

# Disiplinler Arası Bir Bilim Dalı: Biyoinformatik

A.Ceren Akın, Büşra Bürçe, Burak Çevirici, Bengisu Şahin, Eda Şahin, Yağmur Şahin

Danışman: Doç. Dr. A. Canan Yazıcı

## ÖZET

Biyoinformatik, biyolojik sistemler ve olaylardan elde edilen verinin değerlendirilmesi amacıyla biyoloji, tıp bilimleri, bilişim teknolojileri, matematik ve biyoistatistiğin entegrasyonu sonucu doğan yeni ve disiplinler arası uygulamalı bir bilim dalıdır. Başlangıçta biyolojik dizilerin analizi için geliştirilmiş olan biyoinformatik günümüzde yapısal biyoloji, genomik çalışmalar ve gen ekspresyon çalışmaları, rasyonel ilaç geliştirme, tanı ve tedavi dahil olmak üzere birçok konuyu geniş bir yelpazede kapsamaktadır. Temel hedefi biyolojik deneyler ve gözlemsel çalışmalardan elde edilen verilerin yönetimi, işlenebilir halde depolanması, analizi ve yorumlanmasına olanak sağlamak olan biyoinformatik, bilişim teknolojileri ve biyoistatistik biliminden yoğun bir biçimde yararlanmaktadır. Bu çalışma ile biyoinformatik biliminin ortaya çıkışı, amaçları, oluşumunda yer alan bilim dalları ele alınarak uygulama sürecine ilişkin temel bilgiler verilmiştir.

**Anahtar Kelimeler:** Biyoinformatik, Biyoistatistik, Genom, Veri Tabanı, Bilişim Teknolojileri

## 1. GİRİŞ

Günümüzde biyolojik veriler olağanüstü bir hızda artış gösterirken bu verilerin kullanılabilir hale getirilerek işlenmesi ihtiyacı, bilgi teknolojisine dayanan, çok geniş bir çalışma alanına sahip multidisipliner bir bilim dalının ortaya çıkışını sağlamıştır. Biyoinformatik ve bilişimsel biyoloji çalışmaları, çeşitli tipteki biyolojik bilgilere kolay erişilmesini, bu bilgilerin verimli kullanımı ve yönetimi için gerekli bilgisayar yazılımlarının geliştirilmesini ve bu çalışmalar ile büyük veri kümeleri içindeki ve arasındaki ilişkilerin çeşitli algoritmalar ve biyoistatistiksel yöntemler ile açığa çıkarılmasını kapsamaktadır (2, 4). Biyoinformatik alanında yapılan çalışmalarda, özellikle büyük ölçekteki biyolojik verilerin yönetim ve analizi amacıyla, bilgisayar bilimleri ve biyoistatistik gibi disiplinlerin entegrasyonundan ortaya çıkan bilişim teknikleri biyolojik veriye uygulanmaktadır. Bu alanda yapılan çalışmalarda ağırlıklı olarak biyomoleküllerle ilgili çeşitli analiz yöntemlerinin geliştirilmesi ve heterojen veri kaynaklarının bütünleştirilmesi hedeflenmektedir (9, 14).

## 2. BİYOİNFORMATİĞİN TARİHÇESİ

Tıptaki arařtırmaların ilerlemesi ve yeni yöntemlerin kullanılması merak edilen yeni soruların kaynađı olmuřtur. Uzun süredir bilim adamları birçok organizmanın eksiksiz genom dizisini elde etmek için uğrařmaktadır. Biyoinformatik bilimi de bu çabaların sonucu olarak geliřmiřtir (8). Pauling ve Corey'in 1951 yılında proteinlerin sekonder yapılarının dođru tahmini için geliřtirdikleri yaklařım biyoinformatik için bařlangıç kabul edilebilir. Ancak asıl anlamda biyoinformatik biliminin bařlangıcı olarak, 1966 yılında bilgisayarla moleküler grafiklerin çizimine ait ilk makalenin Scientific American dergisinde yayımlanması olarak kabul edilebilir. Bununla beraber "Biyoinformatik" terimi 1980'li yılların ortalarından sonra kullanılmaya bařlanmıřtır. Temel moleküler ve genetik süreçlerin anlaşılmasında ve karmařık verilerin analizi ve yorumlanması için yeni yöntemler geliřtirilmesinde en etkin kurum olan "National Center for Biotechnology Information (NCBI)" 1988'de kurulmuřtur. Biyoinformatik alanındaki en önemli projelerden biri olan İnsan Genom Projesi (İGP) ise Ekim 1990 da bařlamıřtır. Projenin amaçları:

- ✓ İnsan DNA'sındaki yaklařık 20.000-25.000 genin tanımlanması,
- ✓ İnsan DNA'sını oluřturan yaklařık 3 milyar kimyasal baz çiftinin diziliminin belirlenmesi,
- ✓ DNA ve protein bilgilerinin veri tabanlarında tutulması, fonksiyonlarının önceden bilinmesi, hastalıkların nüksetmeden tedavi edilmesinin sađlanması,
- ✓ Veri analizi için araçlar geliřtirilmesi,
- ✓ İlgili teknolojilerin özel sektöre transfer edilmesi,
- ✓ Bu projeden ortaya çıkabilecek, yasal, etik ve sosyal konuların ele alınması olarak özetlenebilir.

İGP çalıřmaları biyoinformatiđin geliřiminde çok önemli bir yere sahiptir. ABD Enerji Bakanlığı ve Ulusal Sađlık Enstitüsü tarafından koordine edilen 13 yıllık bu proje 2003 yılında tamamlanmıřtır. Projeye İngiltere, Japonya, Fransa, Almanya gibi ülkeler de katkıda bulunmuřlardır. Bařlangıçta 2005 yılında bitirilmesi planlanan proje, teknolojiye gerçekeřen geliřmelerle birlikte beklenenden iki yıl önce tamamlanmıřtır. Ancak dizi verilerindeki hataların ayıklanmasının ve veri analizlerinin yapılmasının bir süre daha devam edeceđi düşünölmektedir (5, 10, 13).

Gerçekeřtirilen diđer genom projeleri ile 1000'den fazla organizmanın gen dizileri elde edilmiřtir. Böylece, farklı canlı türlerinin gen dizilimlerini karřılařtıran çalıřmalar da yapılmaya bařlanmıřtır (6).

## 3. BİYOİNFORMATİĞİN TEMEL HEDEFLERİ

Biyoinformatiđin hedefleri üç temel bölümden oluřmaktadır. İlk hedefi, mevcut verilerin arařtırmacıların ulařabileceđi ve yenilerini oluřturduđu da yükleyebileceđi şekilde düzenlenmesidir. Diđer bir ifadeyle veri tabanı oluřturma'dır. Veri toplamak titizlik gerektiren ve kapsamlı bir iř olmakla beraber, kaydedilmiř veriler analiz edilmeden önce tamamen kullanıřsız

olarak kabul edilir. Bu açıdan biyoinformatik sonraki süreçte gözönüne alındığında veri toplama hedefinden çok daha fazlasına ulaşır.

İkinci hedefi, veri analizinde kullanılan araçları ve kaynakları geliştirmektir. Örneğin; aminoasit dizisi belirli bir proteinin, dizi özellikleri belirlenmiş başka bir proteinle karşılaştırılması basit bir yazılım araştırmasından fazlasını gerektirir. Mesela moleküllerin biyolojik içerikleri açısından incelemeler yapılabilirdir. Örneğin, BLAST (Basic Local Alignment Search Tool) aynı ya da farklı organizmalar arasında, nükleotid dizisi ya da protein dizisi karşılaştırmasını yapan bir algoritmadır. Bu tip yazılım kaynaklarının gelişmesi, biyoloji bilgisi gerektirdiği kadar, bilişim alanında da uzmanlık gerektirmektedir. Günümüzde biyoinformatiğin temel hedefi sadece daha etkili veri depolama araçları sağlamak değil, hedefe daha doğru yaklaşan analiz araçları tasarlamak amacıyla bilgileri mantığa uygun olarak düzenlemektir. Bilgisayar bilimi ve biyoloji arasındaki beklenmedik birlik, hayatın kendisinin bir bilgi teknolojisi olduğu gerçeğini göstermektedir (9).

Üçüncü hedef ise elde edilen bilgileri biyolojik açıdan anlamlı bir şekilde analiz etmek ve yorumlamaktır (9). Biyoinformatik bu amaçla biyoistatistikten faydalanmaktadır. Toplanan veriyi düzenleyerek analiz edilebilir hale getirmek, bilinen faktörlere göre değerlendirmek, değişkenler arası ilişkileri belirlemek ve modellemek, veri içindeki örüntüleri bulmak, sonuçları yorumlamak ve genelleştirmek için biyoistatistik yöntemler kullanılmaktadır.

Biyoinformatiğin genel hedeflerini kısaca aşağıdaki gibi özetleyebiliriz:

- Biyolojik problemlerin çözümünde bilişim teknolojilerinin kullanılması
- Moleküler dizilimlerin incelenmesi, her hangi bir biyolojik fonksiyonu artıran veya azaltan moleküllerin tasarlanması
- Üç boyutlu yapısal moleküler veri tabanları oluşturulması
- Veri madenciliği ve metin madenciliği çalışmaları
- Protein dizi çalışmaları, protein yapı ve fonksiyonunun belirlenmesi protein-protein ilişkilerinin analiz edilmesi
- Biyolojik olarak aktif moleküllerin araştırılması
- Bakteri, maya, insan, hayvan, bitki genom projelerinden elde edilen bilgilerin depolanması ve analizi
- DNA sıra ve dizileme çalışmaları
- Makromoleküllerin yapılarının üç boyutlu dizilimlerinin araştırılması
- Tıbbi ya da endüstriyel amaçlı yeni makromoleküller üretilmesi
- Biyolojik bilginin paylaşılmasının kolaylaştırılması
- Bilgisayar ile otomatize edilmiş veri analizi
- Biyolojik faaliyet süreçlerinin simülasyonu
- Karmaşık genetik fonksiyon veya regülasyonun tanımlanması
- Genetik faktörlerin hastalık yatkınlığına etkilerinin ortaya çıkarılması

Kısacası, biyoinformatiğin hedefleri; hızla gelişen ve artan bilgiyi toplamak, yönetmek, dağıtmak, bilgiye en hızlı ve kolay biçimde ulaşılmasını

sağlamak, çok karmaşık yapıda olan biyolojik sistemleri incelemek ve çözebilmektir (7, 13).

#### **4. BİYOİNFORMATİK UYGULAMALARI**

Bu bölümde biyoinformatiğin başlıca uygulama alanlarından bazıları ele alınmıştır.

##### **4.1. Homologları Bulma**

Biyoinformatiğin arkasındaki itici güçlerden biri farklı biyomoleküller arasındaki benzerlikleri aramasıdır. Sistemik veri organizasyonu sağlamadan bağımsız olarak, protein türdeşlerini tanımlamada bazı pratik kullanımları vardır. En belirginini, ilişkili proteinler arasında bilgi aktarımıdır. Örneğin karakteristiği tam olarak belirlenememiş bir protein için onun daha iyi anlaşılması homologlarını bularak elde edilen yeni bilgiler değerlendirilir. Biyokimyasal ya da yapısal veri yetersiz kaldığında, çalışmalar maya gibi düşük seviyeli organizmalarda başlatılabilir ve sonuçlar insan gibi üst düzey organizmalardaki homologlarda uygulanır (3, 9, 12).

##### **4.2. Rasyonel İlaç Geliştirilmesi**

Biyoinformatiğin tıbbi çalışmalardaki uygulamaları rasyonel ilaç tasarımına yardımcı olmuştur. Mesela nükleotid sekansı verilen, kodlanmış proteinin muhtemel amino asit sekansı transkripsiyon yazılımı kullanılarak belirlenebilir. Sekans arama teknikleri model olarak kullanılan organizmada homologları bulmak için kullanılabilir ve bu çalışma sekans benzerliğine dayalı olarak gerçekleşir. Bu deneysel çalışmalar, başka organizmalarda insan proteininin yapısını modellemeyi mümkün kılar. Son olarak, bağlantı algoritmaları gerçek protein üzerindeki biyolojik etkinliği test etmek için, biyokimyasal tahlillere öncülük edebilir ve modellenen proteinin yapısına bağlanabilen molekülleri tasarlayabilir (3, 9, 12). Gen ekspresyon analizleri ile farklı hastalıklardan etkilenen hücrelerin ekspresyonları derlenerek, sağlıklı hücrelerinki ile kıyaslanmakta ve aradaki farklılıklardan hastalık teşhisi ve hedef ilaç dizaynında yararlanılmaktadır.

##### **4.3. Veri Tabanları**

Biyoinformatiğin en önemli işlevlerinden biri, çoklu veri kaynaklarından elde edilen bilgilerin birleştirilebilmesine olanak sağlamasıdır. Örneğin, bir proteinin üç boyutlu yapı tahmini; proteinin işlevi, farklı genomlarda bulunması ve diğer moleküllerle etkileşimi ile ilgili verilerle birlikte incelendiğinde daha faydalı olur. Ancak dizinlerdeki ve dosya biçimlerindeki farklılıklardan dolayı bu bilgi kaynaklarına erişmek ve verimli kullanmak konusunda sıkıntılar yaşanabilmektedir.

Temel düzeyde, bu sorunun çözümüne yönelik olarak birçok veri kaynağına erişim sağlanabilecek şekilde veri kaynaklarını birleştirme çabası vardır. Bu noktada, iki önemli erişim sisteminden söz edilebilir. Bunlardan biri düz dosya veri tabanlarının birbirlerine endekslenmesine

olanak sađlayan Dizi Eriřim Sistemidir (SRS, Sliding Rail System). Bu sistem; kullanıcıya nkleik asit, protein dizisi, protein motifi, protein yapısı ve bibliyografik veri tabanlarına eriřim, bađlantı ve bunları kullanım imkanı sađlamaktadır. Diđeri de DNA'ya, protein dizilerine, genom harita verilerine, 3D makromolekler yapılara ve PubMed bibliyografik veri tabanına benzer yollarla eriřim sađlayan Entrez sistemidir (3, 9, 12).

Kısa srede ok yksek miktarlarda veri reten, endstri dzeyinde gen ekspresyonu, protein-protein iliřkisi, biyolojik olarak aktif molekl arařtırmaları, bakteri, maya, hayvan ve insan genom projeleri gibi biyolojik deneylerin dođurduđu talepler sonucunda, bu alandaki biliřim uygulamaları neredeyse takip edilemez bir hızda geliřmiřtir.

Bilginin toplanması, depolanması ve eriřimine olanak sađlayan pek ok veri tabanı ve veri bankası mevcuttur. Bunlardan bazıları ařađıdaki gibi zetlenebilir:

1. Bibliyografik Veri Tabanları: Metin Madenciliđi temeline oturtulmuř genel literatr veri tabanı tipleridir. MEDLINE, PUBMED, rnek olarak verilebilir.
2. Taksonomik Veri Tabanları: Organizmaların sınıflandırılmasını merkezileřtirmek isteyen dizi esasına ve hiyerarřiye dayanır. TBİTAK veri tabanları taksonomik alıřmalarda kullanılmaktadır.
3. Nkleotid Dizisi Veri Tabanları: Nkleikasit dizi verileri depolanmaktadır. BLASTN, BALSTX gibi veri tabanları nkleotid dizilerini belirlemek iin kullanılmaktadır.
4. Genomik Veri Tabanları: Mikro biyologların ve genetikilerin ilgilendikleri organizmalar iin genlerin veya mutasyonların yayımlanmıř katalogları bulunmaktadır. Bu alıřmada kullanılan en popler veri tabanı NCBI'dır. Ayrıca genom dizilerinin yanında protein yapı tahminlerini de ieren bir veri tabanıdır. UCSC de ilk genom taramasını sađlayan veri tabanıdır. Bununla birlikte ENSEMBL aktif olarak kullanılır.
5. Protein Dizisi Veri Tabanları: zel bir organizmanın, protein veya protein grupları ya da zel ailelerle ilgili bilgi depolayan veri toplayıcılarının uzmanlařması ve btn trlerdeki proteini kapsayan evrensel veri tabanları arasındaki farkı ayırt etmek iin gereklidir. Protein dizileri iin SWISSPROT, protein yapısı iinde PDB veri tabanı kullanılır.
6. DNA Veri Bankaları: GenBank(USA), EMBL (Europe), DDBJ (Japan) gibi veri tabanları ve veri bankaları kullanılmaktadır.

Bunlara ilaveten proteomik kaynakları, organel veritabanları, bitki veri tabanları, immünolojik veri tabanları, insan genleri ve hastalıkları veri tabanları gibi daha pek çok veri tabanı günümüzde kullanılmaktadır (13).

## **5. BİYOİNFORMATİK VE DİĞER DİSİPLİNLER**

Biyoinformatik alanına istatistik bilimi ve bilgisayar bilimleriyle uğraşanlar önemli katkıda bulunmaktadır. Temel hedefi biyolojik deneyler ve gözlemsel çalışmalardan elde edilen verilerin yönetimi, depolanması, analizi ve yorumlanmasına olanak sağlamak olan biyoinformatik bu amaçla fazlasıyla biyoistatistikten faydalanmaktadır. Bu zamana kadar çeşitli istatistiksel yöntemler geliştirilmiş ve biyoinformatikte uygulanmıştır. Destek vektör makinelerine (Vopnik, 1995) dayalı yeni yaklaşımlar (Ewens ve Grant, 2005) öne sürülmüştür. Bunlar, biyoinformatikteki verilerin çoğuyla ve bilgisayar ile karakterizedir. Bu yaklaşımlarla veri üretme yöntemleri ve veri analizi için gerekli istatistiksel gelişmeler yakından takip edilmektedir. Örneğin; fonksiyonel genomik verilerin büyük miktarlarda kullanılabilir bilgi haline dönüştürülmesi biyoistatistikle gerçekleştirilir ve bunlar hayati öneme sahiptir (1, 6).

Program geliştirme, yeni algoritmaların geliştirilmesi, ağ yönetimi ve güvenliği, görselleştirme ve grafik tasarım, çoklu ortam uygulamaları, veri analizi ve yorumu çalışmaları çok açık göstermektedir ki biyoinformatikte bilgisayar bilimleri ve biyoistatistik temel yapı taşlarıdır. Optimal sonuca ulaşabilmek için biyoinformatikte, moleküler biyoloji, genetik, diğer tıp bilimleri, biyoistatistik ve bilgisayar bilimleri arasında güçlü bir entegrasyon söz konusudur (11).

Moleküler biyoloji, Genetik, Sistem Biyolojisi, Protein Mühendisliği, Tıp, Farmasötik araştırmalar, Ekoloji/Populasyon genetiği gibi bilim dalları biyoinformatikten yoğun fayda sağlayan ve aynı zamanda da çözüm aradıkları soruların yarattığı itici güçle bu bilimin ortaya çıkışına neden olan ve gelişimine katkı sağlayan bilim dallarıdır.

## **6. SONUÇLAR**

Biyoinformatik, biyolojik araştırmalara yeni bir boyut ve derinlik kazandırmıştır. Biyoinformatik, canlı sistemleri ayrıntılı olarak incelemeyi aynı zamanda da pek çok sistem arasındaki geçerli olan ortak ilkeleri ortaya çıkarmayı ve bazı sistemlerin kendine özgü özelliklerini vurgulamak amacıyla bunları diğer sistemler ile karşılaştırmayı mümkün kılmaktadır (2,9). Günümüzdeki hızlı veri artışı ile birlikte biyoinformatiğin bilim dünyasındaki yeri giderek önem kazanmakta ve biyolojik araştırmalar için vazgeçilmez hale gelmektedir. Başlangıçta biyolojik dizilerin analizi için geliştirilmiş olan biyoinformatik günümüzde yapısal biyoloji, genomik çalışmalar ve gen ekspresyon çalışmaları, rasyonel ilaç geliştirme, tanı ve tedavi dahil olmak üzere birçok konuyu geniş bir yelpazede kapsamaktadır. Biyoinformatik bilimindeki ilerlemelerin katkısıyla hastalıkların önlenmesinde ve tedavisinde önemli gelişmeler sağlanması beklenmektedir.

Bu alandaki arařtırmalar ve geliřmeler büyük bir hızla devam etmekte ve bu kapsamda disiplinler arası çalıřmalarda büyük bir önem kazanmaktadır.

## **KAYNAKLAR**

1. Arhipova I. The Role of Statistical methods in Computer Science and Bioinformatics. Seventh International Conference on Teaching Statistics (ICOTS-7) "Working Cooperatively in Statistics Education", Salvador, Brazil, July 2-7 2006;1-4
2. Atalay R.Ç, Neden Biyoinformatik?,Avrasya Dosyası,Moleküler Biyoloji ve Gen Teknolojileri Özel,Sonbahar 2002,Cilt 8,Sayı:3,129-141
3. Attwood T.K., Parry-Smith D.J., Introduction to Bioinformatics, Prentice Hall, Harlow, 1999;240
4. Baxevanis, A. D., Ouellette, B. F. F., Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 2nd Edition, John Wiley & Sons Inc., New York (NY); 2001:356.
5. Collins F.S., Morgan M., Patrinos A., The Human Genome Project: Lessons from Large-Scale Biology, Science 2003; 300;5617:286-290
6. Karabulut E, Karağaođlu E, Biyoinformatik ve Biyoistatistik, Hacettepe Tıp Dergisi, 2010; 41:162-170
7. Kumaresan V, Bhatt P, Palanisamy R, Gnanam AJ, Pasupuleti M, Arockiaraj J, A Murrel Cysteine Protease, Cathepsin L: Bioinformatics Characterization, Gene Expression and Proteolytic Activity, Biologia, 2014; 69(3): 395-406
8. Kuonen D, Challenges in Bioinformatics for Statistical Data Miners, Bulletin of the Swiss Statistical Society, 2003; 46:10-17
9. Luscombe NM, Greenbaum D, Gerstein M, What is Bioinformatics: A Proposed Definition and Overview of the Field, Method Inform Med, 2001; 4:346-358
10. Polat M, Karahan A.G, Multidisipliner Yeni Bir Bilim Dalı: Biyoinformatik ve Tıpta Uygulamaları, S.D.Ü. Tıp Fak. Derg., 2009; 16(3):41-50
- 11.Suarez C, Manteiga W, Statistics in Biomedical Research, ARBOR Ciencia, Pensamiento y Cultura, 2007; 725:353-361
12. Tandon P, Bhattacharjee P, Bioinformatics: An overview, Plant Tissue Culture and Molecular Markers, I.K. International Publishing House Pvt. Ltd., Green Park Extension Uphaar Cinema Market, New Delhi, India 2009;29-49.
- 13.[www.hasanbalik.com/dersler/eem101/10-11sunular/4-hafta-biyoinformatik.pdf](http://www.hasanbalik.com/dersler/eem101/10-11sunular/4-hafta-biyoinformatik.pdf)
- 14.[www.muhendislik.khas.edu.tr/biyoinformatik-ve-genetik/boeluem.-hakkinda-2.html](http://www.muhendislik.khas.edu.tr/biyoinformatik-ve-genetik/boeluem.-hakkinda-2.html)