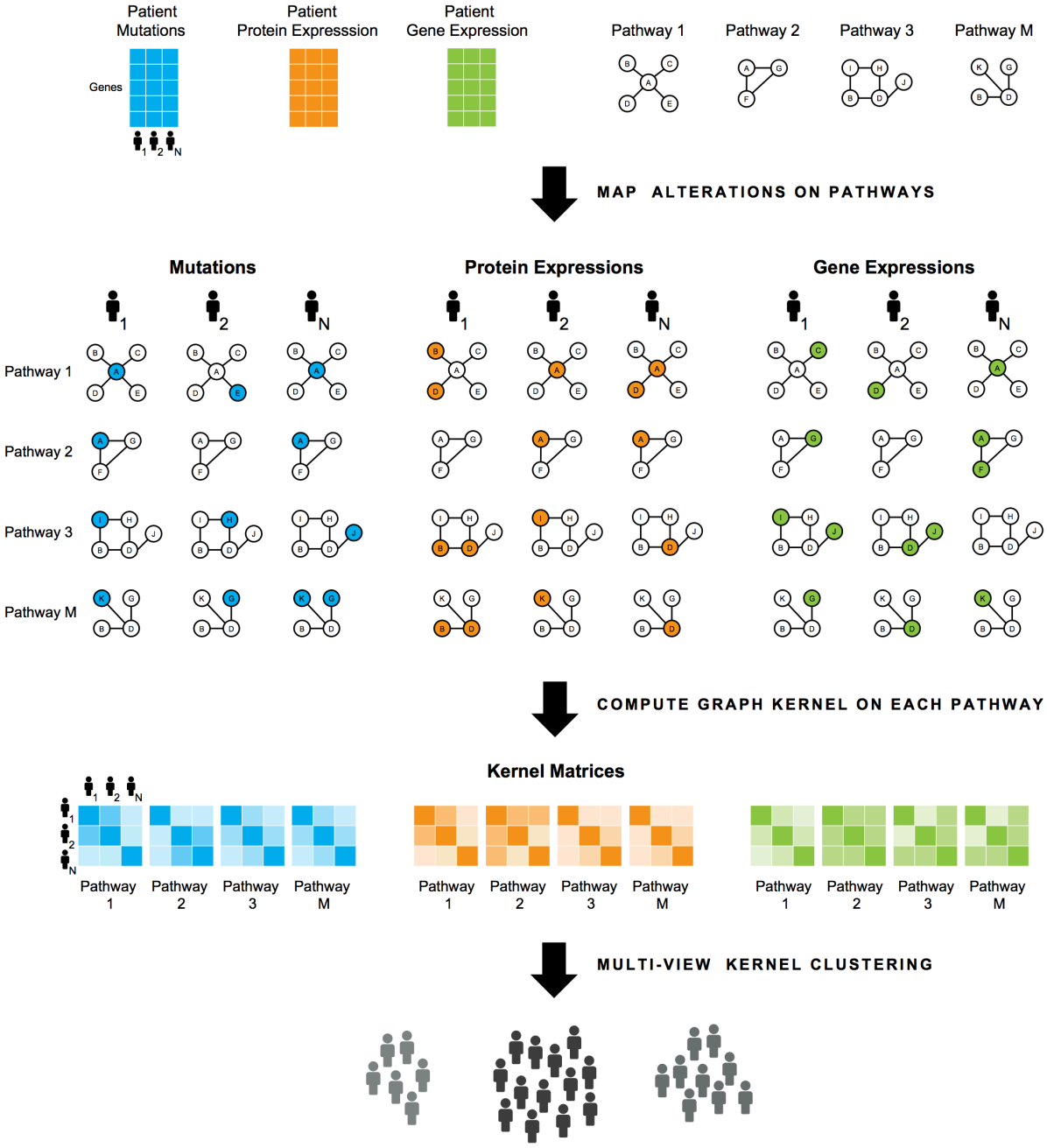



Patient Mutations
Patient Protein Expresssion
Patient Gene Expression
Genes
1
2
N
Pathway 1
Pathway 2
Pathway 3
Pathway M
MAP ALTERATIONS ON PATHWAYS
Mutations
Protein Expressions
Gene Expressions
Pathway 1
Pathway 2
Pathway 3
Pathway M
COMPUTE GRAPH KERNEL ON EACH PATHWAY
Kernel Matrices
Pathway 1
Pathway 2
Pathway 3
Pathway M
MULTI-VIEW KERNEL CLUSTERING

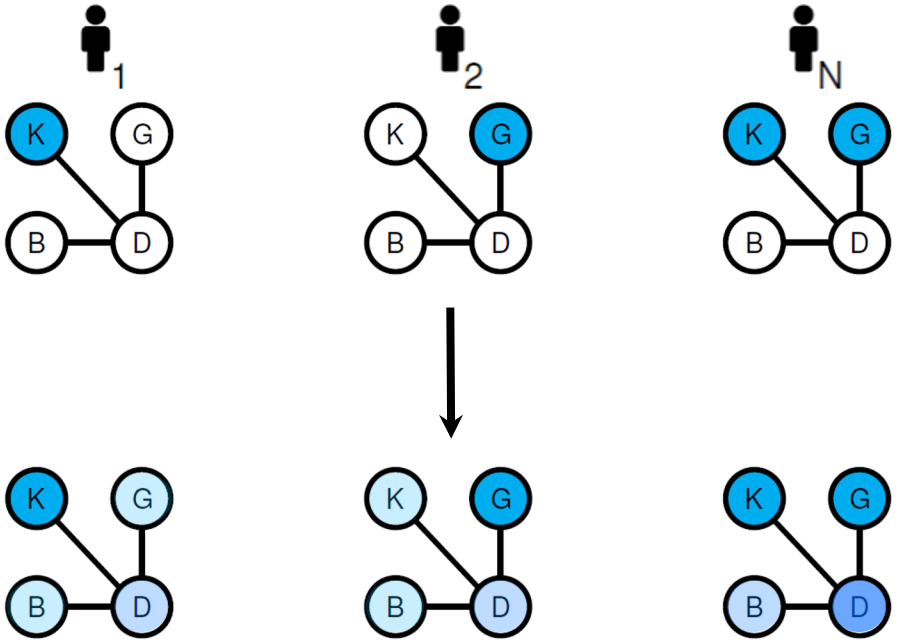

Patient Mutations
Patient Protein Expresssion
Patient Gene Expression
Genes
1 2 N
Pathway 1
Pathway 2
Pathway 3
Pathway M
MAP ALTERATIONS ON PATHWAYS
Mutations
Protein Expressions
Gene Expressions
Pathway 1
Pathway 2
Pathway 3
Pathway M
COMPUTE GRAPH KERNEL ON EACH PATHWAY
Kernel Matrices
Pathway 1
Pathway 2
Pathway 3
Pathway M
MULTI-VIEW KERNEL CLUSTERING


Patient Mutations
Patient Protein Expresssion
Patient Gene Expression
Genes
1
2
N
Pathway 1
Pathway 2
Pathway 3
Pathway M
MAP ALTERATIONS ON PATHWAYS
Mutations
Protein Expressions
Gene Expressions
Pathway 1
Pathway 2
Pathway 3
Pathway M
COMPUTE GRAPH KERNEL ON EACH PATHWAY
Kernel Matrices
Pathway 1
Pathway 2
Pathway 3
Pathway M
MULTI-VIEW KERNEL CLUSTERING


Patient Mutations
Patient Protein Expresssion
Patient Gene Expression
Genes
Pathway 1
Pathway 2
Pathway 3
Pathway M
MAP ALTERATIONS ON PATHWAYS
Mutations
Protein Expressions
Gene Expressions
Pathway 1
Pathway 2
Pathway 3
COMPUTE GRAPH KERNEL ON EACH PATHWAY
Kernel Matrices
Pathway 1
Pathway 2
Pathway 3
Pathway M


Patient Mutations
Patient Protein Expresssion
Patient Gene Expression
Genes
1
2
N
Pathway 1
Pathway 2
Pathway 3
Pathway M
MAP ALTERATIONS ON PATHWAYS
Mutations
Protein Expressions
Gene Expressions
Pathway 1
Pathway 2
Pathway 3
Pathway M
COMPUTE GRAPH KERNEL ON EACH PATHWAY
Kernel Matrices
Pathway 1
Pathway 2
Pathway 3
Pathway M
MULTI-VIEW KERNEL CLUSTERING

Düzgünleştirilmiş Kısa Yol Çizge Çekirdeği: Etiketleri Yayma Adımı

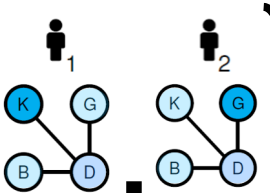


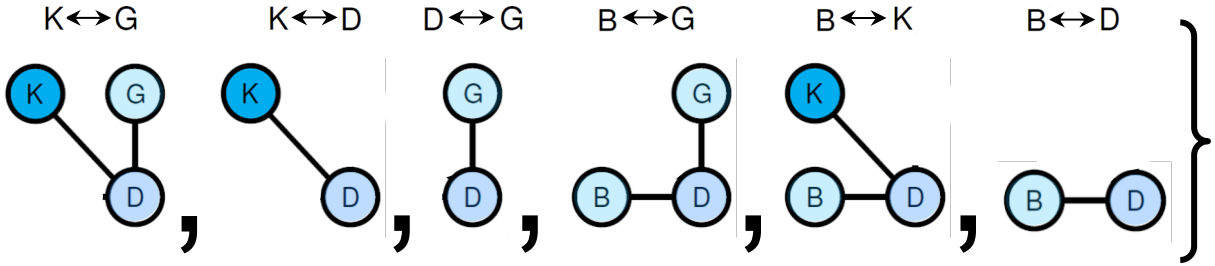
Etiketleri yolak çizgeleri üzerinde komşulara dağıt

$$\mathbf{S}_g^{(t+1)} = \alpha \mathbf{S}_g^{(t)} \mathbf{A}_g + (1-\alpha)\mathbf{S}_g^{(0)}$$

Network Propagation
(Cowen et al. 2017, Nature Reviews Genetics)

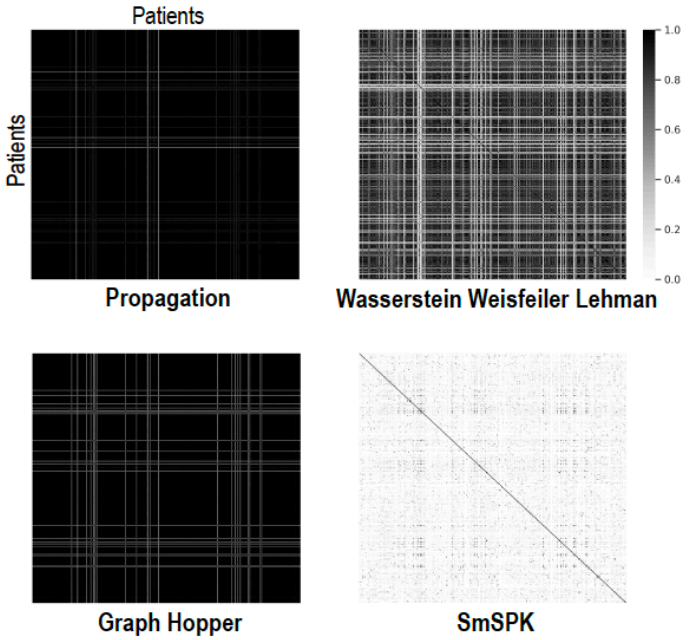
Düzgünleştirilmiş Çizge Çekirdeği Hesaplaması

$$\mathcal{K}\left(\text{  }\right) = \sum_{p=1}^{P} \mathbf{s}_p^{(1)} \cdot \mathbf{s}_p^{(2)}$$

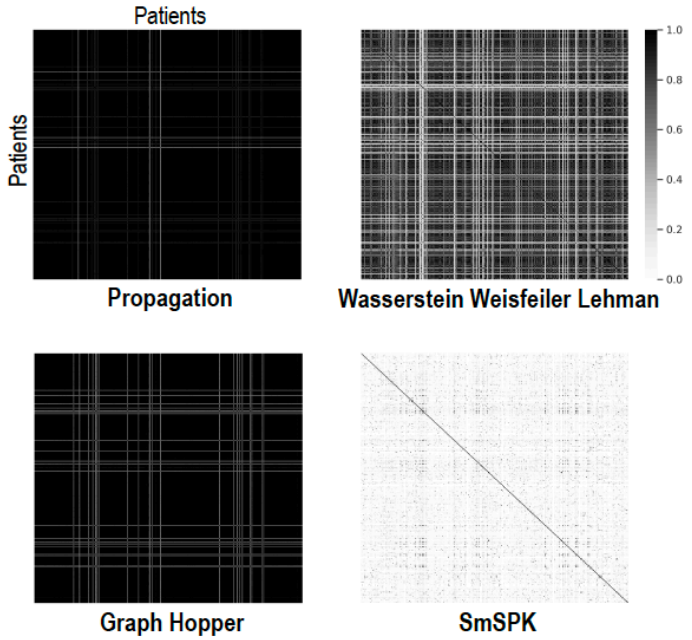
$$P = \left\{ \text{  } \right\}$$

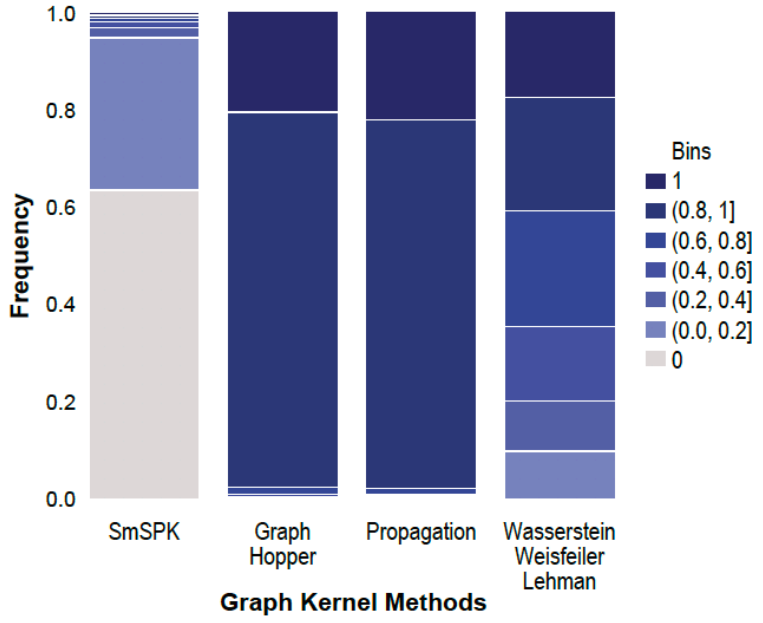
All pairs of shortest paths

Yeni bir Çizge Çekirdeğine İhtiyaç Var mı?



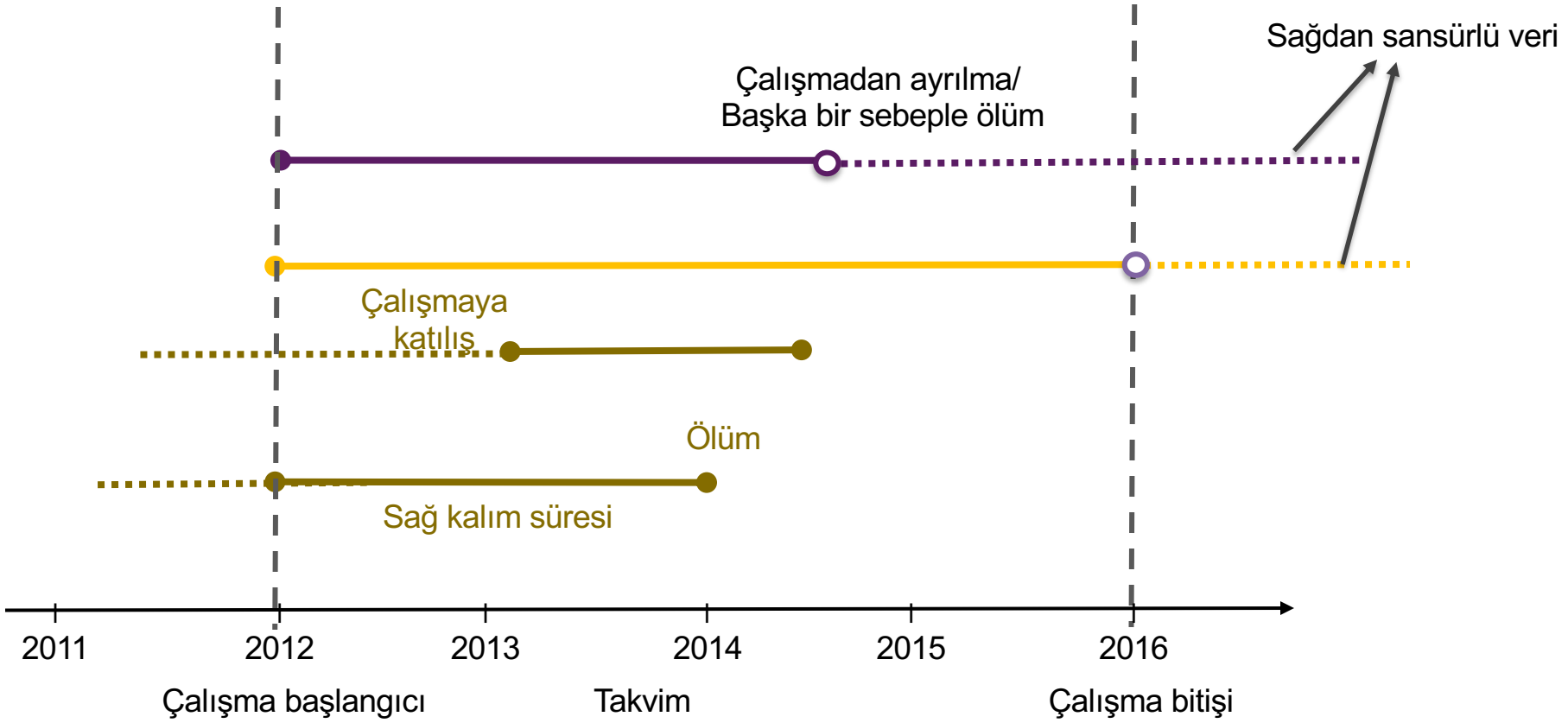
Yeni bir Çizge Çekirdeğine İhtiyaç Var mı?



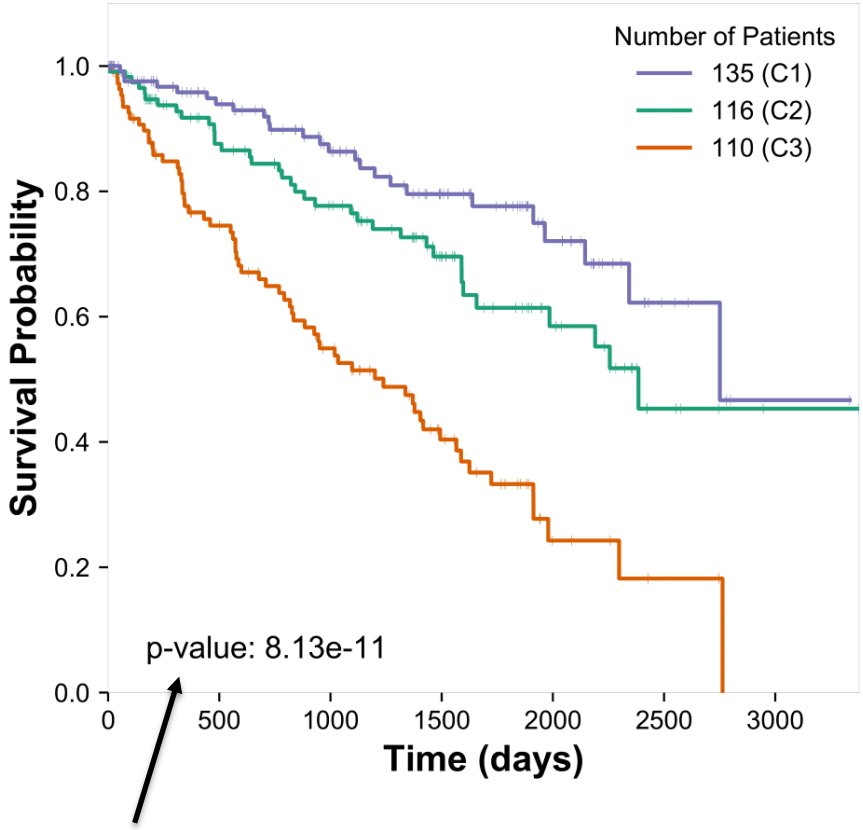


Diğer çizge çekirdekleri çizge yapısındaki benzerliklere odaklandıkları için hastları hep benzer hesaplıyor.

Öbeklemeyi Değerlendirme
Sağkalım Verisi



Öbeklemeyi Değerlendirme

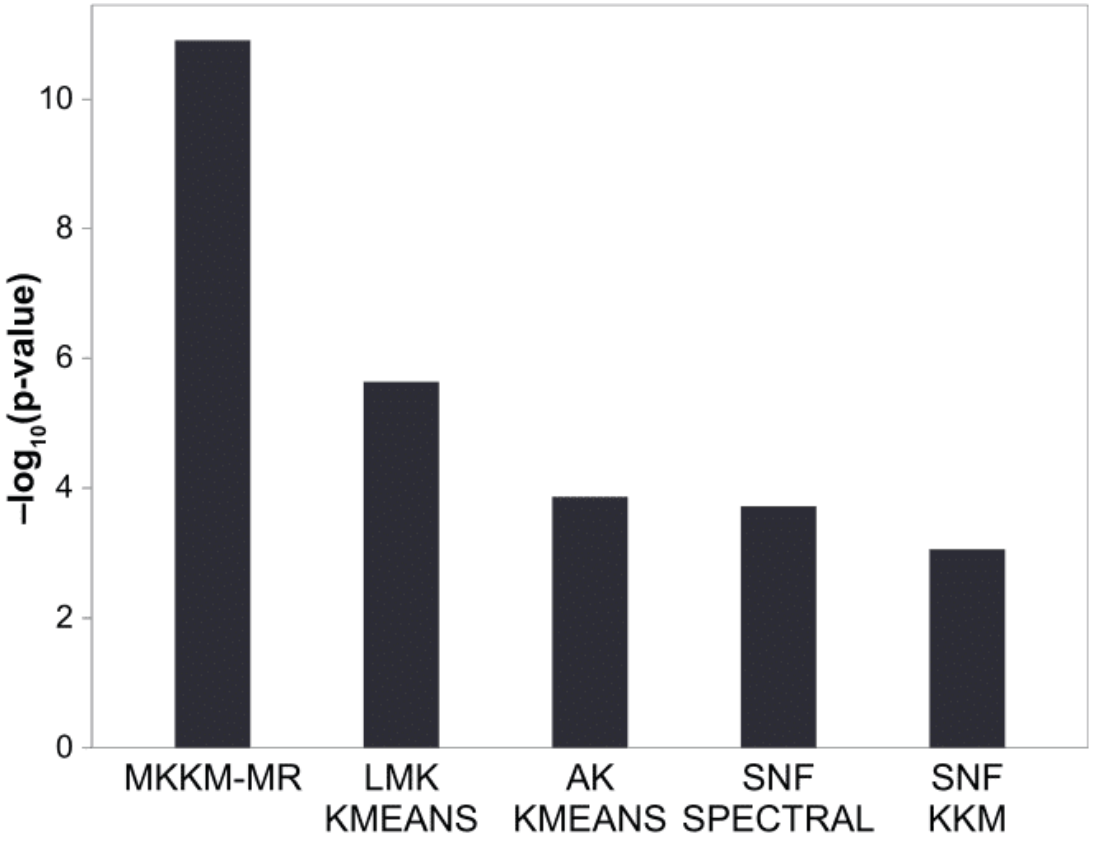


Farkı gruplardaki hastaların sağkalım dağılımlarının biribirinden ne kadar ayrıldığını için log-rank testinin döndürdüğü p değeri.

Çoklu Bakış Açılı Kümele Adımı için Alternatif Metotlar

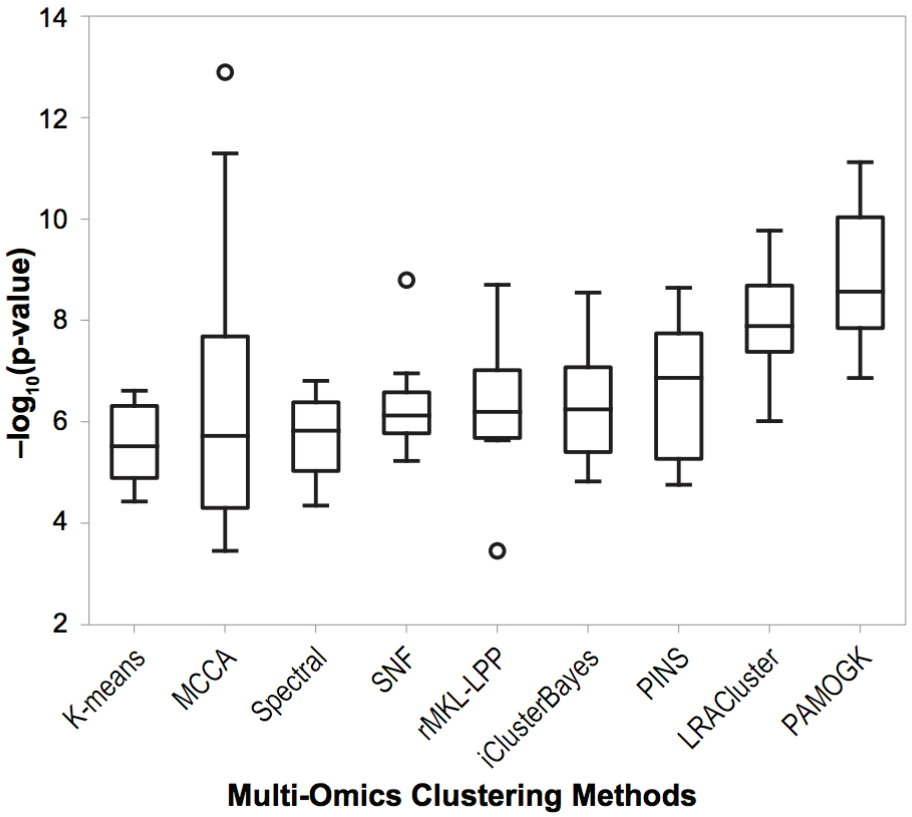
- Average Kernel K-Means (**AKKM**)
- Multiple Kernel K-Means with Matrix Induced Regularization (**MKKM-MR**) (Liu et al. 2016, AAAI'16)
- Localized Multiple Kernel K-Means (**LMKKM**)(Gönen et al. 2016, AAAI'16)
- Spectral Clustering (**SNF K-Means**) (Wang et al. 2014, Nature Methods)
- Kernel K-Means (**SNF Spectral**) (Wang et al. 2014, Nature Methods)

Farklı Çoklu-Bakış Açılı Metotlardan Hangisini Kullanmalıyız



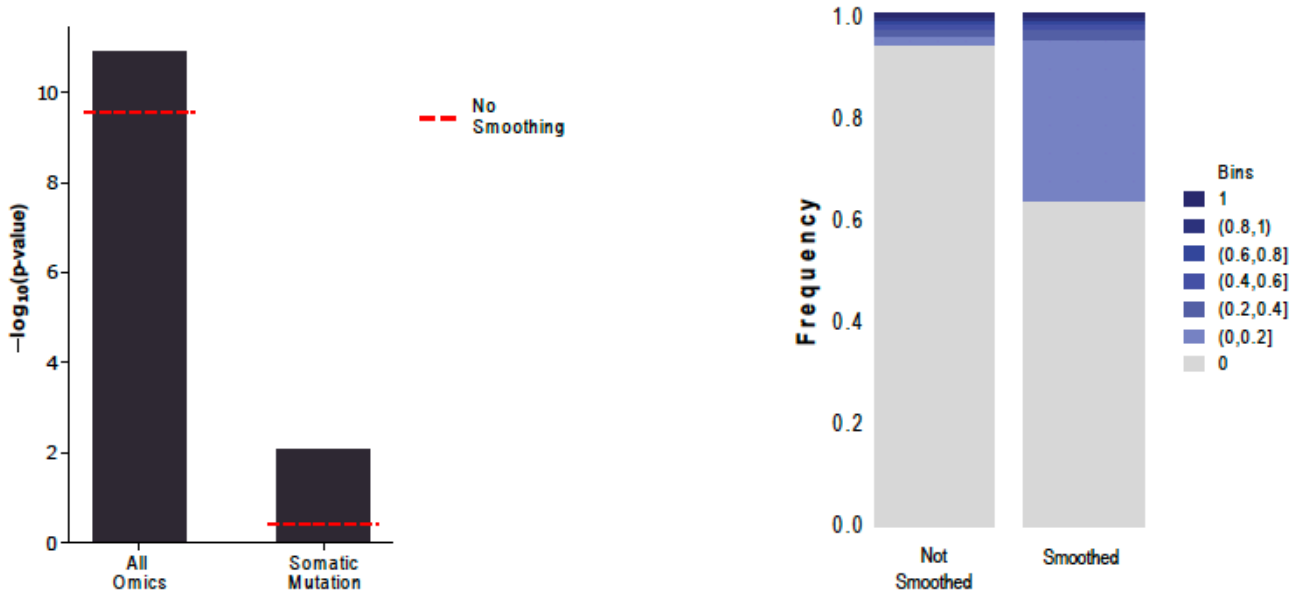
Bu problemde MKKM-MR en iyi gruplamayı yapan çoklu bakış açılı, çekirdekli algoritma oldu.

Diğer Metotlarla Karşılaştırma



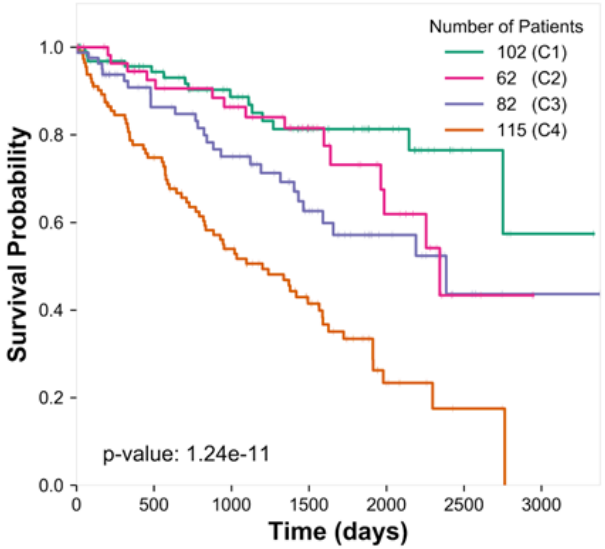
Benchmark from: Rappaport, Shamir. 2018, Nucleic Acid Research

Etiketleri Yaymanin Etkisi

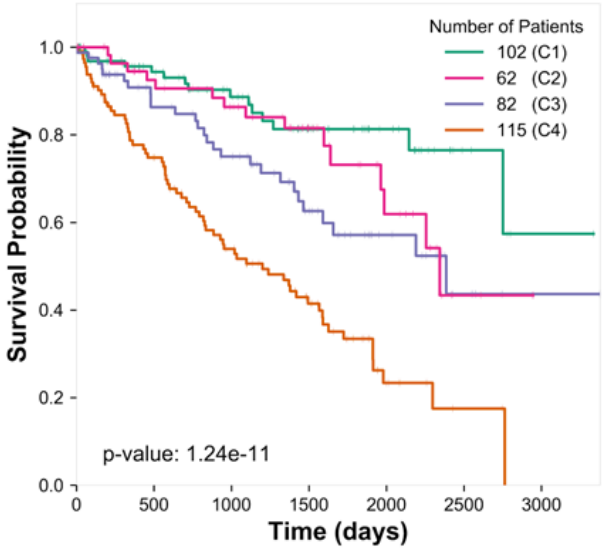


Moleküler değişiklikleri çizge üzerinde yaymak işe yarıyor, özellikle mutasyon verilerinden yararlanmada.

Bulunan KIRC Hasta Grupları

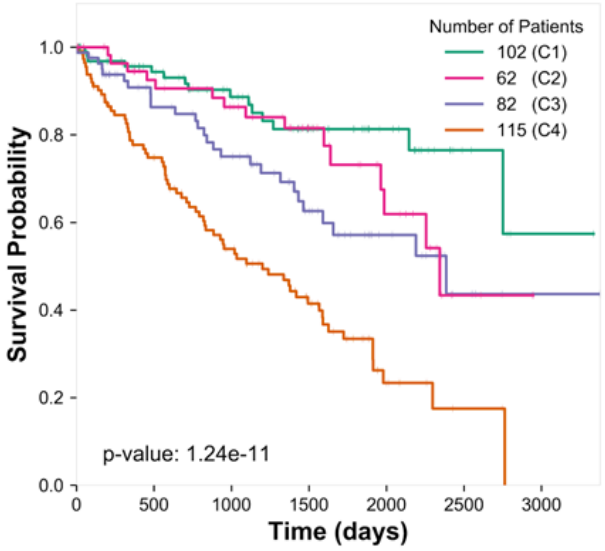


Bulunan KIRC Hasta Grupları

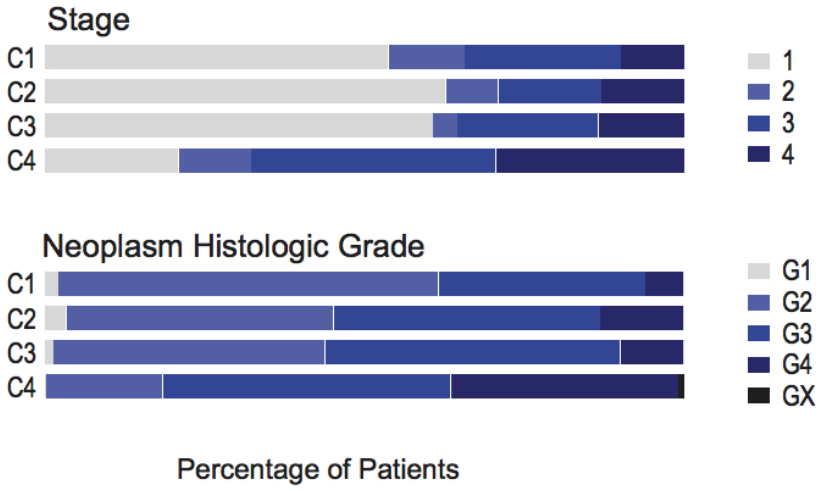


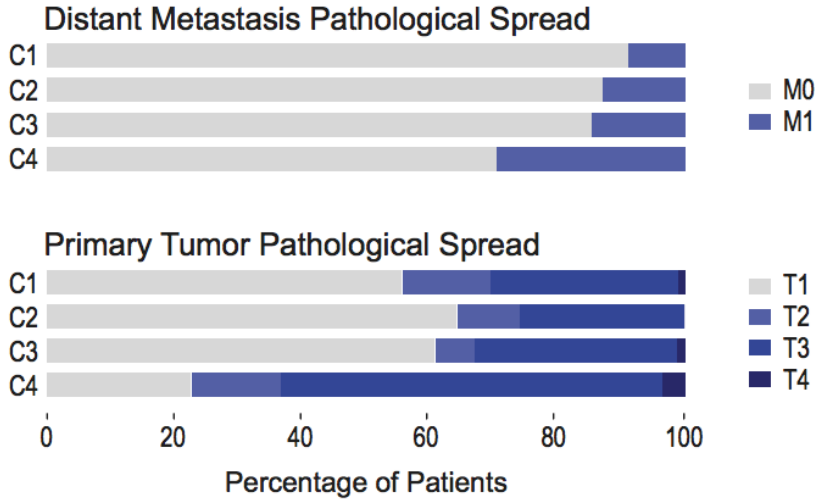
Clinical Parameter	Test	p-value
Age	One-Way ANOVA	2.200e-01
Gender	X^2	4.080e-01
Tumor Stage	X^2	3.476e-08
Primary Tumor Pathological Spread	X^2	1.349e-07
Distant Metastasis Pathological Spread	X^2	3.766e-04
Neoplasm Histologic Grade	X^2	2.104e-09

Bulunan KIRC Hasta Grupları



Clinical Parameter	Test	p-value
Age	One-Way ANOVA	2.200e-01
Gender	X^2	4.080e-01
Tumor Stage	X^2	3.476e-08
Primary Tumor Pathological Spread	X^2	1.349e-07
Distant Metastasis Pathological Spread	X^2	3.766e-04
Neoplasm Histologic Grade	X^2	2.104e-09





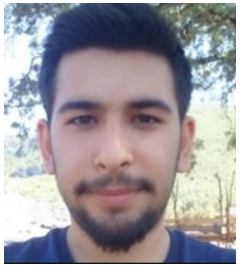
Sonuç

- Çizge çekirdekleri çoklu omik verilerine biyolojik ağları entegre etmede fayda getirir
- Farklı ağlar kullanılabilir
- Etkileşimlerin yönleri ve tipleri entegre edilebilir

Teşekkürler

TUBITAK) **3501 Career Award - Grant #117E140**

TUBITAK-BIDEB 2210 Fellowship Program



Yasin İ. Tepeli
Sabancı Universitesi

Ali Burak Ünal
Önceden : Bilkent Üniversitesi
Şimdi Tübingen Universitesi

M. Furkan Akdemir
Bilkent Üniversitesi

Bütün **TastanLab Üyeleri**

PAMOGK

Available at:
github.com/tastanlab/pamogk

otastan@sabanciuniv.edu