



Bioinformatika és genomanalízis az orvostudományban

Bevezetés

Cserző Miklós

2018

A mai előadás

- A kurzus menete
- Hol találkozunk bioinformatikával
- Mi a bioinformatika
- Miért van bioinformatika
- A számítógépekről
- A bioinformatika szerepe és jövője
- Mivel foglalkozik a bioinformatika

A kurzus menete

- 13 előadásból áll a kurzus
- A félév 14. hetén előrehozott szóbeli vizsga
- Az előadásokhoz feladatok kapcsolódnak
- Aki 10, ill. 7 feladatra értékelhető megoldást küld – megajánlott jegyet kap
- Mail: cserzo.miklos@med.semmelweis-univ.hu
 - Az előadás anyagok letölthetők:
SE > Élettan > Oktatás > PhD kurzusok > ... Jelszó: eti3_403



A kurzus előadásai

1. Bevezetés
2. Biológiai adatbázisok
3. Szekvencia illesztés
4. Többszörös szekvencia illesztés
5. Keresés adatbázisokban
6. Integrált biológiai adatbázisok
7. Biológiai modellek
8. Szekvencia elemzés – elmélet
9. Szekvencia elemzés – gyakorlat (5)
10. Korreláció, klaszterezés
11. Nagy áteresztésű technikák
12. Rendszerbiológia
13. Molekula-modellezés alapjai (-)



Hol találkozhatunk bioinformatikával?

- Orvosbiológiai kutatás
 - Napi rutin
 - Bizonyos területek elképzelhetetlenek nélküle
- Kiemelt kórházak különleges részlegei
 - Bizonyos betegségek esetén
 - Diagnosztikában
 - Terápiában
- Körzeti orvosok, sarki gyógyszerészek?

Updated February 19, 2015: 23andMe provides ancestry-related genetic reports and uninterpreted raw genetic data only. We intend to add some health-related genetic reports in the future once we have a comprehensive product offering. At this time, we do not know which health reports might be available or when they might be available. For more information, please go to the [health](#) page. [Close Alert](#)

sign in

register kit



welcome

ancestry

research

how it works

buy

help



Find out what your DNA says about you and your family.

- Learn what percent of your DNA is from populations around the world
- Contact your DNA relatives across continents or across the street
- Build your family tree and enhance your experience with relatives

order now

\$99

★★★★★ 2.7 (1275)

Mi a bioinformatika

Biológiai vonatkozású adatok

- Gyűjtése
- Tárolása
- Kezelése
- Elemzése



Biológiai rendszerek viselkedésének

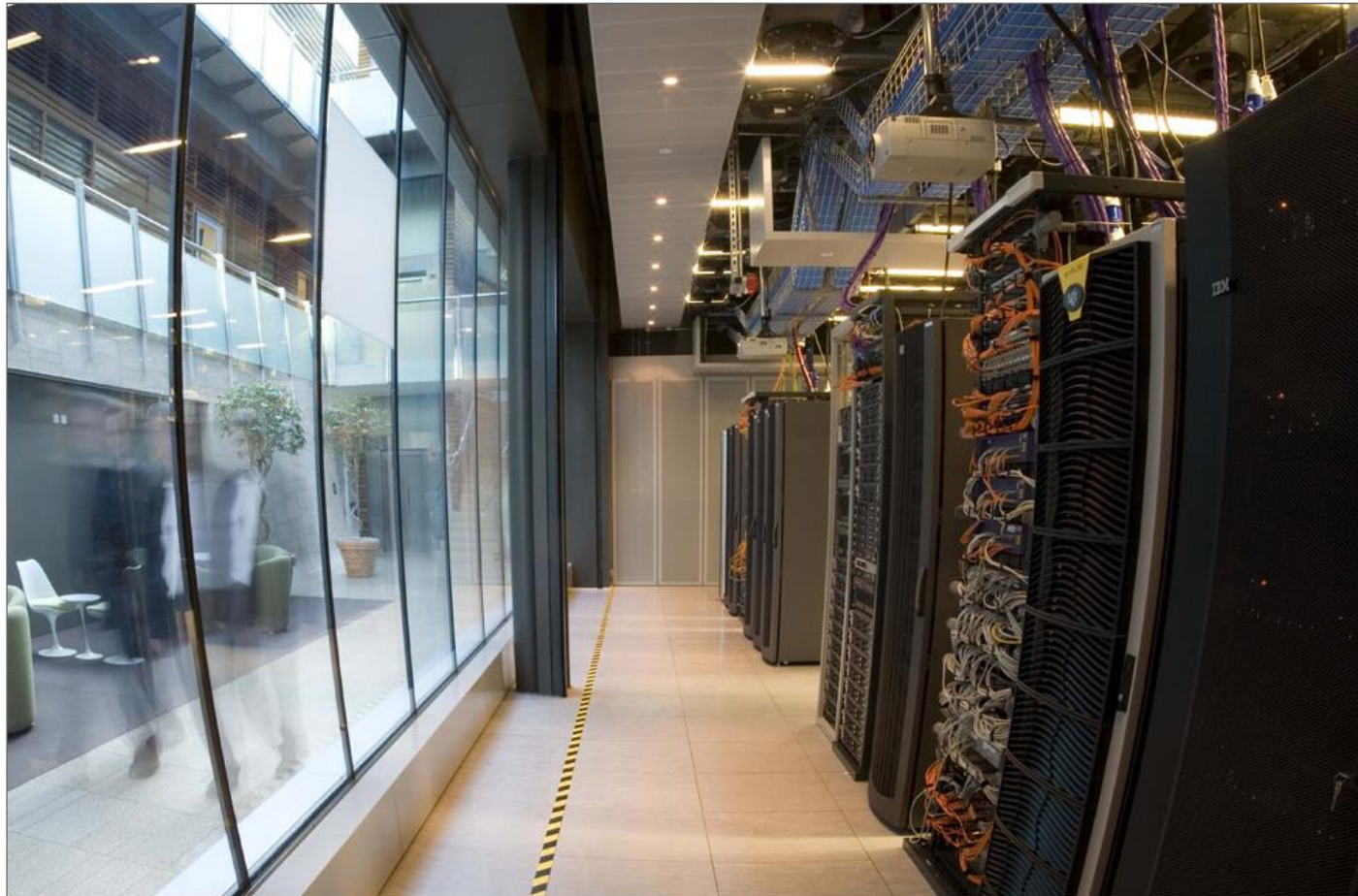
- Számítógépes modellezése
- Tudományos előrejelzése (jóslása)

new buildings. These would be home to some of the most important genetic discoveries of the 20th and 21st Centuries. A second building development to extend the Campus' facilities was opened in 2005, creating a state-of-the-art new home for staff amenities and a data centre to house the growing data storage needs of the Sanger Institute and EBI.

Throughout the
facilities are

Gallery

Click an an in



The Data Centre. The Institute's computer systems can store up to 4 petabytes of data [Credit: Genome Research Limited]

Image 3 of 31 [< Prev](#) [Next >](#)

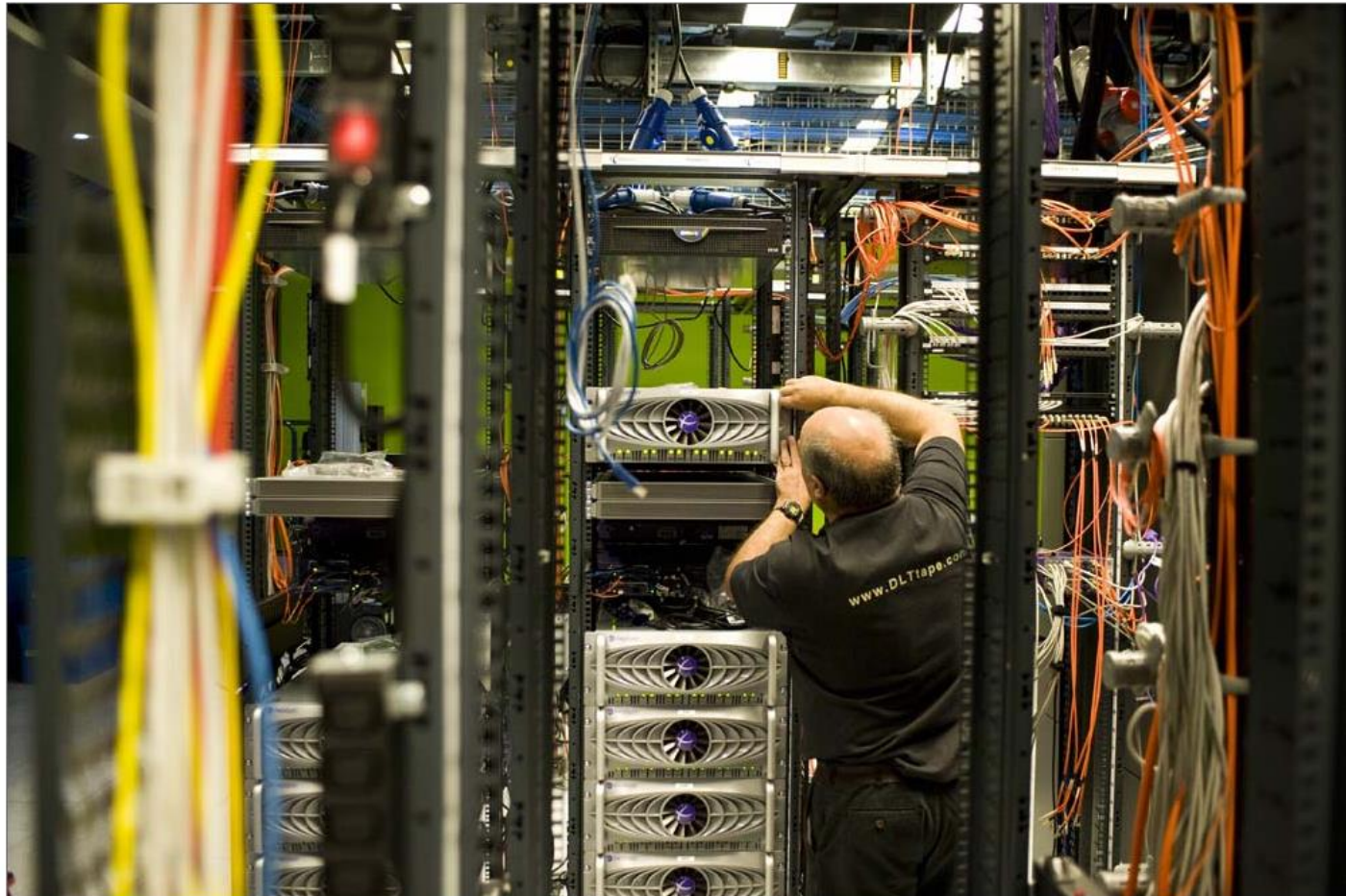
[close](#) or Esc Key

new buildings. These would be home to some of the most important genetic discoveries of the 20th and 21st Centuries. A second building development to extend the Campus' facilities was opened in 2005, creating a state-of-the-art new home for staff amenities and a data centre to house the growing data storage needs of the Sanger Institute and ERI.

Throughout the
facilities are

Gallery

Click an an in



The Centre has a total of 4 petabytes of data storage capacity [Credit: Genome Research Limited]

Image 5 of 31 < Prev Next >

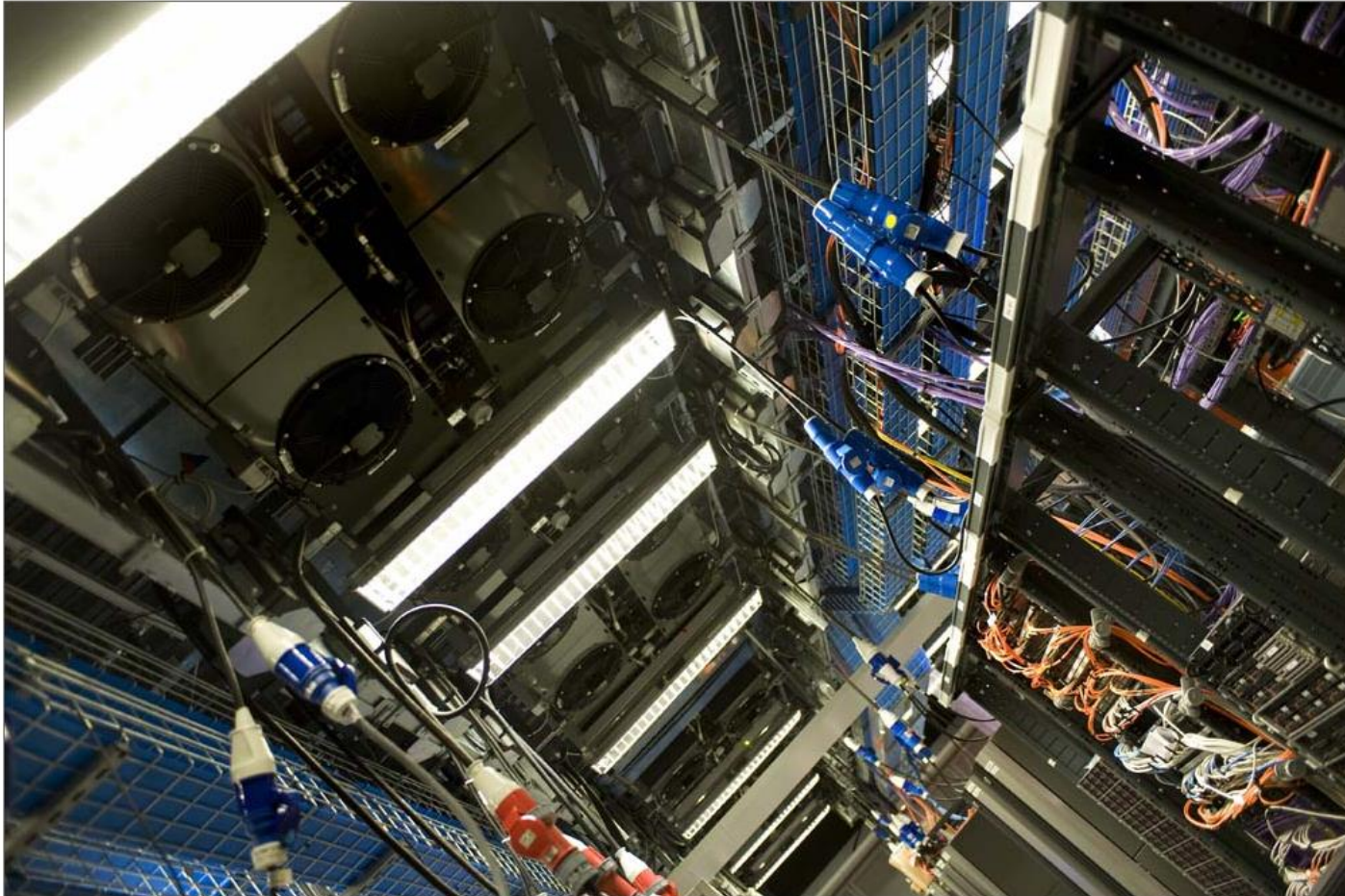
[close](#) or Esc Key

new buildings. These would be home to some of the most important genetic discoveries of the 20th and 21st Centuries. A second building development to extend the Campus' facilities was opened in 2005, creating a state-of-the-art new home for staff amenities and a data centre to house the growing data storage needs of the Sanger Institute and ERI.

Throughout the
facilities are

Gallery

Click an image



Air cooling system in the Data Centre [Credit: Genome Research Limited]

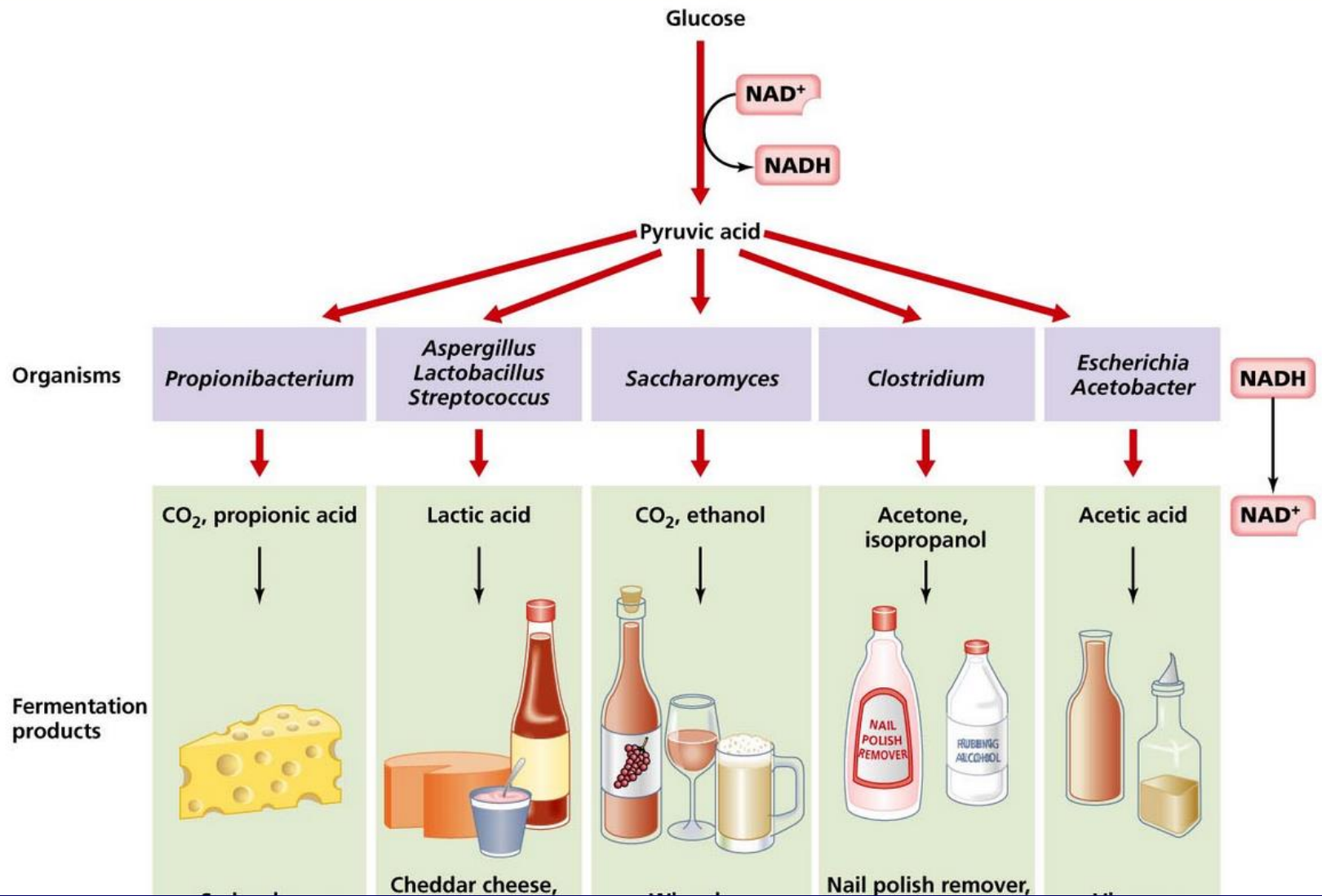
Image 6 of 31 [Prev](#) [Next](#)

[close](#) or Esc Key

Paulien Hogeweg véleménye:

It seemed to us that one of the defining properties of life was information processing in its various forms, e.g., information accumulation during evolution, information transmission from DNA to intra- and intercellular processes, and the interpretation of such information at multiple levels. At a minimum, we felt that that information processing could serve as a useful metaphor for understanding living systems. We therefore thought that in addition to biophysics and biochemistry, it was useful to distinguish bioinformatics as a research field (or what we termed a “work concept”).

Propelled by the exponential increase of sequence data, the term bioinformatics became mainstream in the late 1980s, coming to mean the development and use of computational methods for data management and data analysis of sequence data, protein structure determination, homology-based function prediction, and phylogeny. But the rich insights obtained from the massive sequencing projects, and the related bioinformatic analysis to unravel function and evolution, is not really the “roots of bioinformatics”, but rather the “trunk of bioinformatics”, and not the subject of this article.



PEARSON

LabBench Activity

[Back to LabBench Main](#)



LabBench Contents

Animal Behavior

[Introduction](#)

[Key Concepts](#)

[Concept 1: Observing Behaviors: Fruit Fly Mating Behavior](#)

[Concept 2: Observing Behaviors: Pillbug Behavior](#)

[Concept 3: Scientific Sketching](#)

[Concept 4: Response to the Environment](#)

[Analysis of Results](#)

[Design of the Experiment](#)

[The Controls](#)

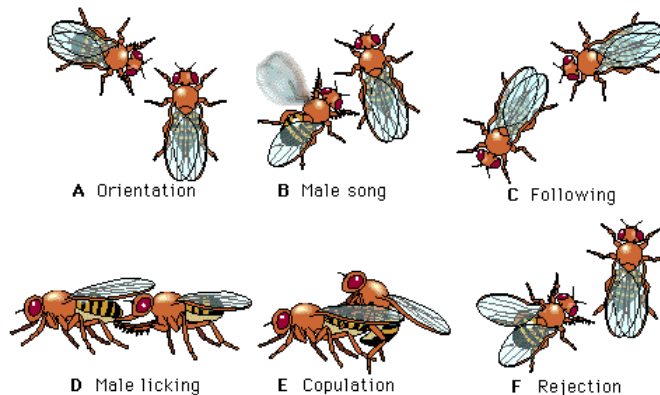
[Sample Size, Results, and Replication](#)

[Other Considerations](#)

[Lab Quiz](#)

Observing Behaviors: Fruit Fly Mating Behavior

There are hundreds of species of fruit flies — so how do members of the same species find each other and signal willingness to mate? Each species has evolved a complex series of behaviors that appear to be genetically programmed. In this activity, you observe the mating dance of *Drosophila melanogaster* and then design your own experiment to investigate some aspect of mating behavior.



Above are sketches of some behaviors you may see. Print the sketches, take them to the laboratory with you, and refer to them as you observe your own flies. After you place your flies in a common vial it will take a few minutes for them to acclimate to their new environment, so be patient and quiet as you wait for the dance to begin. Record your observations.

In the next section, you will observe the behavior of another organism, the pillbug.

[Previous Concept](#)

[Next Concept](#)



Copyright © Pearson Education, Inc. or its affiliates. All Rights Reserved.

[About Pearson Tech Support Product Information](#)

[Privacy Policy Terms of Use Rights and Permissions](#)





Mióta „igazi” tudomány a ...

- Fizika: Newton óta „komoly” tudomány, Einstein jelentősen átformálta, de nem törölte el az alapokat (XVII század)
- Kémia: nehéz egy névhez vagy eseményhez kötni – („flogiszton”, „életerő”, vegyérték, periódusos rendszer) (XIX század vége)
- Biológia: még nehezebb ügy – (Darwin, Mendel, Watson – Crick) (1953)

A matematika és a biológia

- 1978 – „érdekel a természettudomány, de a matekot inkább elkerülöm – biológus leszek”

Heavy use of equations impedes communication among biologists

Tim W. Fawcett¹ and Andrew D. Higginson

School of Biological Sciences, University of Bristol, Bristol BS8 1UG, United Kingdom

Edited[†] by Robert M. May, University of Oxford, Oxford, United Kingdom, and approved June 6, 2012 (received for review April 4, 2012)



A matematika szerepe a tudományokban

Mennyire használja

- Fizikus: „gyenge matematikus”
- Vegyész: „képzetlen fizikus”
- Biológus: „tanulatlan vegyész”

Mire jut vele

- A hidrogén atom teljesen leírható – minden másra közelítő megoldás van:
 - Homogén
 - Kevés komponensű
 - Egyensúlyi
- Az élő rendszer:
 - Inhomogén
 - Sok komponensű
 - Nem egyensúlyi

Miért van bioinformatika?

Az élő rendszer lényegéből fakadóan komplex

- A matematika hagyományos formája nem alkalmas a biológiai rendszerekre
- Helyette van az informatika (alkalmazott matematika, numerikus módszerek, káosz elmélet)
- A matematika beépülése a tudományba természetes folyamat, ez teszi hatékonnyá a tudományt

Recent advances in DNA microarray hybridization technology make it possible to record the molecular biological signals, e.g., mRNA expression levels and proteins' DNA-binding occupancy levels, that guide the progression of cellular processes on genomic scales (1, 2). Biology and medicine today may be at a point similar to where physics was after the advent of the telescope (3). The rapidly growing number of DNA microarray data sets holds the key to the discovery of previously unknown molecular biological principles, just as the astronomical tables compiled by Galileo and Brahe (Fig. 1A) enabled accurate predictions of planetary motions and, later, the discovery of universal gravitation. Just as Kepler and Newton made these predictions and discoveries by using mathematical frameworks to describe trends in astronomical data (Fig. 1B), so future predictive power, discovery, and control in biology and medicine will come from the mathematical modeling of DNA microarray data, where the mathematical variables and operations represent biological reality: The variables, patterns uncov-

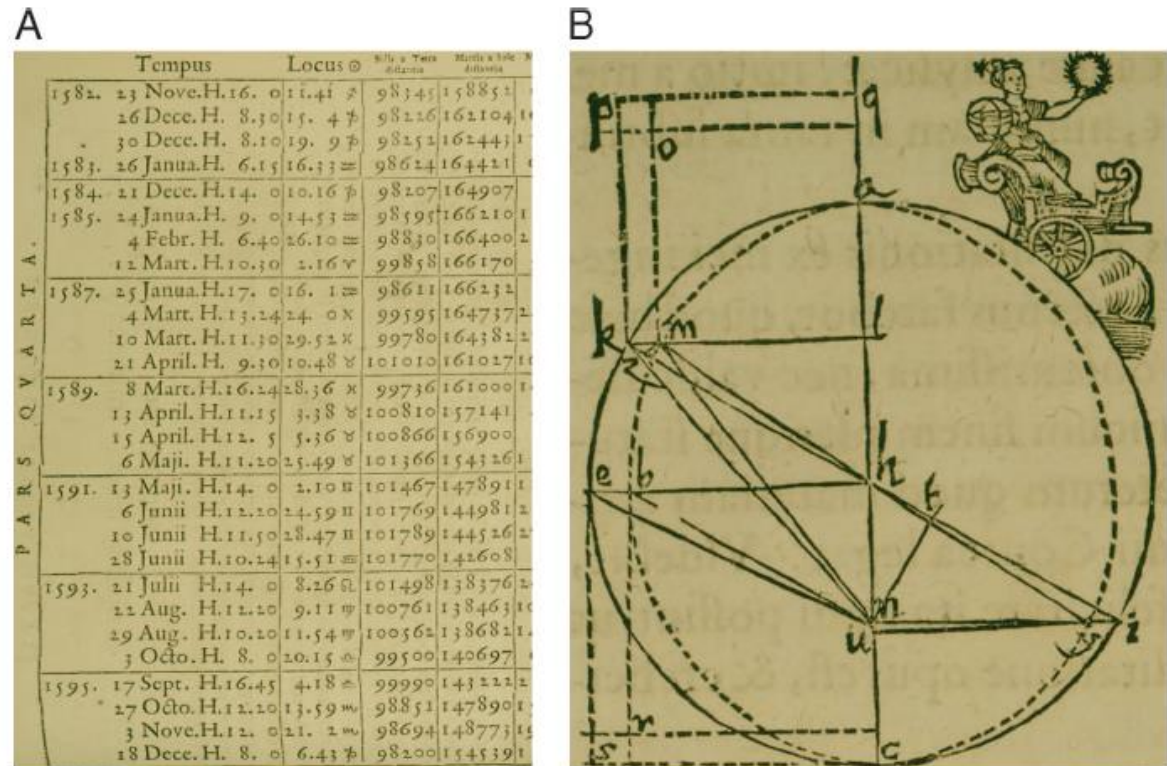


Fig. 1. Kepler's discovery of his first law of planetary motion from mathematical modeling of Brahe's astronomical data. (A) Astronomical table of positions of the sun, Earth, and Mars at different times. (B) Geometrical reconstruction of the orbit of Mars from these data reveals an ellipse with the sun located at one focus. [Images from ref. 3 (reproduced by permission of the Harry Ransom Humanities Research Center of the University of Texas, Austin, TX).]

Élő rendszerek komplexitása

- Hány 6 Kb hosszú DNS fragmens lehetséges?

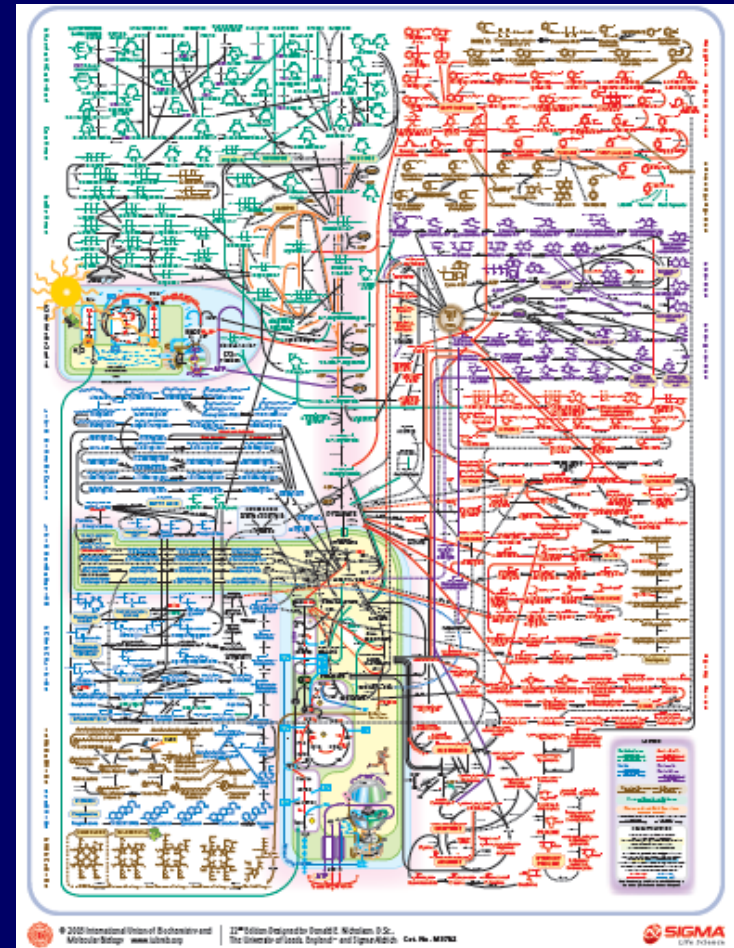
Legyen mind a 4 bázis egyenlő mértékben:

$$6000! / (1500!)^4 =$$

$$5.0 \times 10^{3606}$$

- Metabolikus hálózat

Minden részlete fontos,
de az egész együtt is.



Komplex rendszerek vizsgálata

- A részleteken keresztül értjük meg
- A részek csak együtt adják ki a teljes képet
- Pl.: a geológia használja a mikroszkópot és a teleszkópot is
- Az optikai elv azonos, a vizsgálható tárgy mérete eltér
- A bioinformatika a biológia teleszkópja

Computer – „adattfeldolgozó-gép”

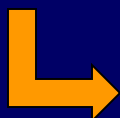
„adattfeldolgozó-

nyam-
nyam

plotty

gép”

ACGTAGCTG
GATCAGTCA
TGCAGTCT
GATCAGTCA
TGCAGTCT
GATCAGTCA
TGCAGTCT
GATCAGTCA
TGCAGTCT
GATCAGTCA



ACGTTAGACT
GGTATCTGTA



Miben jó

Az ember

- Vizuális típus, képekben gondolkodik

Könnyű feladat – egy ember:

- Eltérő arckifejezés
- Más szög
- Más életkor

A gép

- Minden információt számokká alakít

Könnyű feladat

- Jegyezzünk meg 20 millió telefonszámot
- Küldjünk szét 50 millió e-mail-t
- Számoljuk ki a π értékét 5000 milliárd számjegyre

Számítógépet akkor használunk, ha..

- Sok adatunk van
- Ismerjük a feldolgozás eszközeit
- Tudunk ilyet írni
- Értelmezni tudjuk az eredményt
- Figyelembe vesszük a lehetőségeket
 - Perc/másodperc – remek
 - Órák – jó
 - Napok – nem probléma
 - Hetek/hónapok – megoldható
 - Több kéne??? – akkor bajban vagy

A problémák nehézsége – “Skálázás”

Van egy:

- Adott problémám
- Egy lehetséges módszer a megoldásra
- Számítógépem hozzá

Megkapom a megoldást t idő alatt.

- Kétszer bonyolultabb problémához mennyi idő kell?

A bioinformatikus feladata

- A biológia problémáit újrafogalmazza a gép számára kezelhető formában
- Az eredményeket értelmezni tudja a biológiai összefüggéseikben
- Meg tudja ítélni a problémák megoldhatóságát
- Szót tud érteni fizikusokkal, matematikusokkal és informatikusokkal



A bioinformatika jövője

Lincoln D Stein (2008):

Ontario Institute for Cancer Research, Toronto, ON, M5G 0A3, Canada.
Cold Spring Harbor Laboratory, NY 11724, USA.

Abstract

Bioinformatics has become too central to biology to be left to specialist bioinformaticians. Biologists are all bioinformaticians now.

So here is my revised prognosis for the next five years:

Bioinformaticians: gone by 2012. Bioinformatics: stronger than ever.

A kurzus célja a kísérletes biológusok felkészítése erre a szerepre, nem célja specialisták képzése.



Szünet



A bioinformatika tipikus alkalmazásai

- Szekvencia analízis
- Genomi annotáció
- Evolúció-biológia
- Adatbányászat
- Génreguláció analízise
- SNP analízis
- Összehasonlító genomika
- Rendszerbiológia
- Biológiai képfeldolgozás
- Térszerkezeti bioinformatika

Szekvencia analízis

- A leggyakrabban használt bioinformatikai módszer

Alkalmazzák:

- DNS-re, RNS-re és fehérjére
- Egy genomon belül
- Több genom összehasonlítására
- Adatbázis keresések motorja
- Kísérleti eredmények kiértékelésénél

A szekvencia elemzés célja

- A DNS tartalmazza a sejt működéséhez szükséges információ döntő többségét
- Sok genomot ismerünk – mégsem értjük teljesen az élő rendszereket
- A genomi információ több rétegű
- Az eltérő információs szintek kapcsolatban állnak egymással
- A szekvencia elemzés kódfejtés

Alkalmazott eljárások

- Génpredikció
 - Fehérje kódoló
 - RNS-gén
- Szabályozó régiók
 - CpG szigetek
 - TF-kötőhelyek
- Szerkezeti motívumok
- Repetitív elemek keresése
- Szekvenciaillesztés
 - Egy fajon belül
 - Fajok közt
- Keresés adatbázisban
 - Lokális hasonlóság
 - Globális hasonlóság
- Genom rekonstrukció (NGS kísérletek)

Genomi annotáció

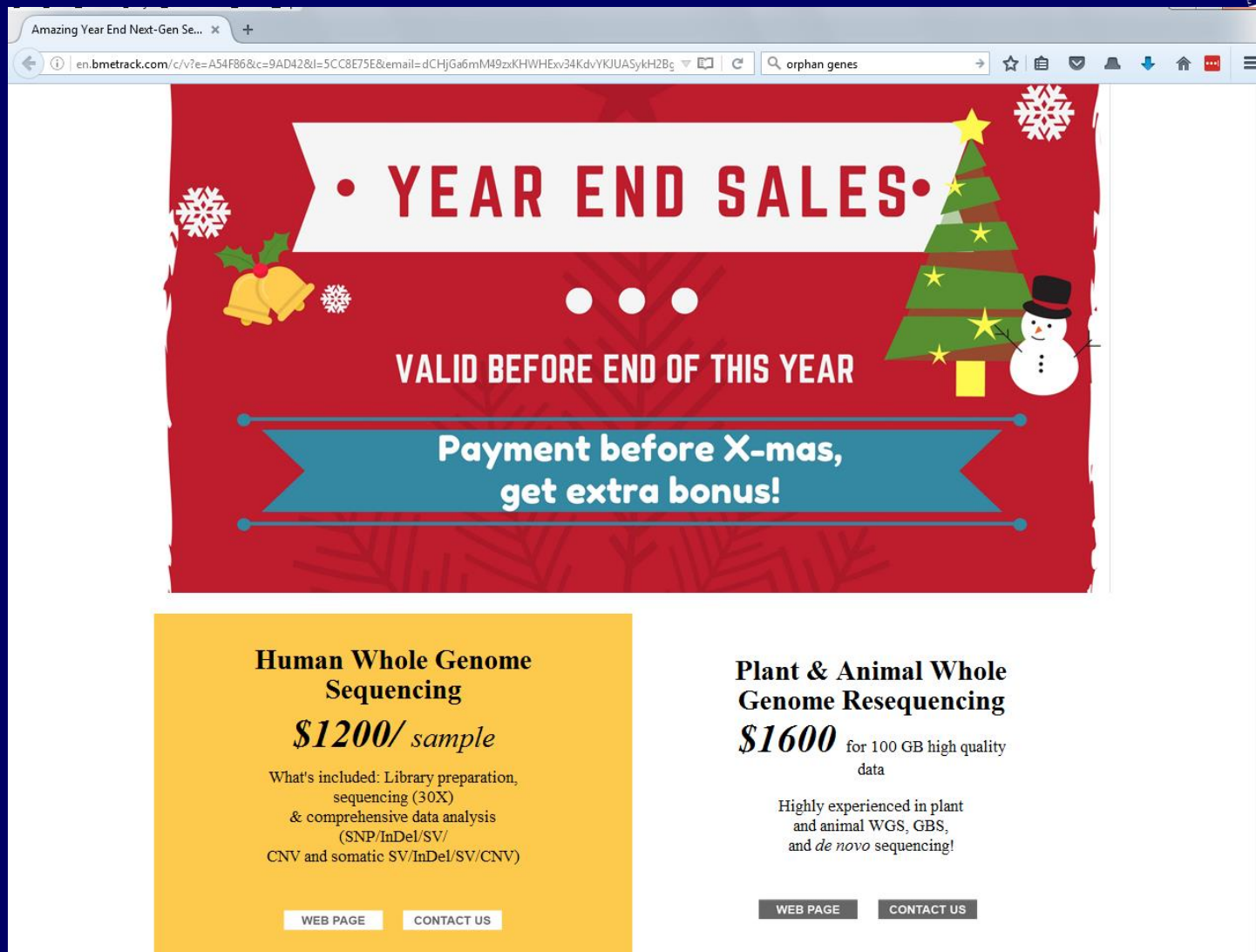
A genom szekvenálás

- Gyors
- Olcsó (cél: „1000 \$-os” genom)

Az igazi kihívás az adatok értelmezése:

- Ismert információ hozzáadása az új genomhoz

```
TCCCAGGCAACTTCCTGGGAACTCCCCAGGCGGAGGACTCCGAGACCTCAGGCCTTCCT
GTCTCCCCCTCCCCCTCCTCTCAAACCCCTCCCCCTCCCTCCACCTCTTCAGTTTGCTCTT
CAAACCTTGCTGGACGCCATTCTATGCTGGGGCCAAGAACACAAAGAGCGGAGGAAGGGAAC
AGGTTAAAGAAAACAAGAAACACAATCAGACCACAGAAAAGCCAGGCAGAAAAGGGTTTCG
ACGGGCAAAAAGAATGTGGCTGTCCAGATAAAGAATGTCTGTCCCGGCCCGGCCTGTGC
TGCAAGTGCGCAACTCACCTAGCCGCTGCCACCCAGGCTCCCGCCACCGCGCAGCCCCG
CCAGCGGCTTCTCGCCTCCCCCTCGCCTCGGATAGGGTTAGGGCCTGAGGTAATAAATG
CAAGGCCTTCAATTCTCCAAGCAGTGCGCAGTGCAATTTTCTTTATTTCTGGGAACTTGC
GCCCAGGTCTCTGTGAGGCCTGCTGTGAGGGATTCTACGCGGGGAGAAGGTGGAGGCTGC
GCAGGTGGAGAAAGGGGCCCCAGAAAGGGGGGCTAGAAAGTGGAGGGCAACGTGGGGGCGGG
GCGGGTATCCCAGAGGGTGCCCTGGAGGGTCTGTAGTTGATGTCTTAAACATGCAGGT
CACTTGTTTCAGAGAACTTTATTTGCTTCTTAGGCCTCGCTAGGAGCATCGGCTGTTTC
AGGACCTGGAGAAAGGCCCCAGCTCTACCCTGAGAGGACGTGTCTCTCCACGCTCTCTCC
GCAAATGCTGTCCCTCTTCCCCAGCCAGGGCCCGGCTCTTCGGTGTGTCTGGGCCATTG
CAACCCCGTCTCCCCACCTCTCCGCATGGCCCTCGCGCCTTGAGACTGGGCAGGGCAGG
CTGATGGAGGGGCGGGAGGGGTGGCGGTTGCCAGGCTAACGTGTCCGTGCGTGGGGGT
CCCCTTGCTCTCGCAGAGGTGTAAGGACAAGTTGAACGCTTTGGCCATCTCGGTGATGAA
CCAGTGGCCAGGAGTGAACTGCGGGTGACCGAGGGCTGGGACGAAGATGGCCACCACTC
AGAGGAGTCTCTGCACTACGAGGGCCGCGCAGTGGACATCACCACGTCTGACCGCGACCG
CAGCAAGTACGGCATGTGGCCCGCTGGCGGTGGAGGCCGGCTTCGACTGGGTGTACTA
CGAGTCCAAGGCACATATCCACTGCTCGGTGAAAGCAGGTAAGCTGGGCCCTGGCCCCCG
GATCCGACCCAAGGAAGGCCATTGGCGCACCTCGGCTTGATTCAAGAGAAAAAGAACTT
GGGGGGAGGCTGAGGGCCAGGAGCAGGGTCGCTGGGCGATGACTGCGTTTCCGCGGTGGA
ACCTGCCCTGTGAGGTGCCGGCCCTCGAAATCACCCCTACCTTTGAGGCCACAGAGCCC
```



Amazing Year End Next-Gen Se... x

en.bmetrack.com/c/v?e=A54F86&c=9AD42&l=5CC8E75E&email=dCHjGa6mM49zKHWHEv34KdvYKJUA5ykH2Bg orphan genes

• YEAR END SALES •

VALID BEFORE END OF THIS YEAR

Payment before X-mas,
get extra bonus!

Human Whole Genome Sequencing

\$1200/ sample

What's included: Library preparation, sequencing (30X) & comprehensive data analysis (SNP/InDel/SV/CNV and somatic SV/InDel/SV/CNV)

WEB PAGE CONTACT US

Plant & Animal Whole Genome Resequencing

\$1600 for 100 GB high quality data

Highly experienced in plant and animal WGS, GBS, and *de novo* sequencing!

WEB PAGE CONTACT US

Az annotált genom

Integrált adatbázis, minden ismert információt tartalmaz

- Kísérletes adatokat rokon fajokból
 - Általános érvényű szabályokból eredő következtetéseket
 - Predikciókat elfogadott modellek alapján
- Figyelem! Eukariótákra és prokariótákra más szabályok vonatkoznak.

Evolúció-biológia

- Az evolúció genetikai nyomait keresi
 - Egyes gének szintjén
 - Teljes genomok összehasonlításával
- Az evolúció folyamatának modellezése a populáció szintjén

Orvosbiológiai vonatkozás:

- Megfelelő biológiai modell kiválasztása pl. kutya, gomba

File Edit View History Bookmarks Tools Help

Exome sequencing of lymph...

genome.cshlp.org/content/early/2015/09/16/gr.194449.115.abstract?paperoc

HOME ABOUT ARCHIVE SUBMIT SUBSCRIBE ADVERTISE AUTHOR INFO CONTACT HELP

Institution: SEMMELWEIS UNIV OF MEDICINE Sign In via User Name/Password

Search for Keyword: Go

Advanced Search

Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background

Ingegerd Elvers¹, Jason Turner-Maier², Ross Swofford², Michele Koltookian², Jeremy Johnson², Chip Stewart², Cheng-Zhong Zhang³, Steven E Schumacher³, Rameen Beroukhim³, Mara Rosenberg², Rachael Thomas⁴, Evan Mauceli⁵, Gad Getz⁶, Federica di Palma⁷, Jaime F Modiano⁸, Matthew Breen⁹, Kerstin Lindblad-Toh^{1,10} and Jessica Alfoldi²

Author Affiliations

* Corresponding author; email: kersli@broadinstitute.org

Abstract

Lymphoma is the most common hematological malignancy in developed countries. Outcome is strongly determined by molecular subtype, reflecting a need for new and improved treatment options. Dogs spontaneously develop lymphoma, and the predisposition of certain breeds indicates genetic risk factors. Using the dog breed structure, we selected three lymphoma predisposed breeds developing primarily T-cell (boxer), primarily B-cell (cocker spaniel), and with equal distribution of B and T-cell lymphoma (golden retriever), respectively. We investigated the somatic mutations in B- and T-cell lymphomas from these breeds by exome sequencing of tumor and normal pairs. Strong similarities were evident between B-cell lymphomas from golden retrievers and cocker spaniels, with recurrent mutations in TRAF3-MAP3K14 (28% of all cases), FBXW7 (25%), and POT1 (17%). The FBXW7 mutations recurrently occur in a specific codon; the corresponding codon is recurrently mutated in human cancer. In contrast, T-cell lymphomas from the predisposed breeds, boxers and golden retrievers, show little overlap in their mutation pattern, sharing only one of their 15 most recurrently mutated genes. Boxers, which develop aggressive T-cell lymphomas, are typically mutated in the PTEN-mTOR pathway. T-cell lymphomas in golden retrievers are often less aggressive and their tumors typically showed mutations in genes involved in cellular metabolism. We identify genes with known involvement in human lymphoma and leukemia, genes implicated in other human cancers, as well as novel genes that could allow new therapeutic options.

Received May 13, 2015.

OPEN ACCESS ARTICLE

ACCEPTED MANUSCRIPT

This Article

Published in Advance
September 16, 2015, doi:
10.1101/gr.194449.115

Genome Res. 2015.
Published by Cold Spring
Harbor Laboratory Press

- Abstract **Free**
- Full Text (PDF) **Free**
- Supplemental Material

Article Category

Research

Services

- Google Scholar
- PubMed/NCBI
- Share
- Metrics

1

See more details
Tweeted by 1

Recent Updates

Follow us on twitter

Genome Research
@genomeresearch 12h
MT@DailyScan:
@genomeresearch:regulation of alt
splicing, host genetic-environ
interactions' effect on gut microbes...
http://t.co/CNqUqIB7R

Genome Research
@genomeresearch 12h
RT @thermophila: DNA-guided
nucleosome positioning in eukaryotes by

Current Issue

September 2015, 25 (9)



From the Cover

Alert me to new issues of
Genome Research

- Advance Online Articles
- Submit a Manuscript
- CR in the News
- Editorial Board
- Permissions
- E-mail Alerts & RSS Feeds
- Recommend to Your Library
- Job Opportunities



swifbiosci.com

Highest

Adatbányászat

- A kísérletes adatok elsősorban közleményekben jelennek meg szöveges formában
- Túl sok az ujság, túl sok a cikk
- Lehetetlen elolvasni minden érdekeset
- Automatizálni kellene a cikkek olvasását:
 - Különleges statisztikai módszerek
 - „számítógépes nyelvészet”

Mesterséges szövegértelmezés

A bioinformatika legfiatalabb területe

- Rövidítések felismerése
- Kifejezések felismerése
- Összefüggések értelmezése
 - „A” gén elnyomja „B” gén expresszióját
 - „B” gén expressziója csökken „A” gén hatására
- Eltérő kísérleti technikák kezelése
- Konfliktusok kezelése

Génreguláció analízise

- A gének expressziója időben nem feltétlenül állandó
- Egyes gének
 - Állandóan be vannak kapcsolva („housekeeping genes”)
 - Igény szerint kapcsolnak ki és be
 - Ciklikusan kapcsolnak ki-be
 - Egyszer bekapcsolnak és úgy maradnak
 - Egyszer kikapcsolnak és úgy maradnak

A génszabályozás jelentősége

- Kóros állapotban, bizonyos szövetekben,
- olyan gének működnek, amelyeknek nem kéne
 - más gének meg nem működnek, pedig kéne

Ezért a génszabályozás megértésének óriási a jelentősége diagnosztikai és terápiás szempontból is.

A génszabályozás modellje

- Az elfogadott modell:
 - Fehérje-DNS kölcsönhatáson alapul
 - A transzkripció indításához különleges fehérjék kellenek (transzkripciós faktorok)
 - Ezek a megfelelő kötőhelyeken keresztül ismerik fel a megfelelő gént a kromoszómán
- A konkurens modell:
 - DNS-RNS-fehérje kölcsönhatáson alapul
 - A felismerést az RNS-DNS kölcsönhatás biztosítja

Alkalmazott módszerek

- mRNS szekvenálás
- EST szekvenálás
- RNA-Seq
- RT-PCR
- DNS-chip
- Fehérje-chip
- Tandem-MS
- Több módszer esetében is szükség van informatikára már az adatgyűjtés szakaszában is
- A nyers adatok zajosak. A kiértékeléshez kifinomult statisztikai módszerek kellenek.

SNP analízis

- Cél: a pontmutációk feltérképezése és elemzése
- Egy faj két egyede közti eltérések jelentős részben SNP-kre vezethetők vissza
- Gyógyászati jelentőségük:
 - Kapcsolatba hozhatók-e valamilyen betegséggel
 - Esetleg valamilyen betegséggel szembeni védettséggel

Az SNP-k típusai

- A legtöbb SNP-nek két allélje van
- Esetnek kódoló szakaszra vagy azon kívülre
- A kódoló szakaszra eső polimorfizmus nem feltétlen eredményez eltérő fehérjét (silent mutation)
- SNP okozhatja a fehérje lerövidülését (korai stop-kódon
- A nem-kódoló szakaszokra eső SNP-k szintén fontosak

Összehasonlító genomika

- Teljes genomok összehasonlításával foglalkozik
- Nagy léptékű változások elemzésével állít fel evolúciós modelleket
- Inkább az alap kutatás szempontjából fontos
- Igen változatos és kifinomult statisztikai módszereket alkalmaz



Rendszerbiológia

- Viszonylag új területe a bioinformatikának
- Hogeweg eredeti elképzeléséhez áll közel
- Távlati cél: egy sejt minden működését modellezni
- Végző cél: egy többsejtes élőlény teljes modellezése

Alkalmazott megközelítés

- Az élőlény működését hálózatok hierarchiája alapján írja le
- Metabolikus hálózat
 - Viszonylag ismert
- Szignalizációs hálózat
 - Kevésbé ismert, de intenzíven kutatott
- Génszabályozási hálózat
 - Még kevésbé ismert, de még intenzívebben kutatott

Biológiai képfeldolgozás

- A vizuális információ kitüntetetten fontos a biológiában
- A képi információk feldolgozása
 - Időigényes
 - Nagy gyakorlatot igényel
 - Nehéz számszerűsíteni az eredményt
 - Folyamatok követése igen körülményes
- A képfeldolgozás automatizálása ezért fontos

Alkalmazási területek

- Nagy sebességű, nagy pontosságú sub-celluláris lokalizáció analízise
- Morfometria
- Klinikai képfeldolgozás és vizualizálás
- Kísérleti állatok viselkedésének követése automatikusan kiértékelt videó-felvételek segítségével
- Mérések kiértékelése – gén-chip technológia

Térszerkezeti bioinformatika

- Megfigyelés: a térszerkezet és funkció szorosan összefügg
- Ez áll fehérjékre és RNS-ekre is
- A molekulák térszerkezete
 - Mérésekkel meghatározható
 - Modelllezhető
- Vannak térszerkezettel nem rendelkező fehérjék is

Fehérjék térszerkezete

- Mérési módszerek:
 - Röntgen-kristallográfia
 - Nagy felbontású NMR
- Az adatok feldolgozása számítógépet igényel
- Korlátai:
 - A fehérjét kristályosítani kell – „fekete mágia”
 - A kristályos forma biológiai relevanciája kérdéses
 - Transzmembrán fehérjéket egyik módszer sem kezel

Szerkezet becslése

- Szerkezeti hierarchia:
 - Elsődleges – a szekvencia maga
 - Másodlagos – térszerkezeti elemek
 - Harmadlagos – a másodlagos elemek egymáshoz viszonyított helyzete
 - Negyedleges – fehérje komplexek szerkezete

A hierarchia szintjei egyre több információt tartalmaznak és egyre nehezebb feladat megbízható becslést adni rájuk

Alkalmazott módszerek

- Motívumok, modulok azonosítása
 - Másodlagos szerkezet szintjén működik
- Homológia modellezés
 - Harmadlagos szerkezet szintjén
 - Nagyfokú homológiát igényel
- Ab initio módszerek
 - Molekula mechanika
 - Molekula dinamika

További módszerek

- Dokkolás, drog tervezés
 - Negyedleges szerkezet szintjén
- „threading” technika
 - Harmadlagos szerkezet szintjén
- RNS térszerkezet predikció
 - Összefügg a stabilitással és funkcióval
 - Egyszerűnek látszik, de nagyon nehéz
 - A lehetőségek száma rendkívül nagy (kombinatorikai robbanás)



Köszönet nyilvánítás:

- Dr Miklós István
- Prof. Hunyady László, Prof. Falus András
- Prof. Patthy László, Prof. Simon István
- Dr Szalai Csaba
- Wikipedia, Google

Ajánlott irodalom

- Hogeweg, P. "The roots of bioinformatics in theoretical biology." PLoS Comput Biol 7(3): e1002021.
- Stein, L. D. (2008). "Bioinformatics: alive and kicking." Genome Biol 9(12): 114.
- Alter, O. "Discovery of principles of nature from mathematical modeling of DNA microarray data" **PNAS 103 (44): 16063–16064**
- Fawcett, T.W. and Higginson, A.D. "Heavy use of equations impedes communication among biologists" **PNAS 109(29): 11735–11739**



Feladat 1.

Mit vársz a kurzustól, mi érdekel a leginkább, mi legfontosabb számodra a bioinformatikában, mit tudtál eddig a bioinformatikáról?

A megoldást e-mail-ben kérem a következő előadás előtt.

Tárgymező: 'Kurzus – Feladat 1.'