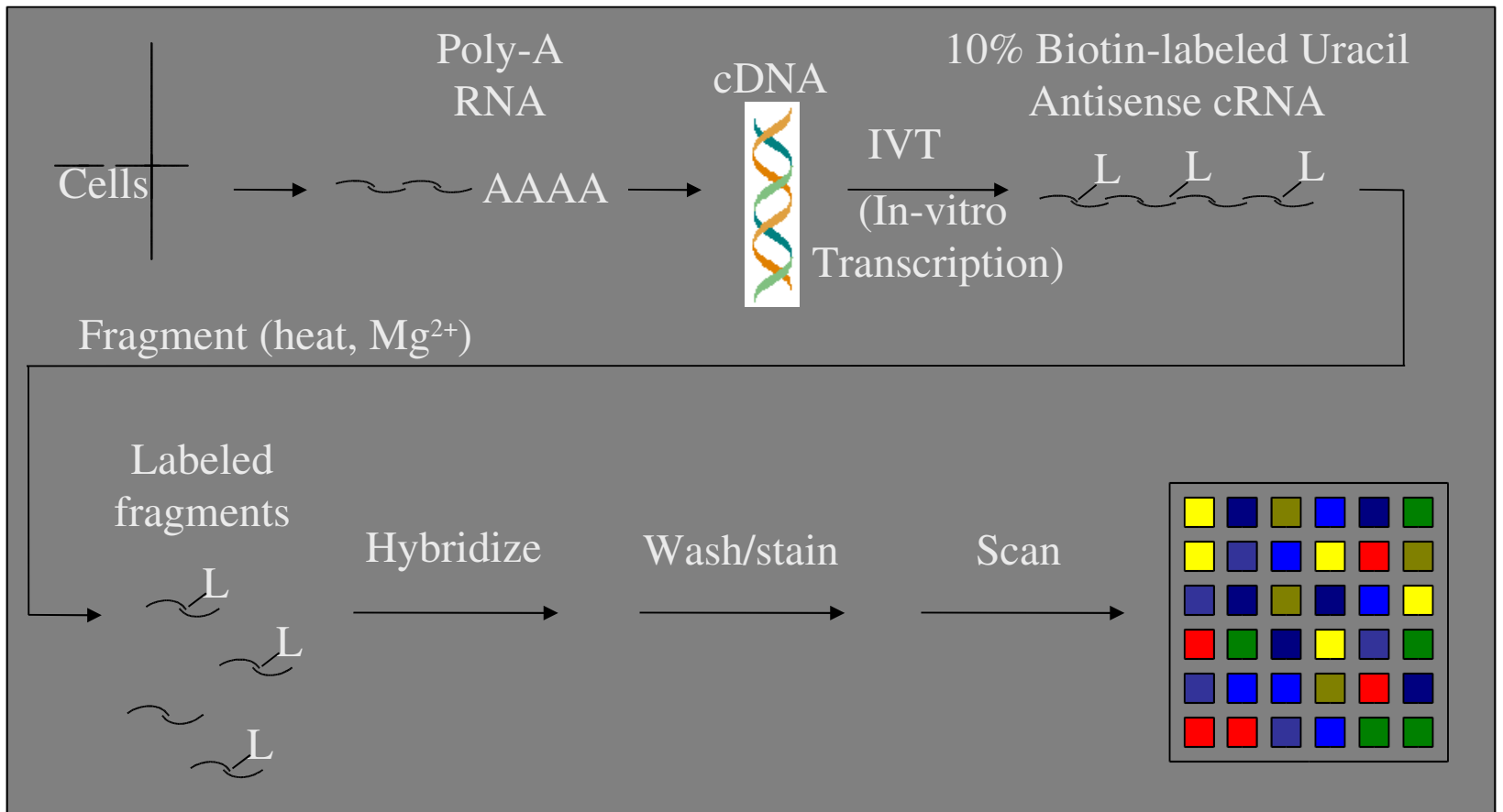


Osnova

- Analýza expresních dat
- Analýza dat v literatuře
- Ariadne Genomics



Analýza expresních dat

- nejčastěji slouží k vytvoření obrazu o expresi genů
- možno je považovat za jakýsi molekulární fenotyp
- využití tam, kde fenotyp neposkytuje dostatek informací
(výzkum funkce genů, medicínská diagnostika)
- vyžadují použití méně běžných metod statistiky ($p \gg n$)
- redukce rozměru daného problému, shlukování

Analýza expresních dat

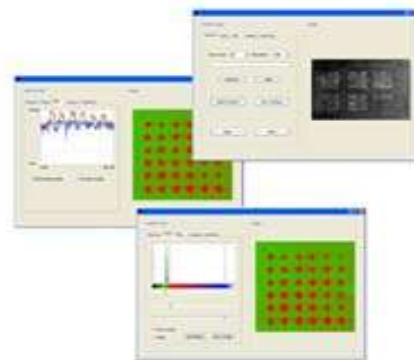


Analýza expresních dat



- izolace mRNA,
příprava cDNA,
značení

- příprava sklíček se sondami,
hybridizace, omývání



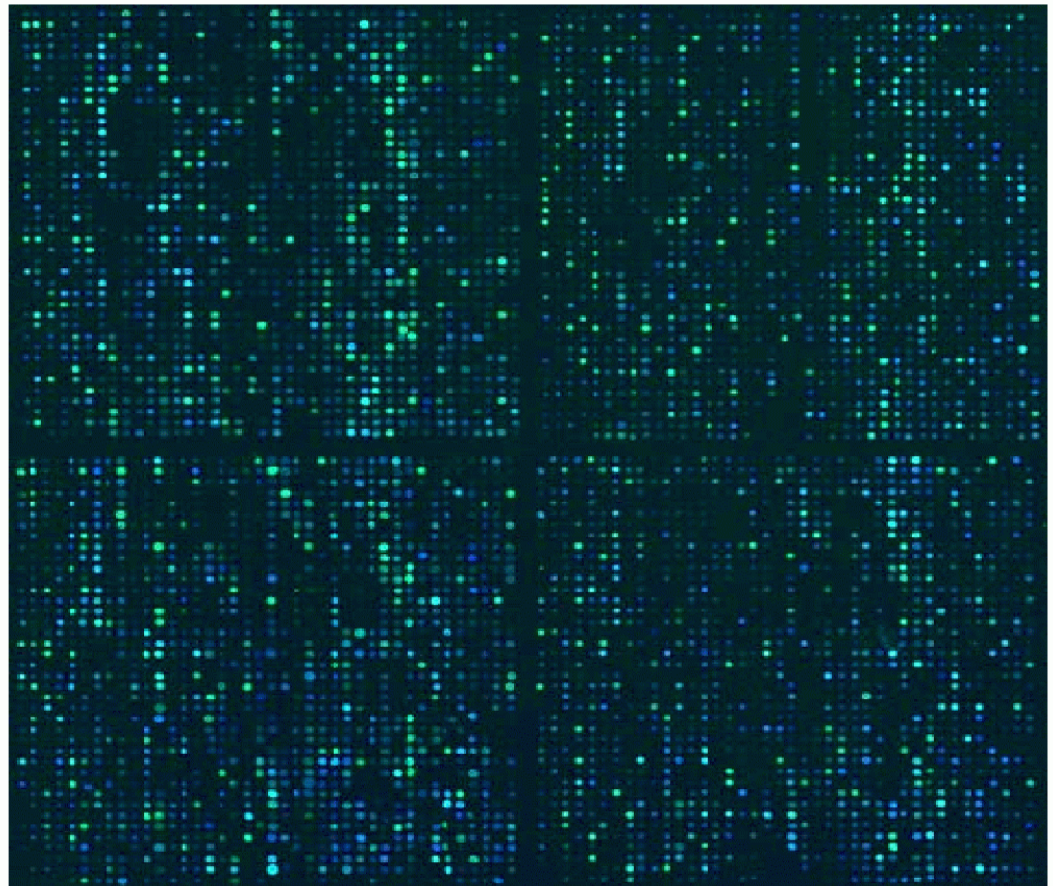
- čtení signálu na jednotlivých
pozicích po hybridizaci,
primární data

- analýza dat

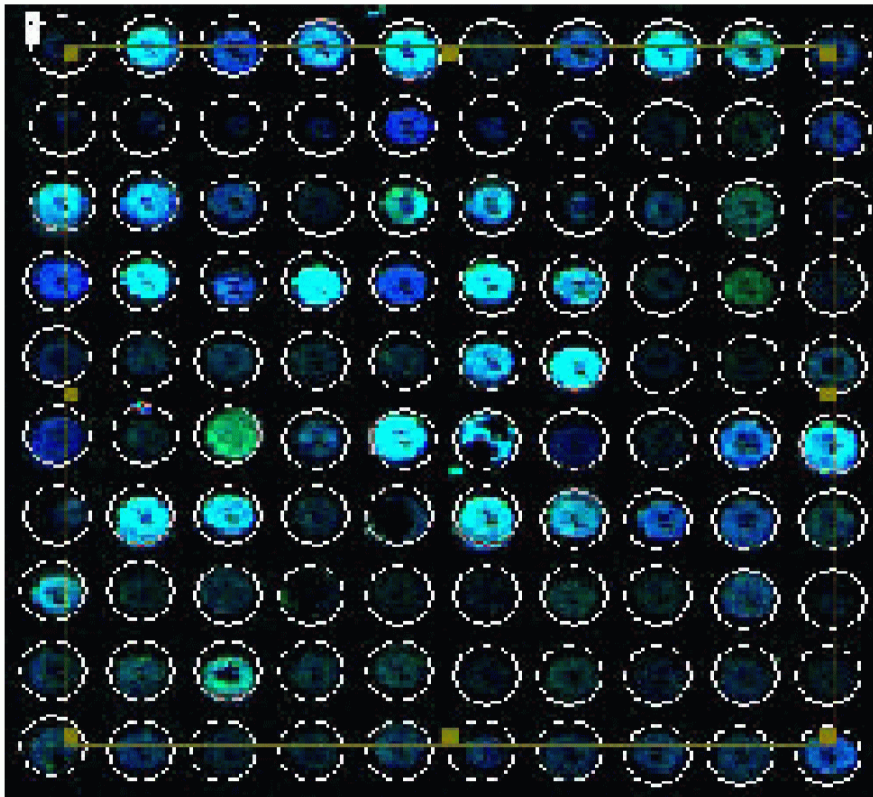
Analýza expresních dat

PRIMÁRNÍ DATA

Jednotlivé pozice odpovídají různým sekvencím DNA, intenzita signálu stupňu jejich přítomnosti ve vzorku.

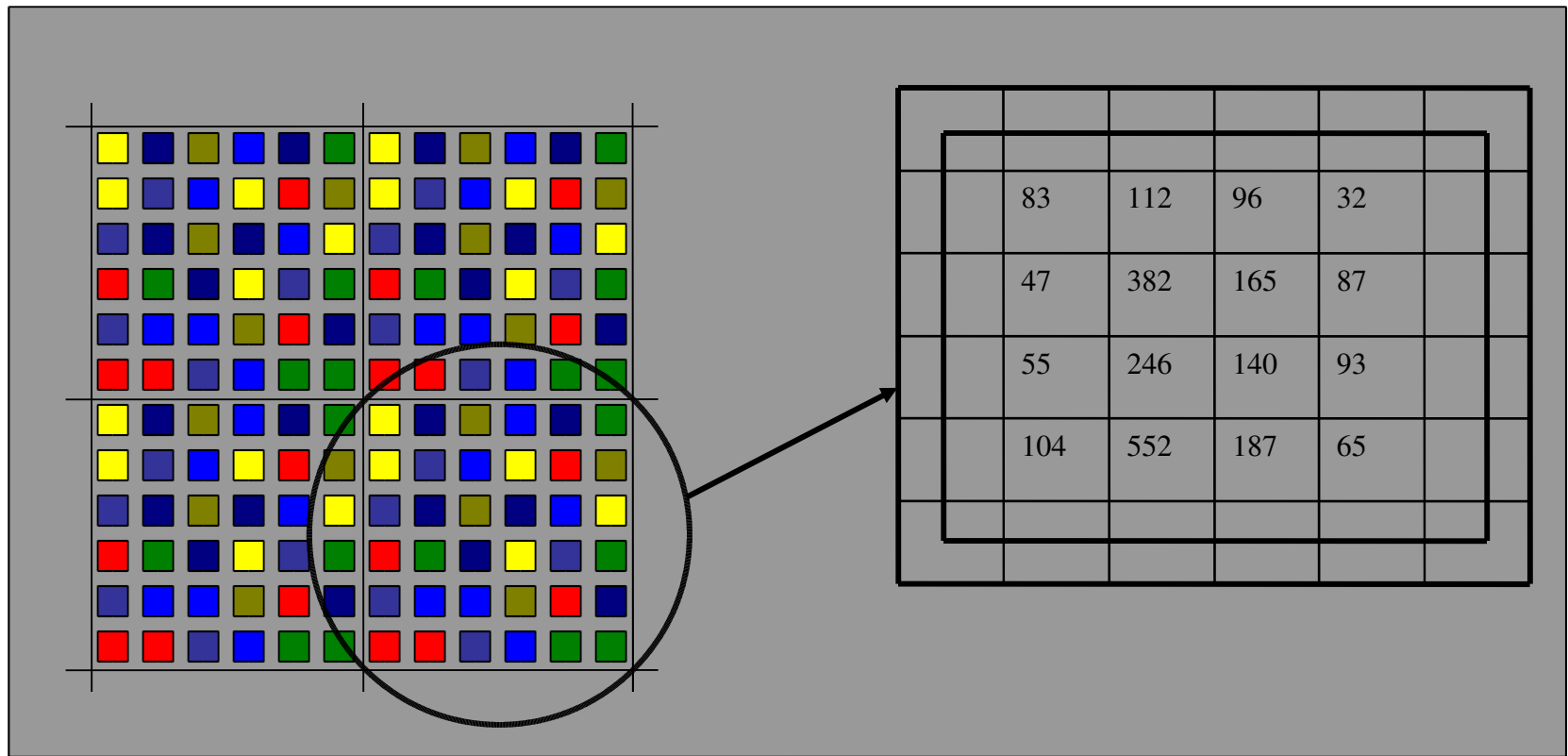


Analýza expresních dat



- rozpoznávání obrazu

Analýza expresních dat

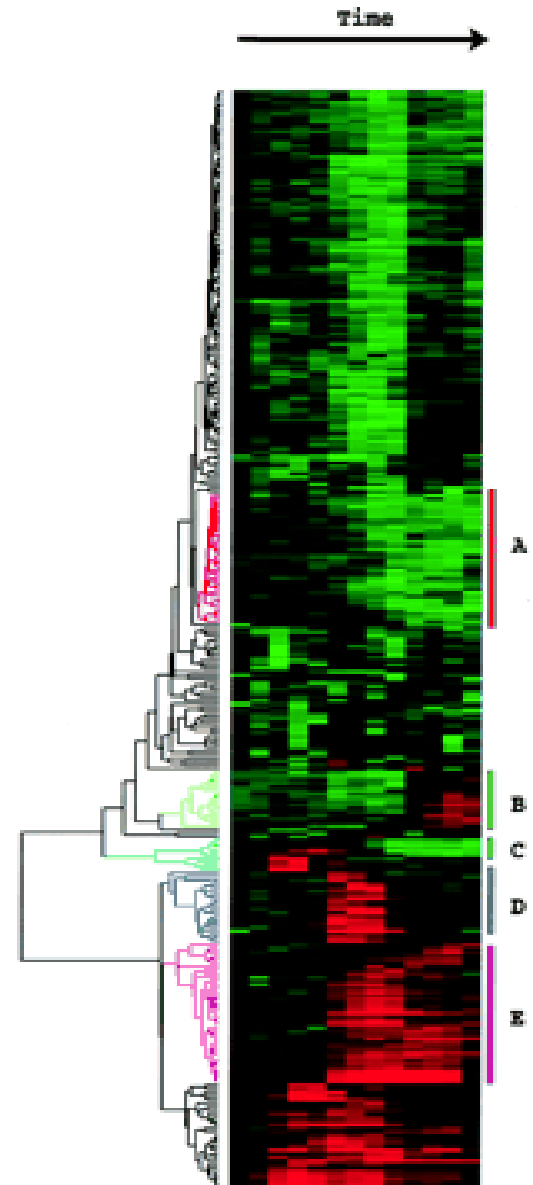


Analýza expresních dat

- měření s nízkým počtem opakování
- rozpoznávání obrazu
- převod na matici číselných hodnot
- normalizace
- statistika jednotlivých pozic (průměrná hodnota, rozptyl, eliminace extrémních hodnot, atd.)
- kontrast: dvě sady experimentálních podmínek
- shluková analýza (clustering)
- analýza v kontextu funkce
- analýza v kontextu sekvence (gen, promotor)
- analýza časových řad (koregulované geny, periodičita)

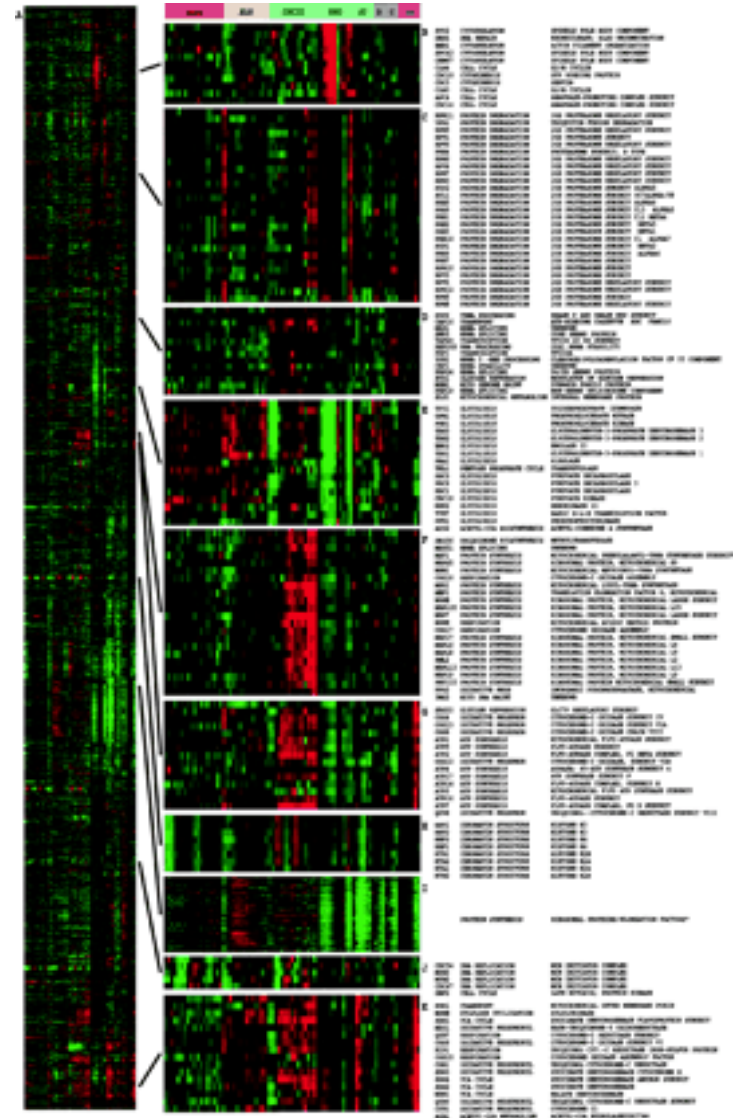
Analýza expresních dat

Analýza časových řad

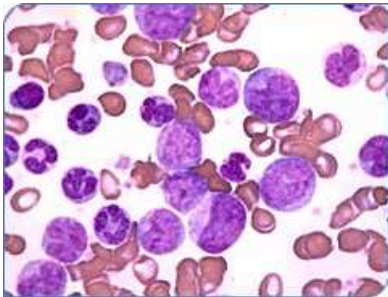


Analýza expresních dat

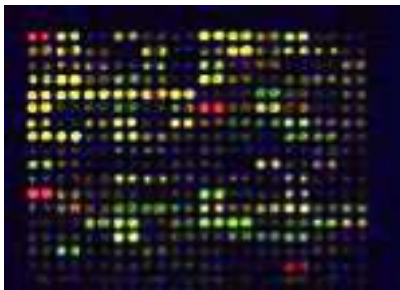
Komplexní data z více časových řad



Analýza expresních dat – medicína



Rakovina – nekontrolované dělení buněk v důsledku mutace (onkogény, supresory nádorů).



Měření genové exprese nachází uplatnění v diagnostice.

- o jaký druh rakoviny se jedná ?
- jaká je prognóza pro daný druh ?

Analýza expresních dat - medicína

Rakovina kostní dřeně – leukemie

Myelogenní, lymfocytická, akutní, chronická ?

Acute Myelogenous Leukemia (AML) Acute Lymphocytic Leukemia (ALL)

Není produkováno dostatečné množství krvinek

Anemia -> slabost, infekce, snížená sráženlivost krve

AML: 10,000 případů ročně

ALL: 3,500 dospělí / 2,400 novorozenci

AML nebo ALL ?



Analýza expresních dat - medicína

Diagnóza ve specializované laboratoři

ALL: 58% pravděpodobnost přežití

AML: 14% pravděpodobnost přežití

Léčba: chemoterapie, transplantace kostní dřeně

ALL: corticosteroidy, vincristin, methotrexat, L-asparaginasa

AML: daunorubicin, cytarabin

Správná diagnóza je velice důležitá pro nasazení správné léčby

Microarray – mohlo by podporovat rozhodování

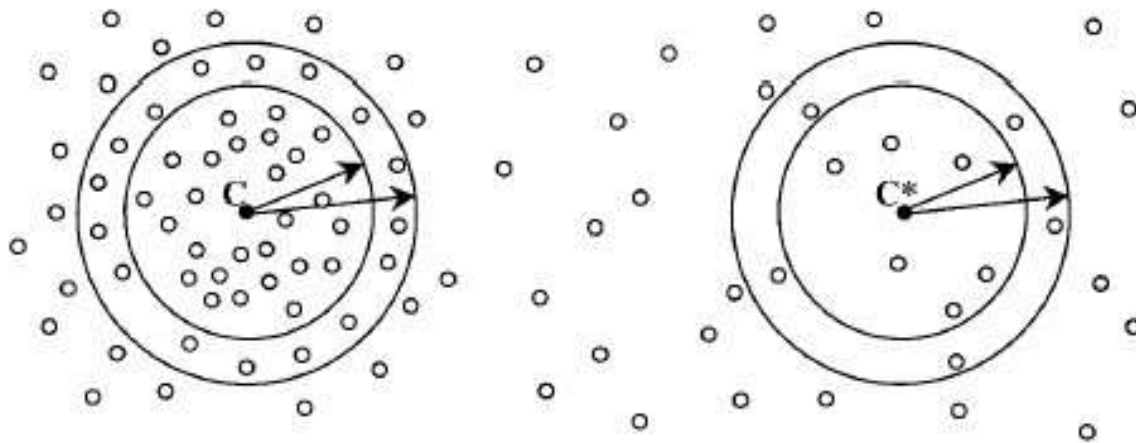
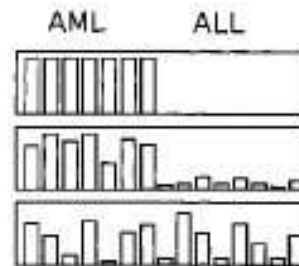


Analyza expresních dat

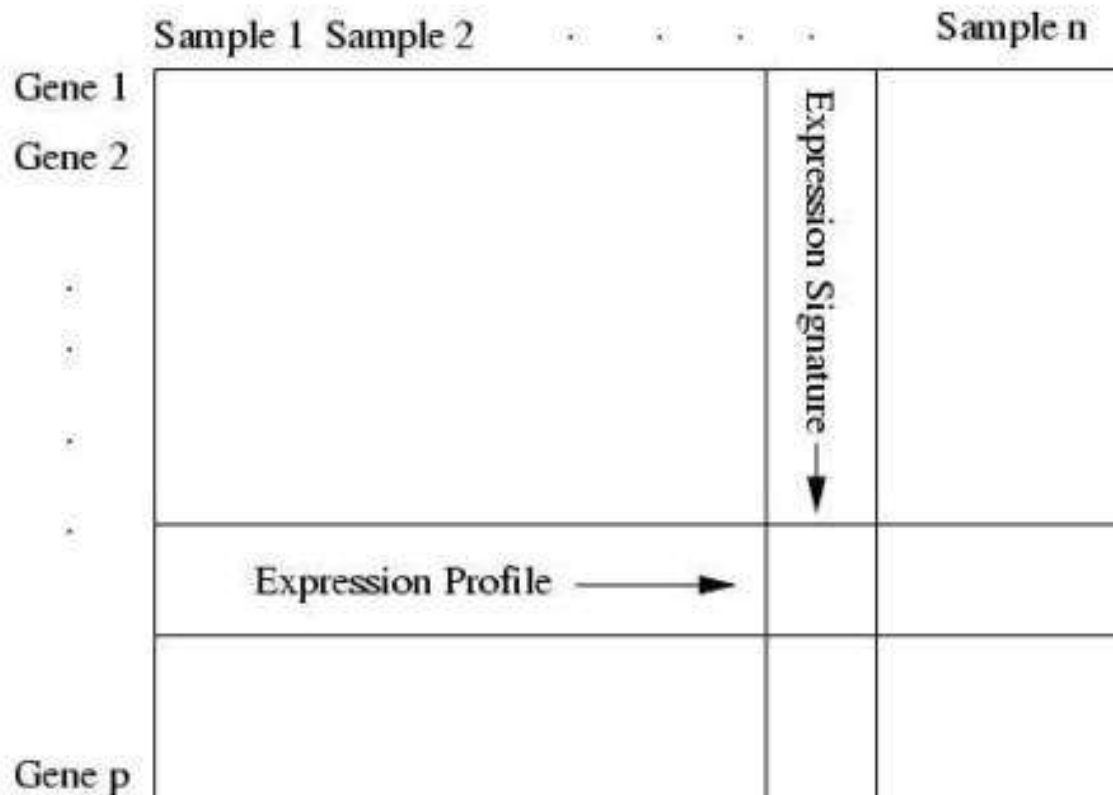
$$c = (1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$$

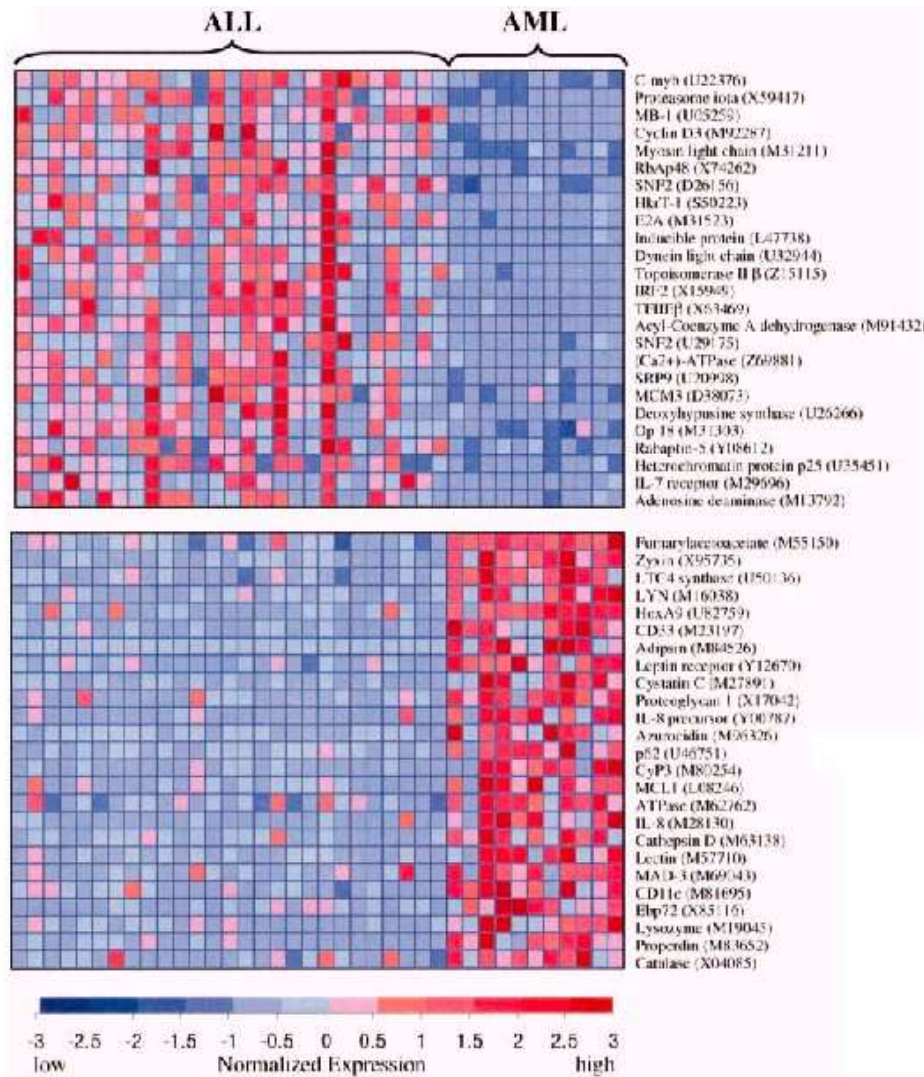
$$\text{gene}_1 = (e_1, e_2, e_3, \dots, e_{12})$$

$$\text{gene}_2 = (e_1, e_2, e_3, \dots, e_{12})$$



Analýza expresních dat



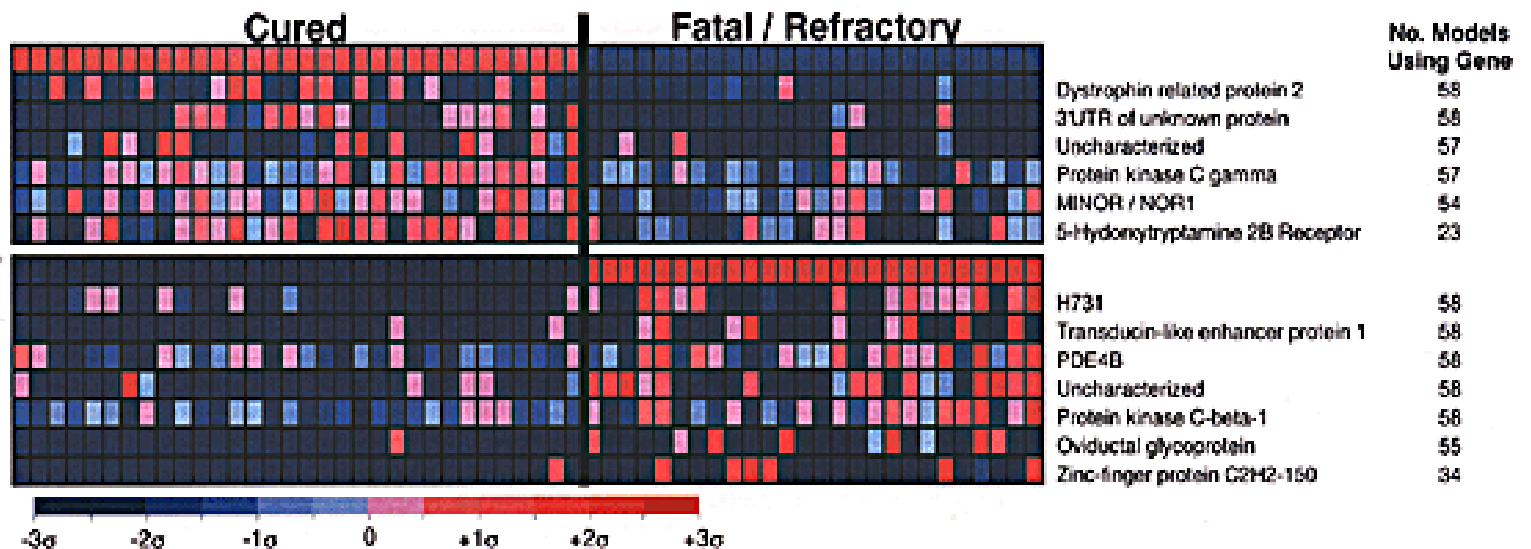


Analýza expresních dat

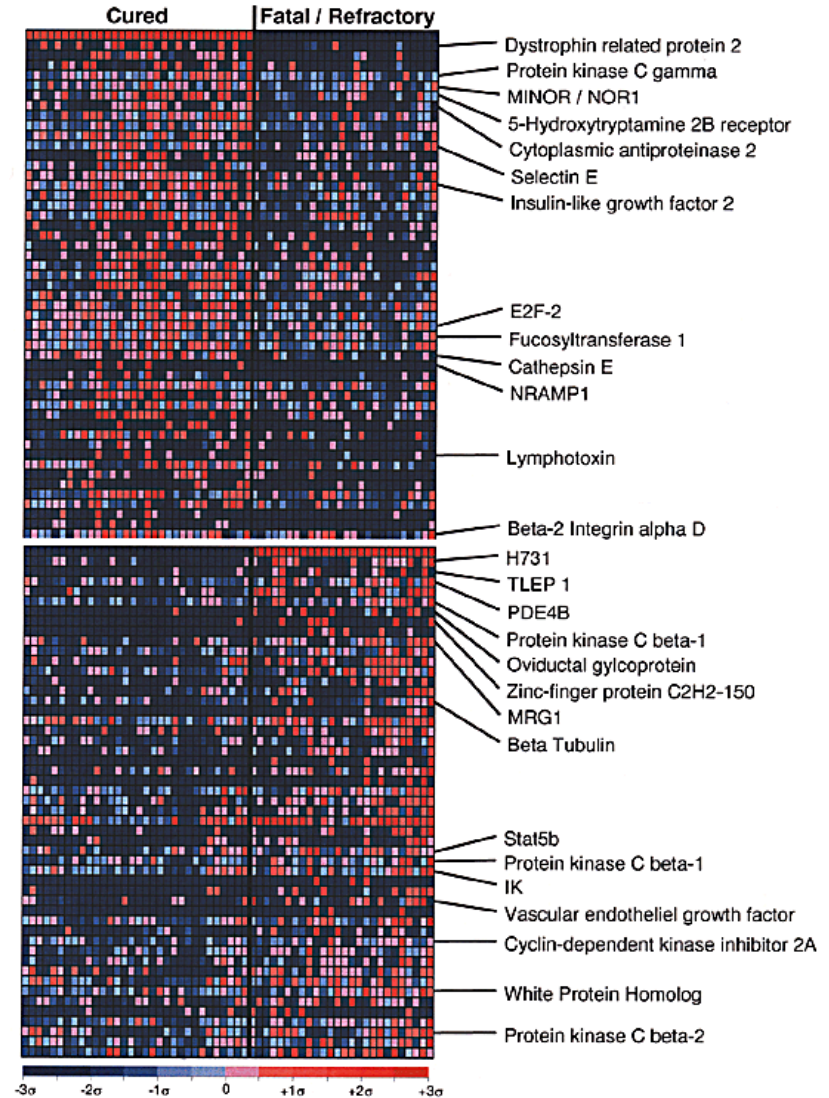
Řádky:
jednotlivé geny

Sloupce:
vzorky, pacienti

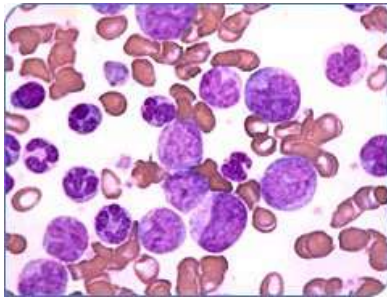
Analýza expresních dat



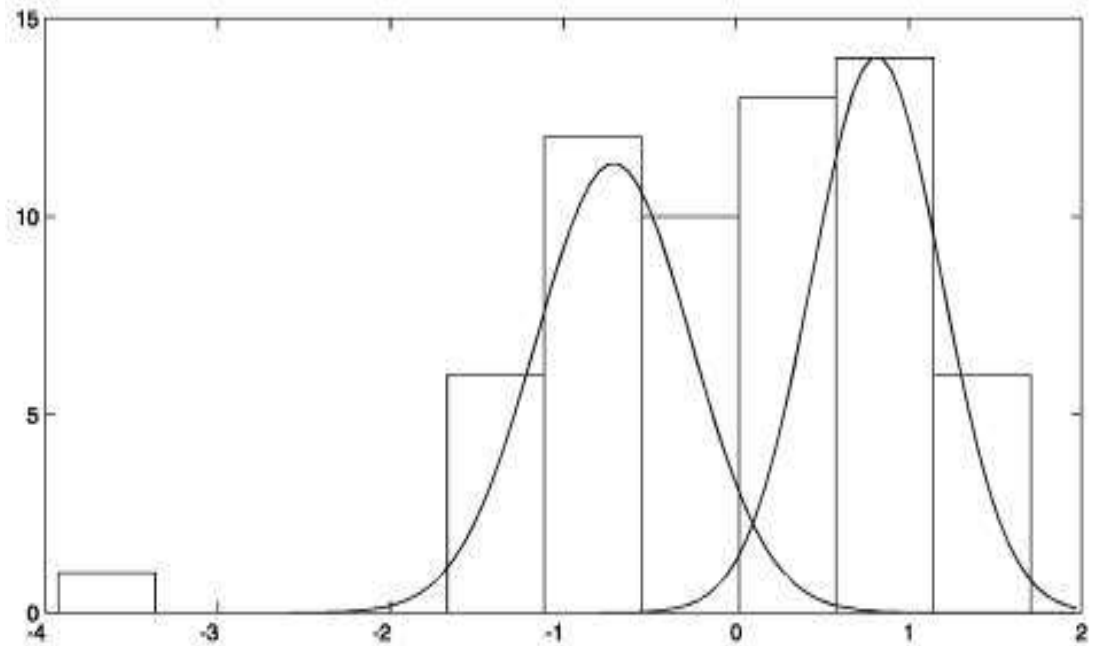
Analýza expresních dat



Analýza expresních dat



Problém – data jsou smíšená



Analýza expresních dat

Statistický program/programovací prostředí R

<http://www.stat.uni-muenchen.de/~strimmer/notes/rexpress.html>

Jazyky a prostředí používané bioinformatiky

UNIX

Perl/Python

C/C++/Java

R/Mathlab/Mathematica

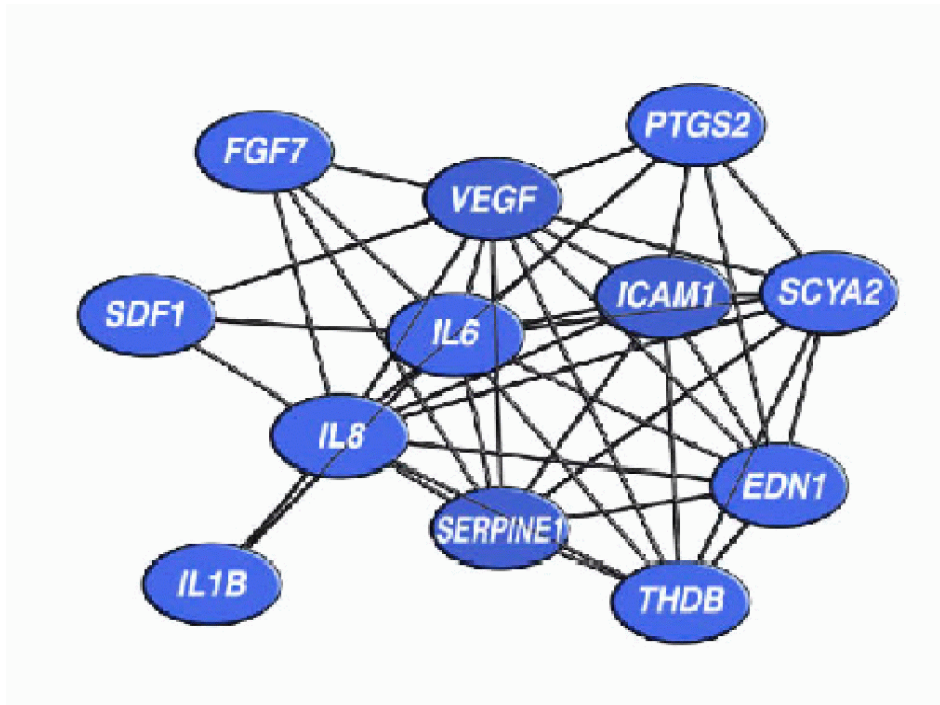
CGI, client/server

SQL

Analýza dat v biomedicínské literatuře

Hlavním uložištěm našich znalostí o aktivitě a funkci genů/proteinů je vědecká literatura. Bionformatickou analýzou těchto textů lze identifikovat skryté závislosti, které mohou být způsobeny vzájemnou interakcí genů/proteinů, jejich společnou funkcí nebo jinými vztahy, které nebyly předmětem cíleného zkoumání.

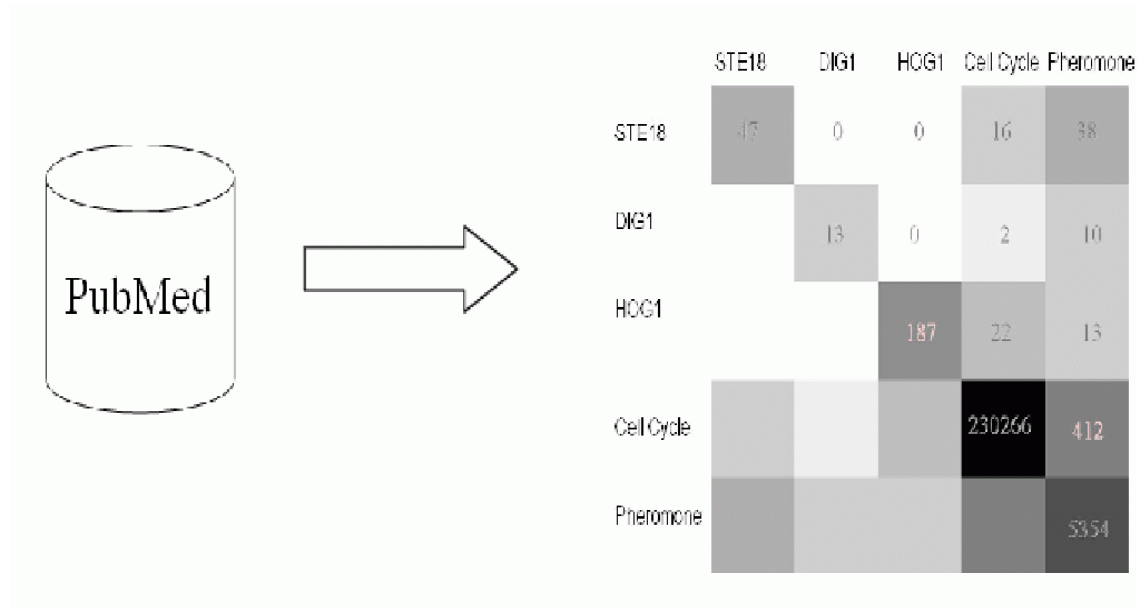
Analýza dat v biomedicínské literatuře



Graf znázorňující
spoluvýskyt proteinů ve
vědeckých člancích.

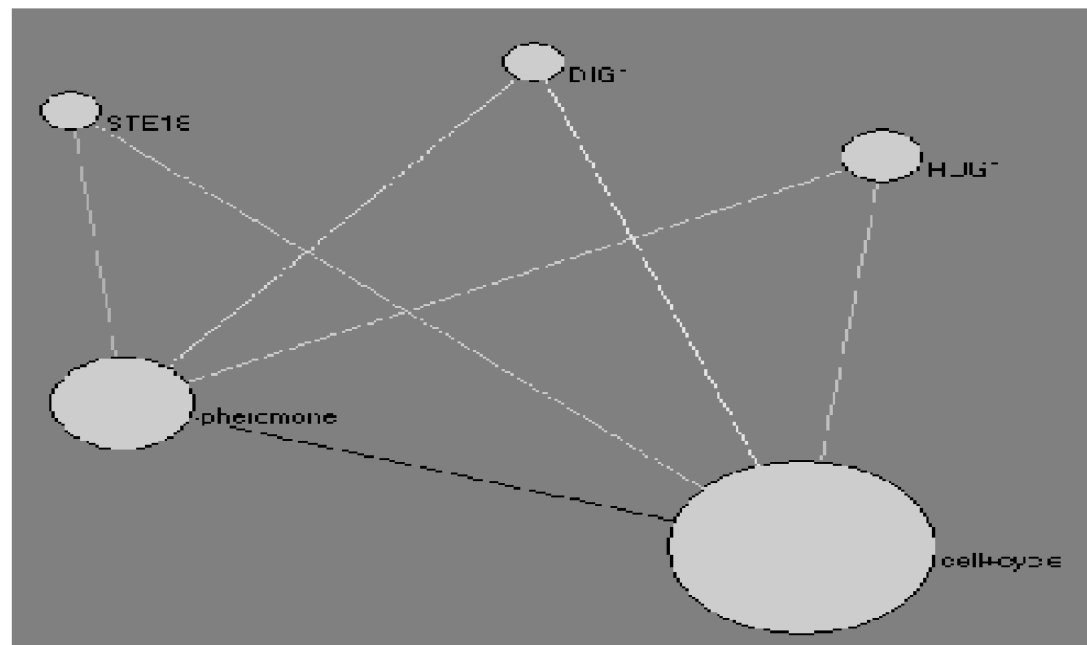
Každá hrana grafu
představuje samostatný
článek v databázi
PubMed.

Analýza dat v biomedicínské literatuře



Konstrukce matice spoluvýskytu jednotlivých termínů, kterými mohou být například názvy genů, funkce z GO a pod.

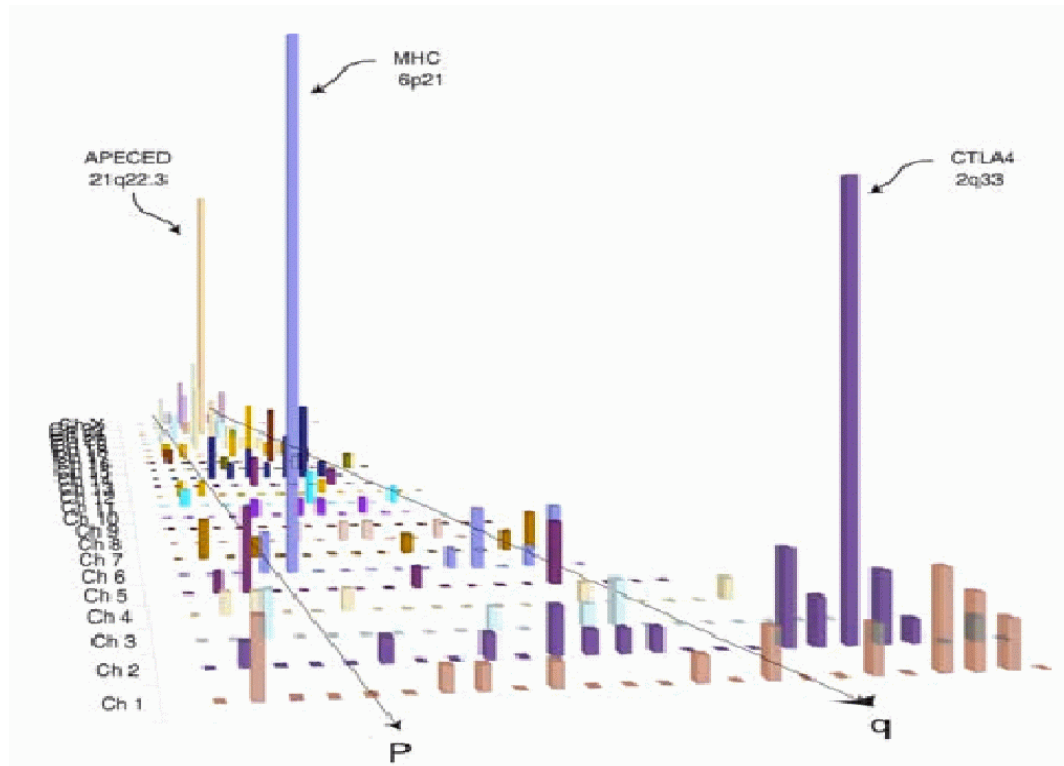
Analýza dat v biomedicínské literatuře



Analýza dat v biomedicínské literatuře

PubMatrix <http://pubmatrix.grc.nia.nih.gov>

Analýza dat v biomedicínské literatuře

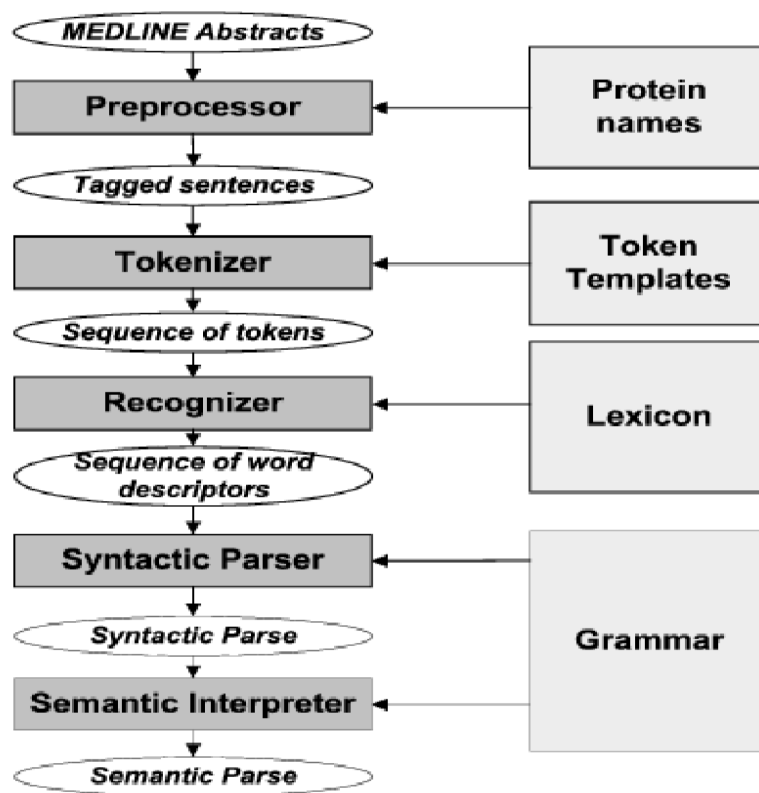


Společný výskyt výrazu “autoimunita” a názvů chromozómů nebo jejich částí.

Analýza dat v biomedicínské literatuře

- tokenizace a vyhledání základního tvaru ((3,4)-metyl 1-hexanol)
- identifikace názvů genů a proteinů (ID)
- mapování volného textu na GO kategorie (sensitivita 60%, specificita 90%)
- mapování ID dvou proteinů vyskytujících se ve volném textu na zjištěný vztah (např. protein A je enzymem a protein B substrátem reakce B-B' katalyzované proteinem A a popsané termínem GO)

Analýza dat v biomedicínské literatuře



5-40% chyb na jednotlivých stupních analýzy

Fig. 1. The components and processing steps of the MedScan system.

Ariadne Genomics <http://www.ariadnegenomics.com/>



[Home](#) [Site Search](#) [Support](#) [Newsletter](#)

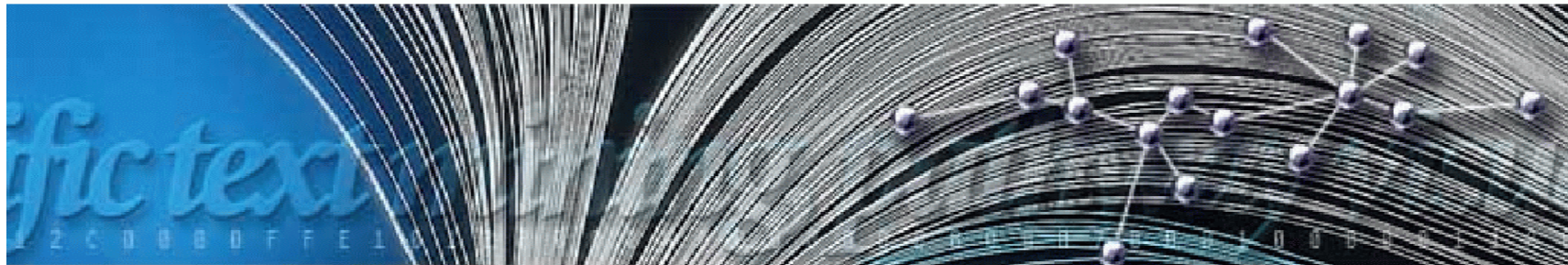
[COMPANY](#)

[PRODUCTS](#)

[TECHNOLOGIES](#)

[SERVICES](#)

[DOWNLOADS](#)



Ariadne Genomics develops user-friendly software tools for biologists in the areas of pathway analysis, sequence data management and automated scientific text processing. Ariadne Genomics products incorporate proprietary Natural Language Processing (NLP) and statistical algorithms designed to functionally interpret novel genetic information.