

MUNI | RECETOX

Specializace

Biomedicínská bioinformatika

přednáška 28.11.2022

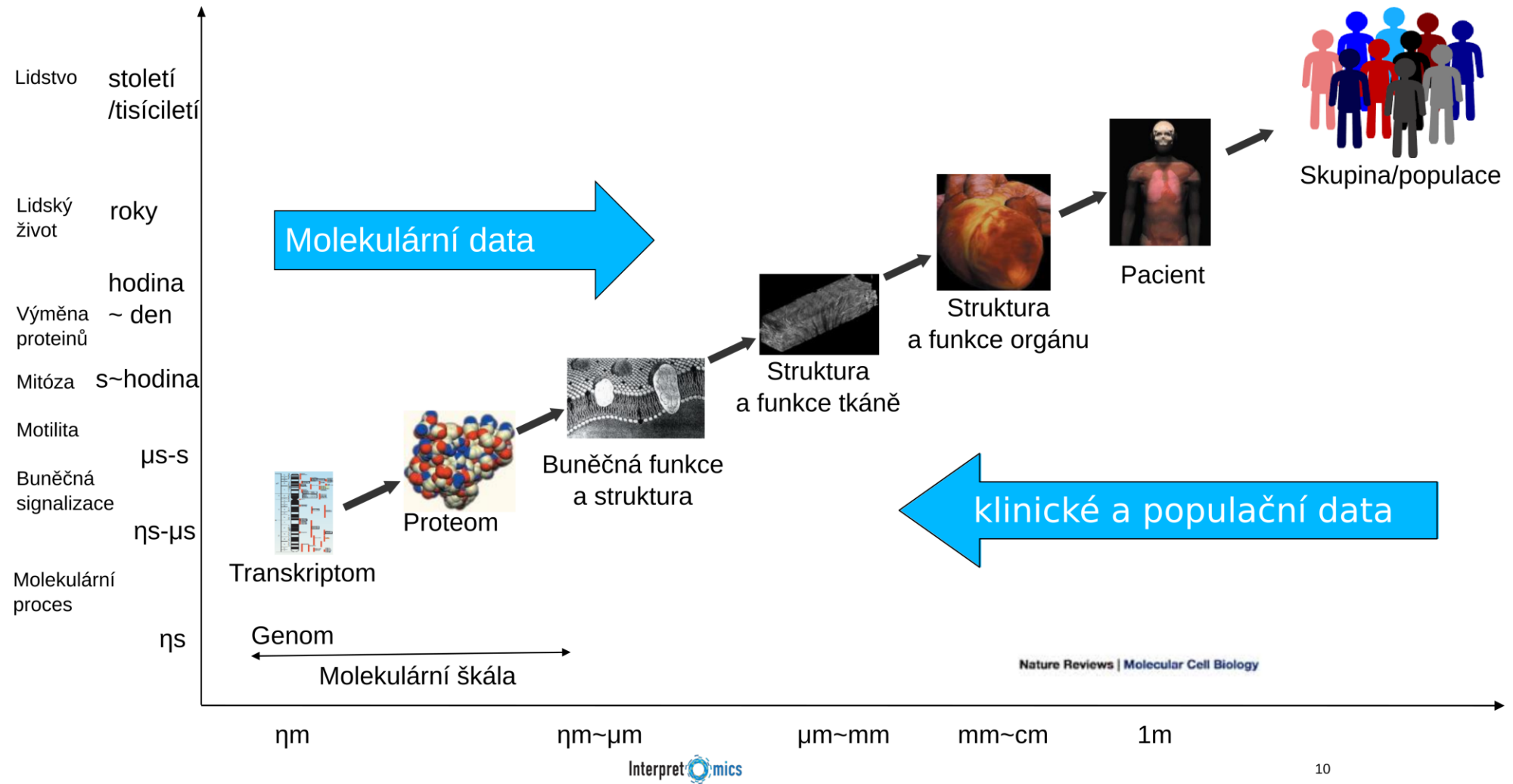
Eva Budinská (budinska@recetox.muni.cz)



Biomedicína

- ... **odvětví lékařské vědy** využívající přírodovědecké principy především fyziologie a biologie v klinické praxi.
- Jedná se o **interdisciplinární obor**, který kombinuje obsah a problémy **experimentální medicíny** s metodami **molekulární biologie** a **buněčné biologie**.
- Zaměřuje se na **molekulární a buněčné základy života** a **patologické změny**, dále zkoumá **příčiny** nemocí, **prevenci** a **kauzální léčbu**.





Moderní biomedicína a molekulární data

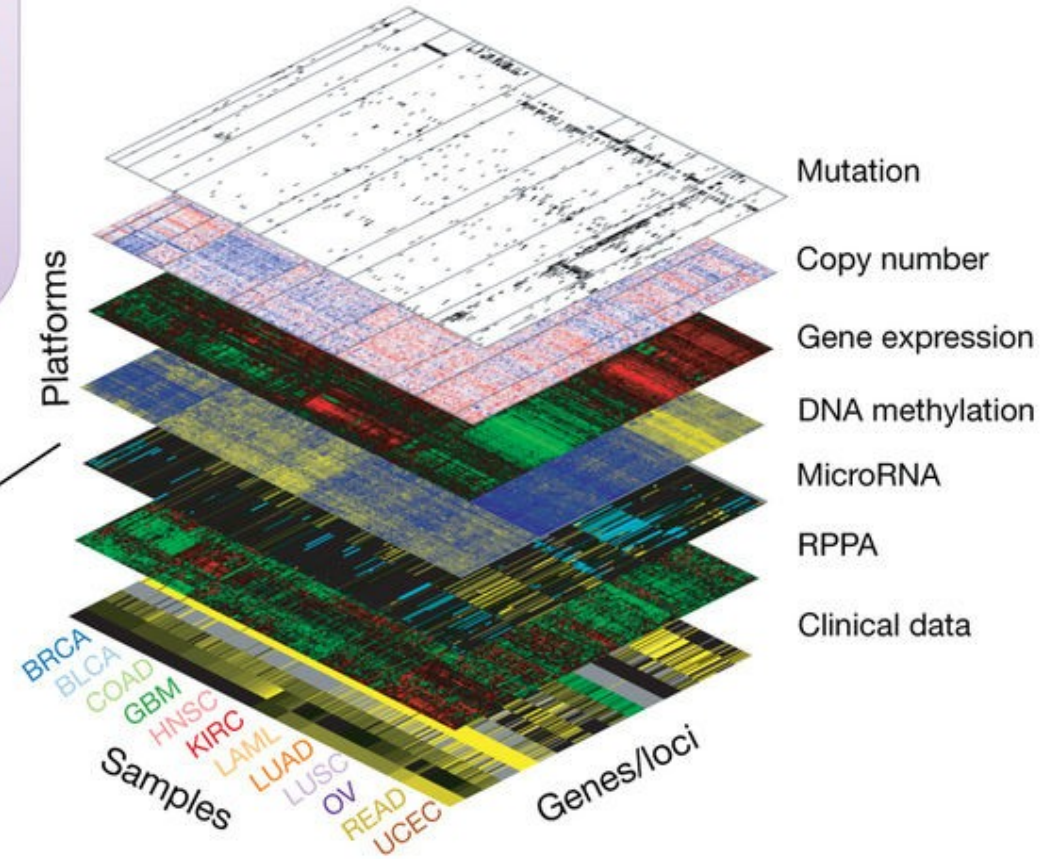


Omicsové experimenty ...

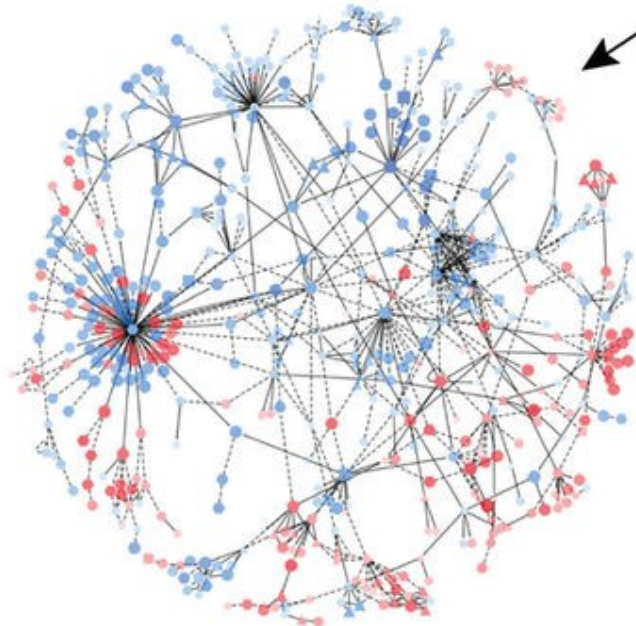
12 tumor types



Omics characterizations



Thematic pathways

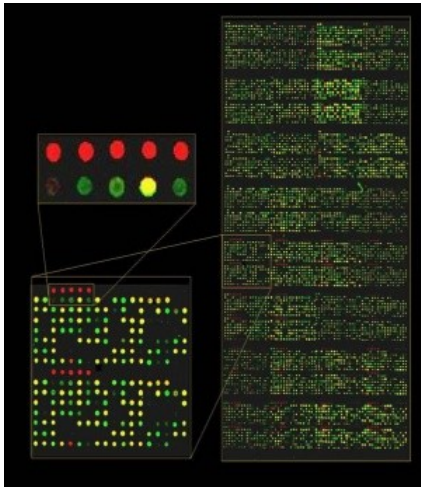


The Human Cancer Genome Atlas (TCGA) projekt

Data z omicsových experimentů

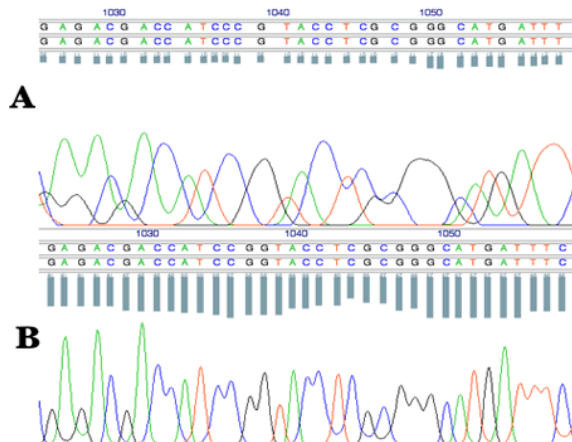
- Moderní vysocepokryvné molekulární technologie produkují obrovské tabulky komplexních dat

Mikročipy



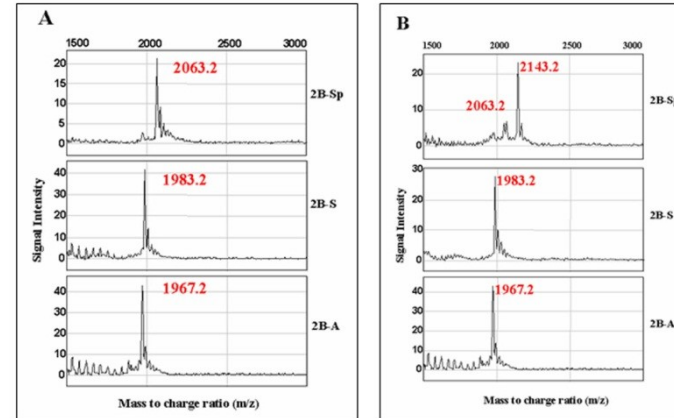
- ☑ Desítky až tisíce genů nebo transkriptů na vzorek

Sekvence DNA



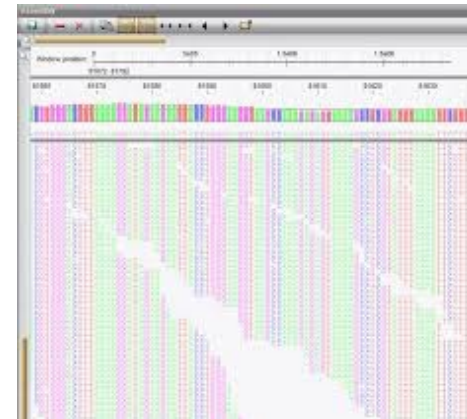
- ☑ Genom s biliony nukleotidů na vzorek

Hmotnostní spektrometrie



- ☑ Tisíce spekter proteinů, metabolitů nebo malých molekul na vzorek

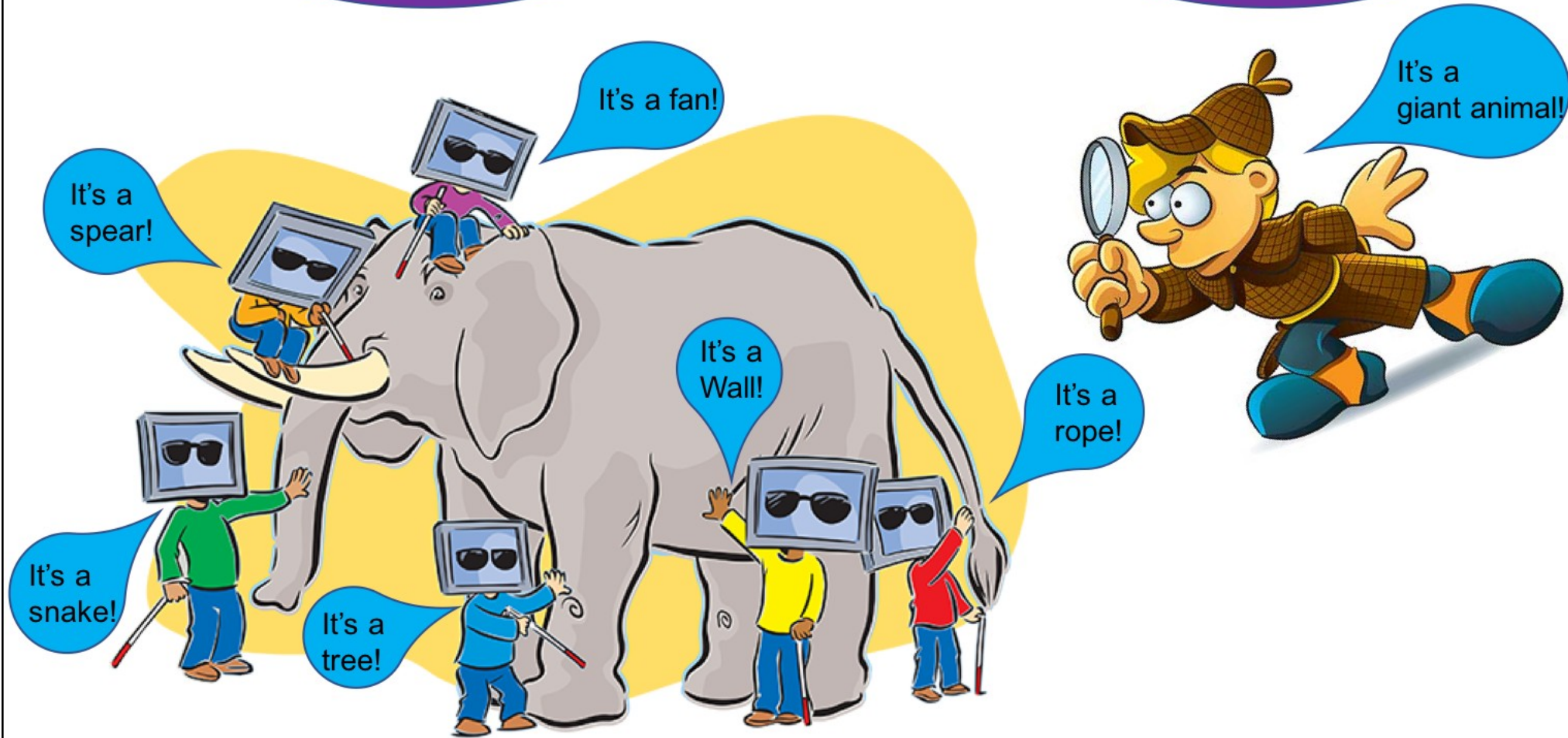
Sekvence nové generace



- ☑ Miliony krátkých čtení DNA na vzorek

Single Omics

Multi-Omics



Co je to biomarker?

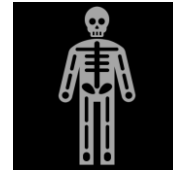
Biologický marker (biomarker):

Charakteristika, která je objektivně měřena a hodnocena jako indikátor normálních biologických procesů, patogenních procesů nebo farmakologických odpovědí na terapeutický zásah.

Biomarkerem může být



Molekula a její stav
(mutace DNA,
hodnota exprese
miRNA, zvýšená
hladina proteinu...)



Aktivita buněk v
konkrétních
oblastech (lymfocyty
v invazivním frontu
nádoru)



**Přítomnost
mikroorganismu**



Proces (zvýšená
proliferace,
přítomnost stromální
reakce v nádoru, ...)



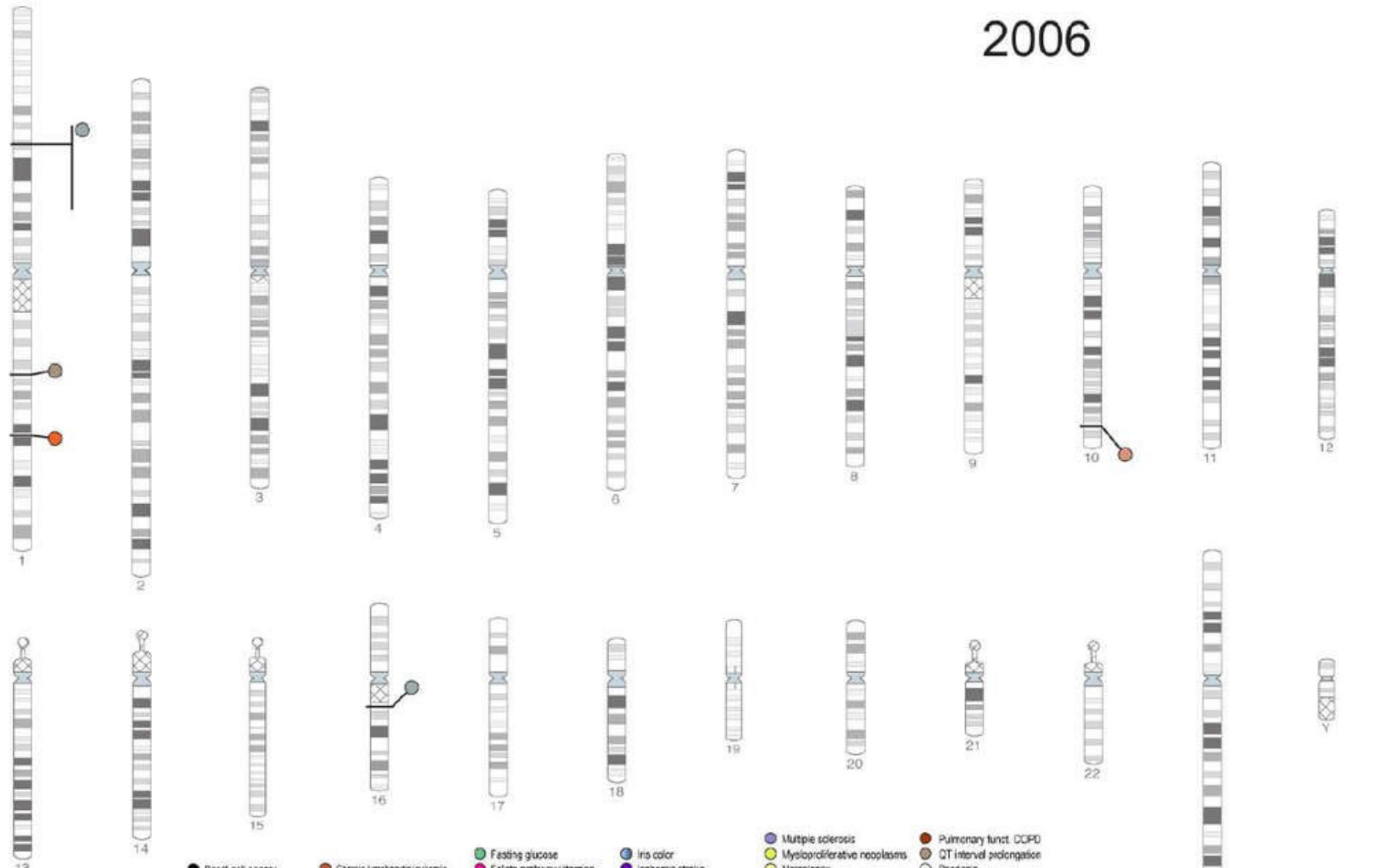
....



**Využití jednotlivých
biomarkerů v
rozhodovacím
PRAVIDLE
(modelu/testu)**



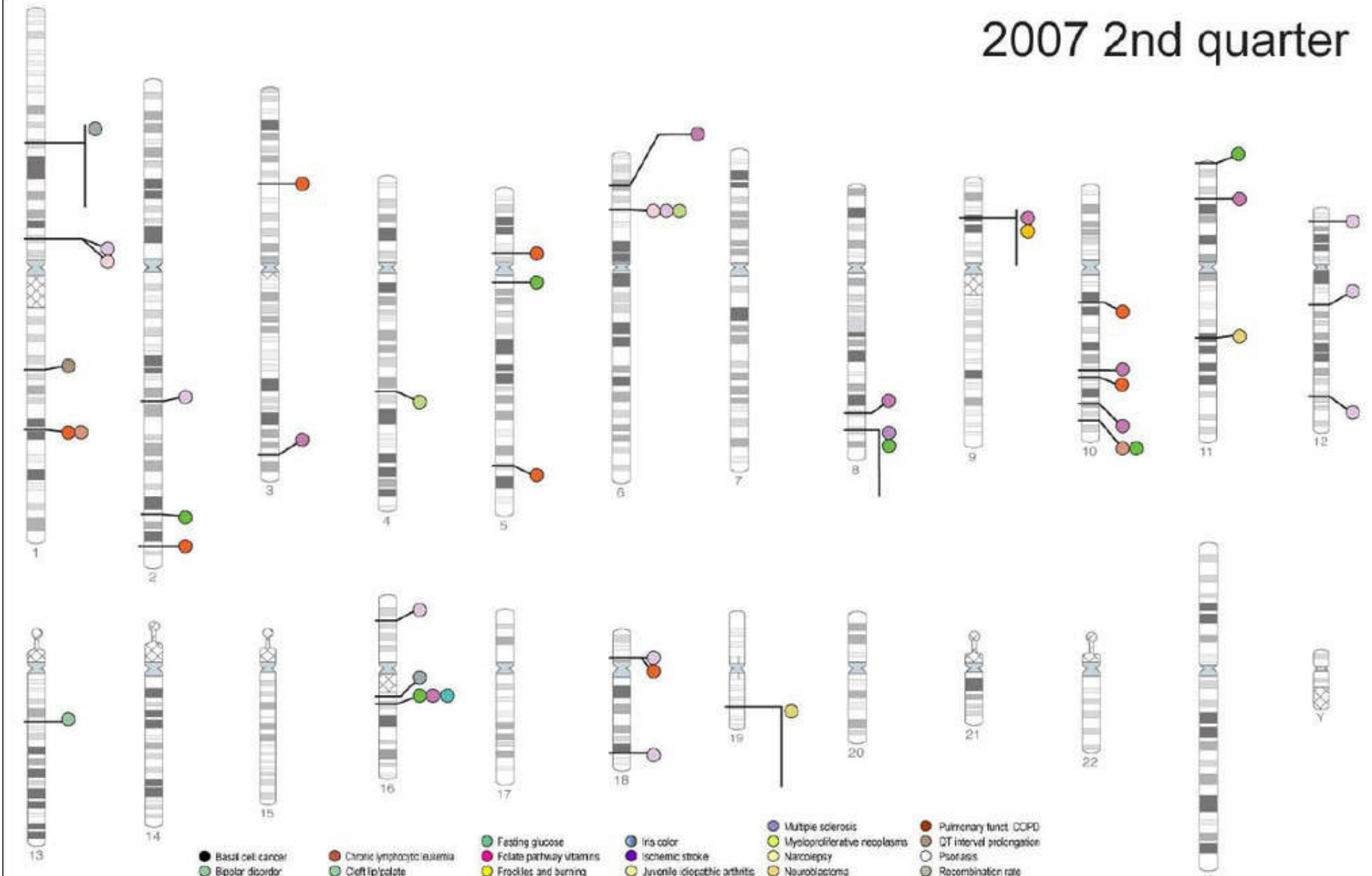
2006



- | | | | | | | |
|--|-------------------------------|-------------------------------|---------------------------------|---------------------------------|--------------------------------|---|
| ● Adiponectin levels | ● Basal cell cancer | ● Celiac lymphocytic leukemia | ● Fasting glucose | ● Iris color | ● Multiple sclerosis | ● Pulmonary funct. COPD |
| ● Age-related macular degeneration | ● Bipolar disorder | ● Cleft lip/palate | ● Folate pathway vitamins | ● Ischemic stroke | ● Myeloproliferative neoplasms | ● QT interval prolongation |
| ● AIDS progression | ● Bladder cancer | ● Cognitive function | ● Freckles and burning | ● Juvenile idiopathic arthritis | ● Narcolepsy | ● Psoriasis |
| ● Alzheimer disease APOE ϵ 4 carriers | ● Blood or brown hair | ● Colorectal cancer | ● Galstones | ● Kidney stones | ● Neuroblastoma | ● Recombination rate |
| ● Amyotrophic lateral sclerosis | ● Blood pressure | ● Coronary disease | ● Hair color | ● LDL cholesterol | ● Nicotine dependence | ● Renal function |
| ● Asthma | ● Blue or green eyes | ● Creutzfeldt-Jakob disease | ● Heart rate | ● Liver cancer | ● Obesity | ● Restless legs syndrome |
| ● Atopic dermatitis | ● BMI, waist circumference | ● Crohn's disease | ● Height | ● Liver enzymes | ● Osteoarthritis | ● Rheumatoid arthritis |
| ● Atrial fibrillation | ● Bone density | ● Drug-induced liver injury | ● Hepatitis B | ● Malaria | ● Other metabolic traits | ● Serum bilirubin |
| ● Attention Deficit Hyperactivity Disorder | ● Breast cancer | ● Eosinophil count | ● Hirschsprung's disease | ● Male pattern baldness | ● Panic disorder | ● Soluble ICAM-1 |
| ● Autism | ● C-reactive protein | ● Essential tremor | ● HDL cholesterol | ● MCP-1 | ● Peripheral arterial disease | ● Statin-induced myopathy |
| | ● Carbonyl/sulphhydryl levels | ● Exfoliation glaucoma | ● Idiopathic pulmonary fibrosis | ● Mean platelet volume | ● Plasma LF (a) levels | ● Stroke |
| | ● Cellar disease | ● F cell distribution | ● Inflammatory bowel disease | ● Monarchs & menopause | ● Primary biliary cirrhosis | ● Systemic lupus erythematosus |
| | | | ● Intraocular aneurysm | | ● Prostate cancer | ● Systemic lupus erythematosus in women |
| | | | | | ● Protein levels | ● Testicular germ cell tumor |
| | | | | | | ● Thyroid cancer |
| | | | | | | ● Total cholesterol |
| | | | | | | ● Triglycerides |
| | | | | | | ● Skin pigmentation by reflectance spectroscopy |
| | | | | | | ● Type 1 diabetes |
| | | | | | | ● Type 2 diabetes |
| | | | | | | ● Venous Thromboembolism |
| | | | | | | ● Vitamin B12 levels |
| | | | | | | ● Warfarin dose |
| | | | | | | ● Weight |
| | | | | | | ● YKL-40 levels |

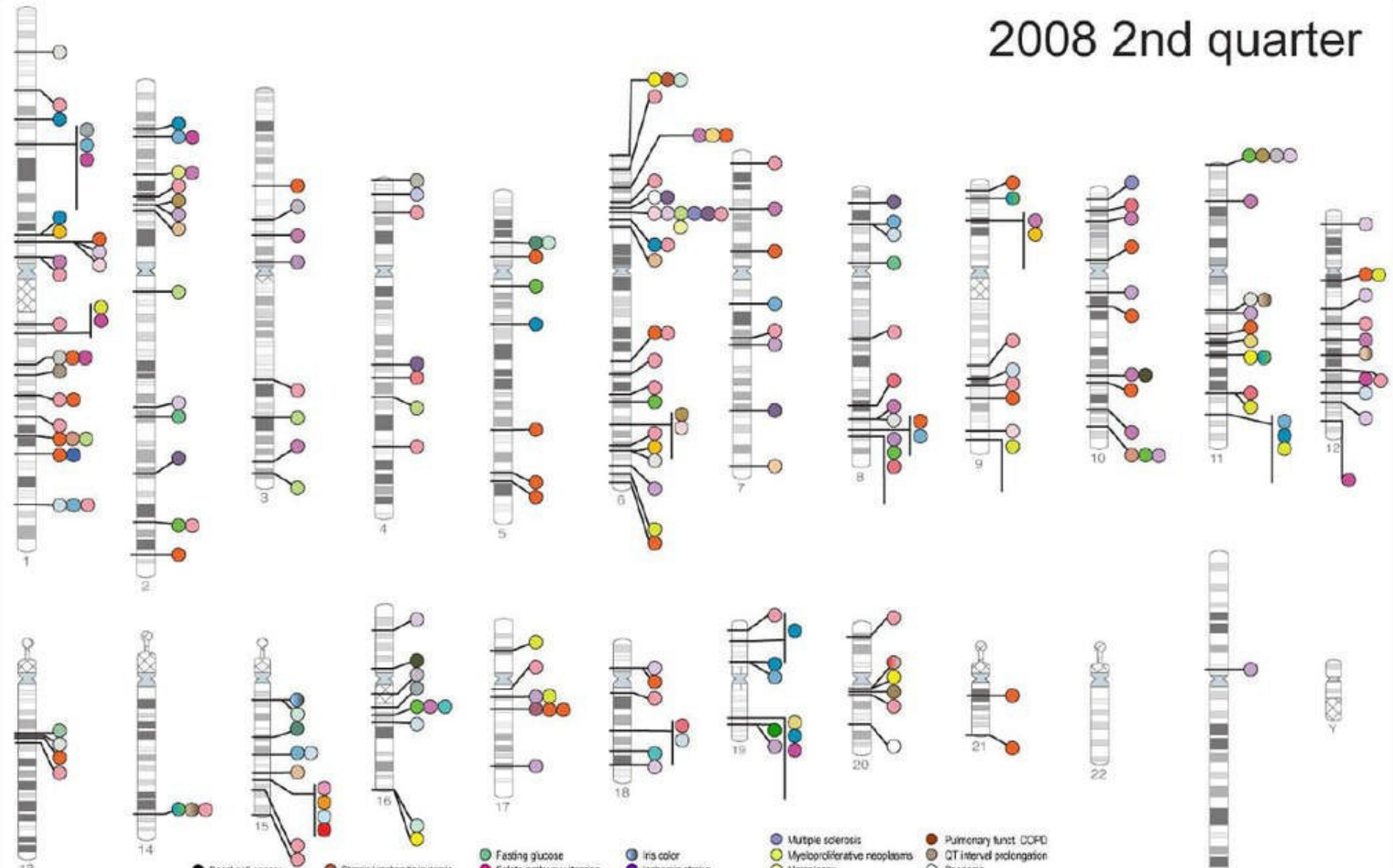
ETOX

2007 2nd quarter



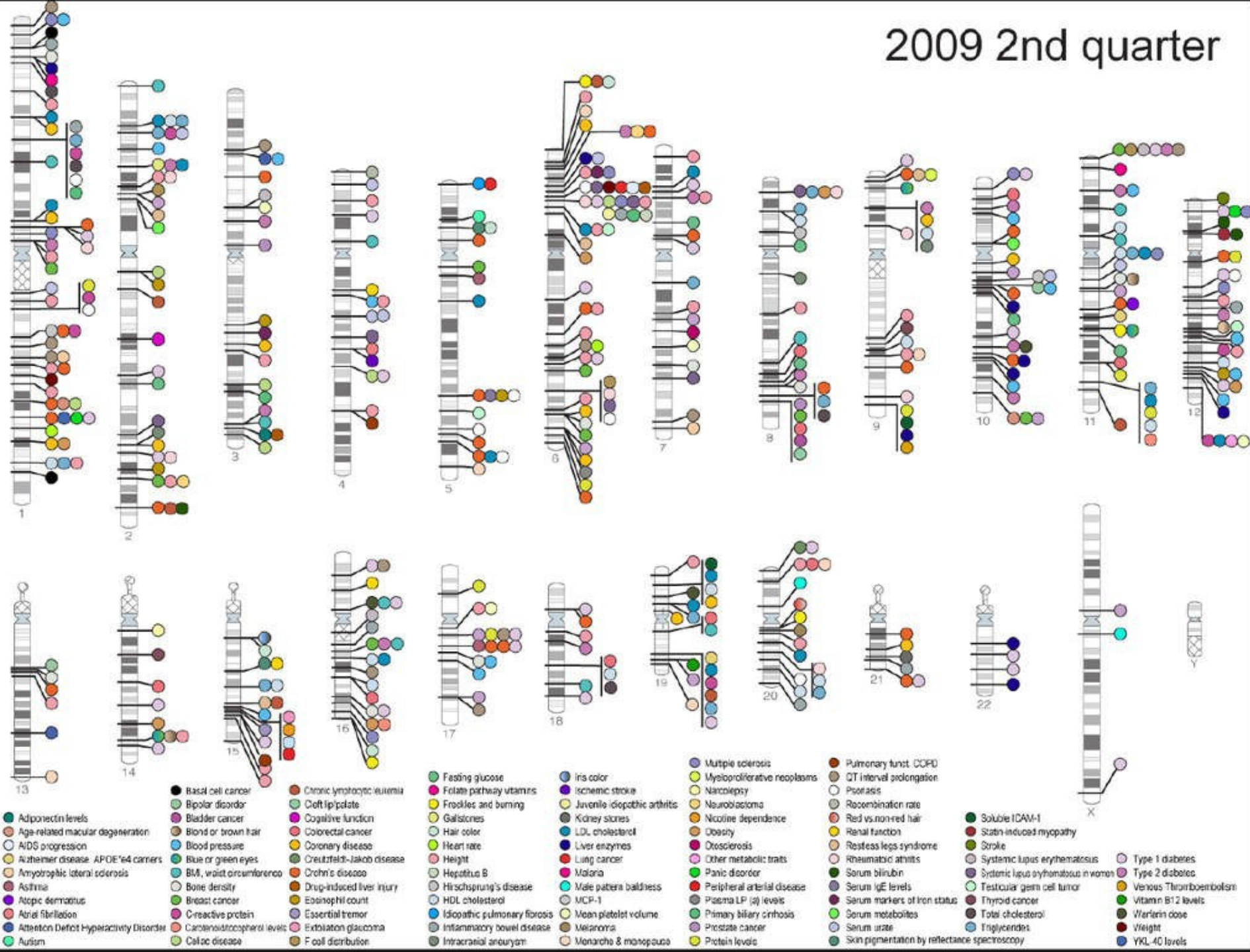
- | | | | | | | | | |
|---|-----------------------------|--------------------------------|---------------------------------|---------------------------------|--------------------------------|---|---|--------------------------|
| ● Adiponectin levels | ● Basal cell cancer | ● Chronic lymphocytic leukemia | ● Fasting glucose | ● Iris color | ● Multiple sclerosis | ● Pulmonary funct. COPD | ● Soluble ICAM-1 | ● Type 1 diabetes |
| ● Age-related macular degeneration | ● Bipolar disorder | ● Cleft lip/palate | ● Folate pathway vitamins | ● Ischemic stroke | ● Myeloproliferative neoplasms | ● QT interval prolongation | ● Statin-induced myopathy | ● Type 2 diabetes |
| ● AIDS progression | ● Bladder cancer | ● Cognitive function | ● Freckles and blemish | ● Juvenile idiopathic arthritis | ● Narcolepsy | ● Psoriasis | ● Stroke | ● Venous Thromboembolism |
| ● Alzheimer disease APOE ϵ 4 carrier | ● Blood or brown hair | ● Colorectal cancer | ● Galstones | ● Kidney stones | ● Neuroblastoma | ● Recombination rate | ● Systemic lupus erythematosus | ● Vitamin B12 levels |
| ● Amyotrophic lateral sclerosis | ● Blood pressure | ● Coronary disease | ● Hair color | ● LDL cholesterol | ● Nicotine dependence | ● Red vs non-red hair | ● Systemic lupus erythematosus in women | ● Warfarin dose |
| ● Asthma | ● Blue or green eyes | ● Creutzfeldt-Jakob disease | ● Heart rate | ● Liver enzymes | ● Obesity | ● Renal function | ● Testicular germ cell tumor | ● Weight |
| ● Atopic dermatitis | ● BMI, waist circumference | ● Crohn's disease | ● Height | ● Malaria | ● Osteoporosis | ● Restless legs syndrome | ● Thyroid cancer | ● YKL-40 levels |
| ● Atrial fibrillation | ● Bone density | ● Drug-induced liver injury | ● Hepatitis B | ● Male pattern baldness | ● Other metabolic traits | ● Rheumatoid arthritis | ● Total cholesterol | |
| ● Attention Deficit Hyperactivity Disorder | ● Breast cancer | ● Eosinophil count | ● HDL cholesterol | ● MCP-1 | ● Panic disorder | ● Serum bilirubin | ● Triglycerides | |
| ● Autism | ● C-reactive protein | ● Essential tremor | ● Idiopathic pulmonary fibrosis | ● Hirschsprung's disease | ● Peripheral arterial disease | ● Serum IgE levels | | |
| | ● Camels/d/ccephalic levels | ● Exfoliation glaucoma | ● Inflammatory bowel disease | ● Intracranial aneurysm | ● Pima LP (a) levels | ● Serum markers of iron status | | |
| | ● Celiac disease | ● F cell distribution | ● Intracranial aneurysm | ● Intracranial aneurysm | ● Primary biliary cirrhosis | ● Serum metabolites | | |
| | | | | | ● Prostate cancer | ● Serum urate | | |
| | | | | | ● Protein levels | ● Skin pigmentation by reflectance spectroscopy | | |

2008 2nd quarter

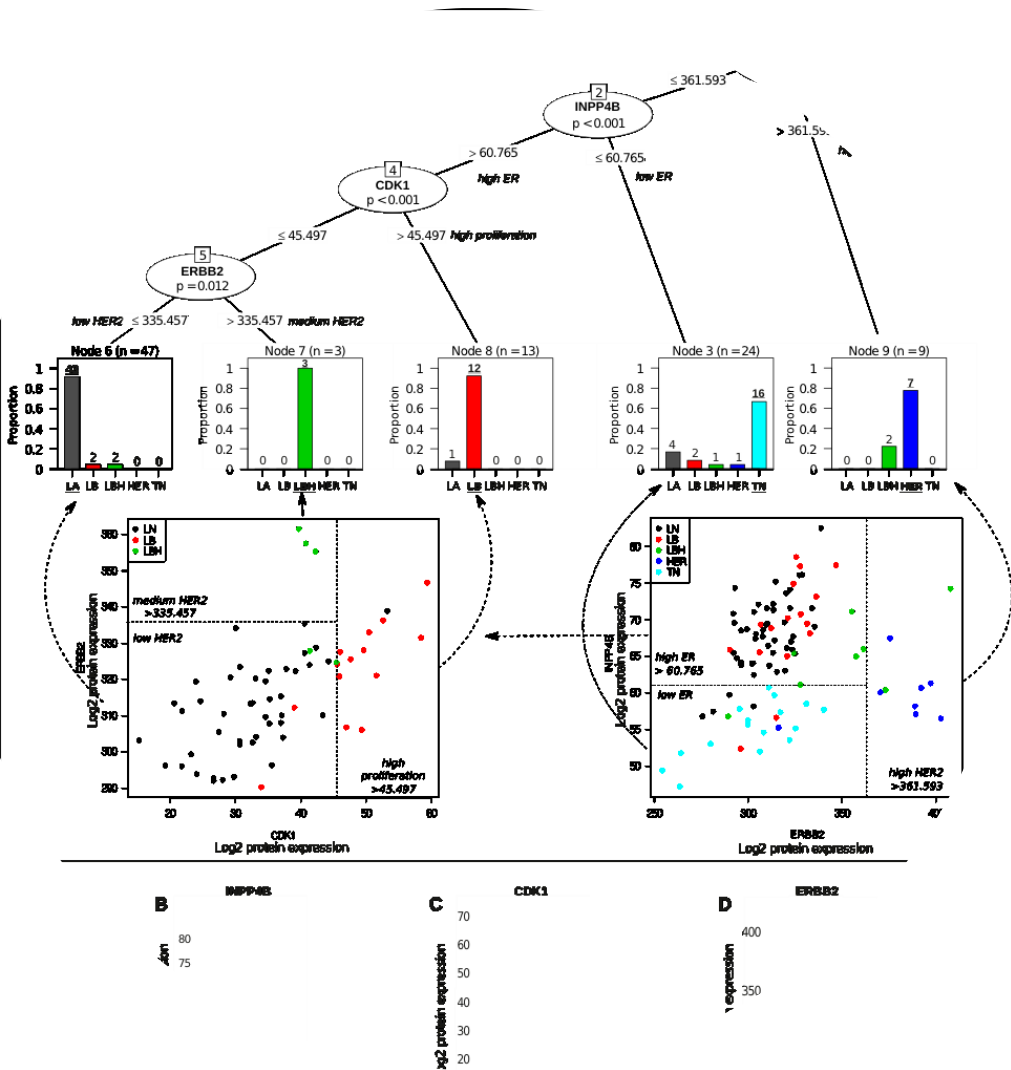


- | | | | | | | | |
|--|--|---|---|---|---|--|--|
| <ul style="list-style-type: none"> ● Adiponectin levels ● Age-related macular degeneration ○ AIDS progression ○ Alzheimer disease APOEϵ4 carriers ○ Amyotrophic lateral sclerosis ○ Asthma ○ Atopic dermatitis ○ Atrial fibrillation ○ Attention Deficit Hyperactivity Disorder ○ Autism | <ul style="list-style-type: none"> ● Basal cell cancer ○ Bipolar disorder ○ Bladder cancer ○ Blood or brown hair ○ Blood pressure ○ Blue or green eyes ○ BMI, waist circumference ○ Bone density ○ Breast cancer ○ C-reactive protein ○ Carotenoid/prophyl levels ○ Celiac disease | <ul style="list-style-type: none"> ○ Chronic lymphocytic leukemia ○ Deft lipoplate ○ Cognitive function ○ Colorectal cancer ○ Coronary disease ○ Creutzfeldt-Jakob disease ○ Crohn's disease ○ Drug-induced liver injury ○ Eosinophil count ○ Essential tremor ○ Exfoliation glaucoma ○ F cell distribution | <ul style="list-style-type: none"> ○ Fasting glucose ○ Folate pathway vitamins ○ Freckles and burning ○ Galactose ○ Hair color ○ Heart rate ○ Height ○ Hepatitis B ○ Hirschsprung's disease ○ HDL cholesterol ○ Idiopathic pulmonary fibrosis ○ Inflammatory bowel disease ○ Intracranial aneurysm | <ul style="list-style-type: none"> ○ Iris color ○ Ischemic stroke ○ Juvenile idiopathic arthritis ○ Kidney stones ○ LDL cholesterol ○ Liver enzymes ○ Lung cancer ○ Malaria ○ Male pattern baldness ○ MCP-1 ○ Mean platelet volume ○ Melanoma ○ Menorcha & menopause | <ul style="list-style-type: none"> ○ Multiple sclerosis ○ Myeloproliferative neoplasms ○ Narcolepsy ○ Neuroblastoma ○ Nicotine dependence ○ Obesity ○ Osteocrosis ○ Other metabolic traits ○ Panic disorder ○ Peripheral arterial disease ○ Plasma LP (a) levels ○ Primary biliary cirrhosis ○ Prostate cancer ○ Protein levels | <ul style="list-style-type: none"> ○ Pulmonary funct. COPD ○ QT interval prolongation ○ Psoriasis ○ Recombination rate ○ Red vs. non-red hair ○ Renal function ○ Restless legs syndrome ○ Rheumatoid arthritis ○ Serum bilirubin ○ Serum IgE levels ○ Serum markers of iron status ○ Serum metabolites ○ Serum urate ○ Skin pigmentation by reflectance spectroscopy | <ul style="list-style-type: none"> ○ Soluble ICAM-1 ○ Statin-induced myopathy ○ Stroke ○ Systemic lupus erythematosus ○ Systemic lupus erythematosus in women ○ Testicular germ cell tumor ○ Type 1 diabetes ○ Type 2 diabetes ○ Venous Thromboembolism ○ Vitamin B12 levels ○ Warfarin dose ○ Weight ○ YKL-40 levels |
|--|--|---|---|---|---|--|--|

2009 2nd quarter



Biomarkery a modely



- Biomarker může být založen na **jediném analytu**, nebo na **jejich kombinaci v modelu** (klasifikátoru)
- Je to právě **kombinace více analytů** (genů, proteinů, metabolitů...), která je typická pro biomarkery z omicsových dat



Co musí biomarker (nebo model) splňovat

Musí být použitelný rutinně v praxi:

- **přesný** (dostatečně citlivý a dostatečně specifický)
- **robustní** (co nejméně omezen technologií měření)
- **reproducibilní** (obecně platný)

Jak se hledá potenciální biomarker v omics datech

Biologická otázka (hypotéza)

Dizajn experimentu

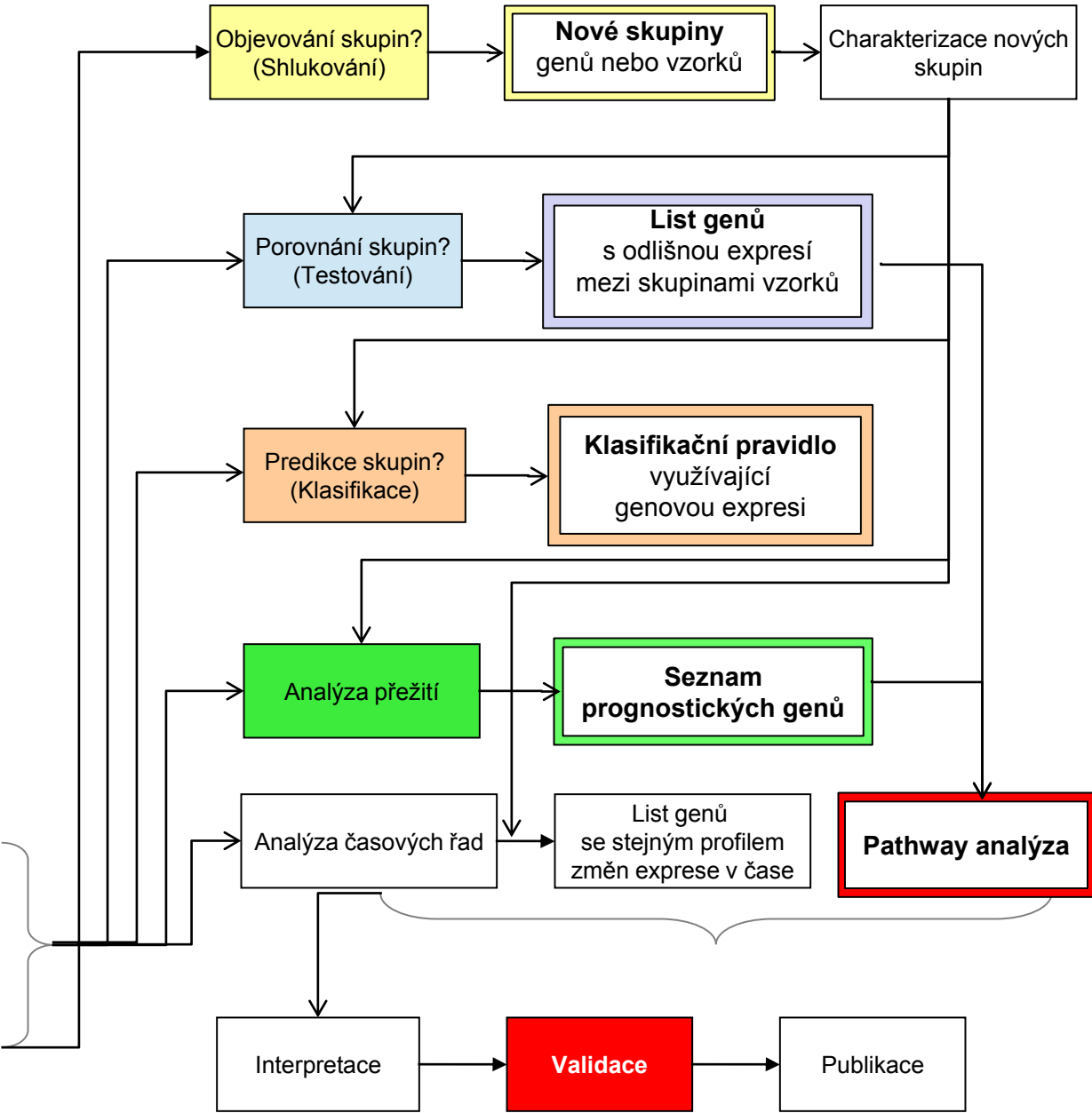
Provedení experimentu (hybridizace mikročipů, hmotnostní spektrometrie...)

N matic základních dat (jedna pro každý z N vzorků)

Kontrola kvality
Normalizace
Sumarizace

Finální datová matice N vzorků a K genů (proteinů)

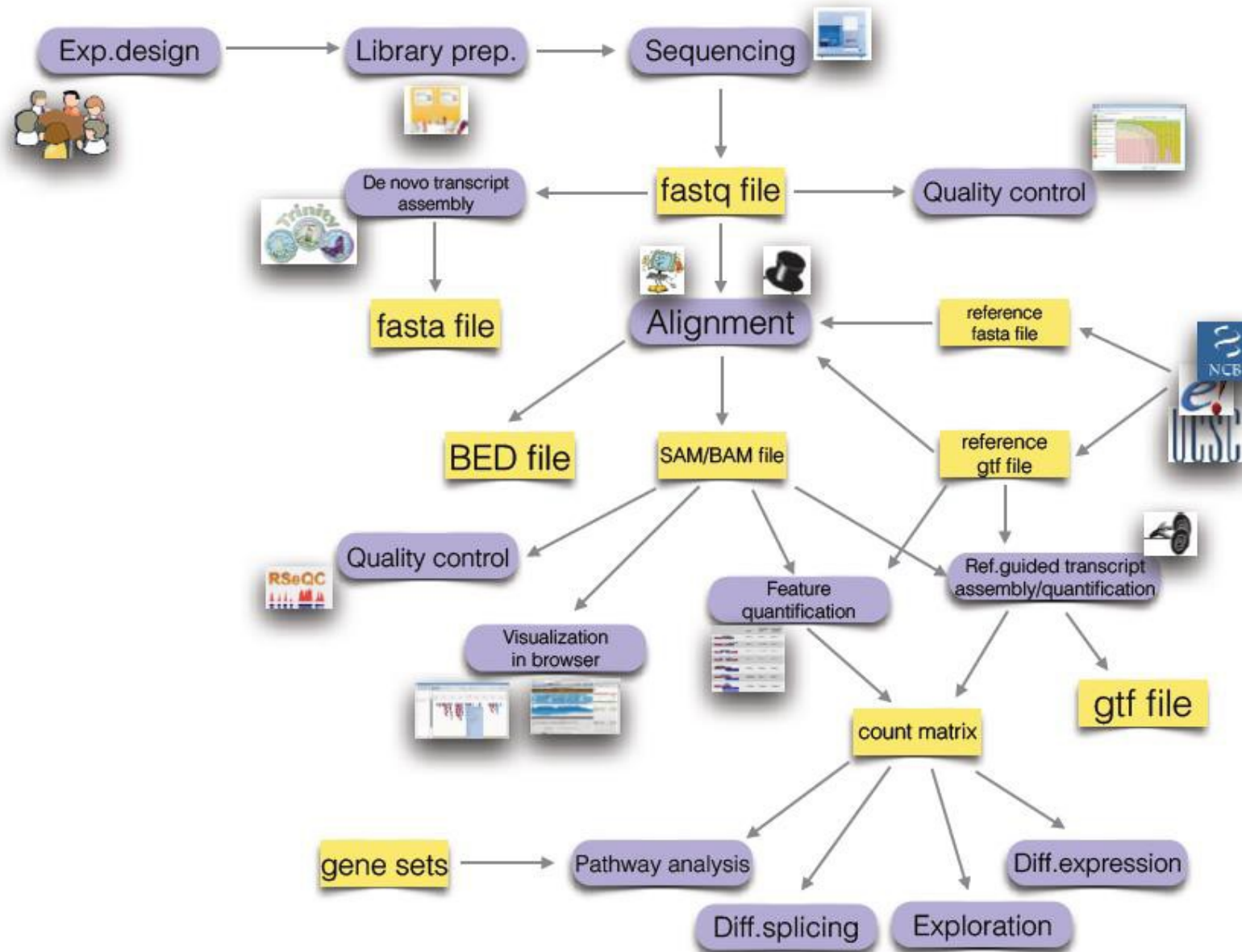
Matice informací o vzorcích N x P (např. klinická data v medicíně)





... zejména
předzpracování těchto
dat a vytváření omics
biomarkerových modelů
má svá specifika a
pravidla!

Postup bioinformatického zpracování RNAseq





Proč jsou omicsová data problematická



Obsahují **množství šumu** (technická i biologická variabilita)

Nerepresentují skutečné hodnoty (koncentrace, počty) sledovaných molekul

Pocházejí z komplexních technologií, které bývají **velice citlivé na vnější vlivy**

Jeich předzpracování je **náročné a vysoce specifické** pro daný typ platformy

Počet vzorků je mnohem **menší než** počet sledovaných **proměnných**.

Zkoumané **proměnné jsou často korelované** a mají mezi sebou komplexní vztahy (geny, proteiny...)

Specifika dat z omics
experimentů



... analýza těchto dat
vytváření omics
biomarkerových
modelů má svá
specifika!





Skandál na Duke university

Severní Karolína, USA





2006 – Anil Potti, nadějný vědec z Duke University publikuje v Nature Medicine s kolegy článek o biomarkerech rezistence na chemoterapeutika v onkologii.

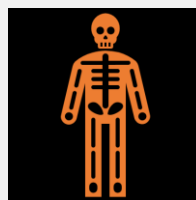
Genomické signatury byly odvozeny z analýzy exprese (mikročipy) senzitivních a rezistentních buněčných linií, výsledky validovány na pacientech.

Obrovský ohlas, v roce 2006 článek zařazen mezi “The Top 6 Genetic Stories of 2006”

Genomic signatures to guide the use of chemotherapeutics

Anil Potti^{1,2}, Holly K Dressman^{1,3}, Andrea Bild^{1,3}, Richard F Riedel^{1,2}, Gina Chan⁴, Robyn Sayer⁴,
Janiel Cragun⁴, Hope Cottrill⁴, Michael J Kelley², Rebecca Petersen⁵, David Harpole⁵, Jeffrey Marks⁵,
Andrew Berchuck^{1,6}, Geoffrey S Ginsburg^{1,2}, Phillip Febbo¹⁻³, Johnathan Lancaster⁴ &
Joseph R Nevins¹⁻³

Using *in vitro* drug sensitivity data coupled with Affymetrix microarray data, we developed gene expression signatures that predict sensitivity to individual chemotherapeutic drugs. Each signature was validated with response data from an independent set of cell



2006 – Biostatistici K. Coombes, J. Wang and K.A. Baggerly se snaží o aplikaci signatur na data výzkumníků z jejich univerzity, ovšem bez úspěchu.



Aktivně konzultují s autory článku.



Čím více se noří do dat, tím více mají pochybností o validitě závěrů a správnosti samotných dat!

2007 – Coombes a kol. publikují v Nature Medicine dopis zpochybňující Pottiho výzkum

(Coombes, Wang, Baggerly. Microarrays: retracing steps, Nature Medicine, 2007)



Reportují tyto chyby:

označení senzitivních a rezistentních buněčných linií nesedí!

tabulka se seznamem významných genů a jejich sond obsahuje systematickou chybu (posun o políčko) – geny nesedí se sondami, po korekci tabulky se podařilo reprodukovat pouze 3 ze 7 seznamů a výsledků senzitivity

Model rezistence na doxacel – podařilo se zreprodukovat pouze 31 z 50 genů publikovaných v článku, ostatních 19 bylo zřejmě přidáno ručně “aby byla validace úspěšná”

Autorský SW (algoritmus), který Potti používá, pracuje s validačními a testovacími daty společně. Po korekci této chyby jsou výsledky validace klasifikátorů špatné – na validačních datech téměř rovné náhodě.

Mezitím vycházejí další články:

Blood (2006), NEJM (2006), JCO (2007), Lancet Oncology (2007), JAMA (2008), PLOS (2008), PNAS (2008), Clin Can Res (2009)

V roce 2009 již **212 citací, několik klinických studií, stovky léčených pacientů**

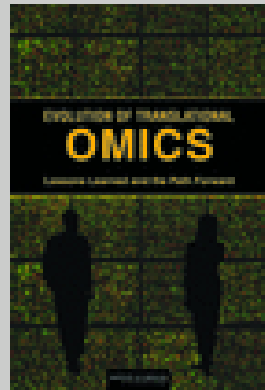
Co bylo dál?

Červenec 2010 – ředitel National Cancer Institute (NCI) Harold Varmus obdržel **dopis** od více než **30 statistiků** a bioinformatiků, ve kterém vyjádřili své obavy nad použitím několika testů založených na genové expresi, které se používali v již probíhajících klinických studiích na Duke University k predikci odpovědi na chemoterapii.

Anil Potti v roce 2010 obviněn z falzifikace výsledků a vyšetřován. Trvá 4 roky a mnoho úsilí, než jsou chyby uznány a články staženy!

V důsledku vznikla komise Institutu medicíny (IOM), cílem které bylo sepsání doporučení pro vývoj testů z omicsových studií

This PDF is available from The National Academies Press at http://www.nap.edu/catalog.php?record_id=13297



Evolution of Translational Omics: Lessons Learned and the Path Forward

ISBN
978-0-309-22418-5

300 pages
6 x 9
PAPERBACK (2012)

Christine M. Micheel, Sharly J. Nass, and Gilbert S. Omenn, Editors;
Committee on the Review of Omics-Based Tests for Predicting Patient
Outcomes in Clinical Trials; Board on Health Care Services; Board on
Health Sciences Policy; Institute of Medicine

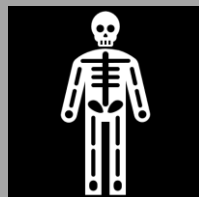
IOM (Institute of Medicine). 2012. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington, DC: The National Academies Press.

MUNI | RECETOX

IOM komise: Specifika testů založených na omics



Testy na bázi omics a ve skutečnosti všechny klinické laboratorní testy podléhají odlišnému regulačnímu rámci než léky



Absence jasného biologického zdůvodnění na rozdíl od většiny ostatních klinických laboratorních testů založených na jediném analytu



Složitost omicsového výzkumu ztěžuje sdílení komplexních datových souborů a výpočetních modelů, což omezuje schopnost ostatních vědců replikovat a ověřovat zjištění a závěry těchto studií

Absence
jasného
biologického
odůvodnění
testů omics
biomarkerů

Biologické zdůvodnění **testu s jedním analytem** je často zcela zřejmé: Test je užitečný, protože gen, RNA, protein nebo metabolit hraje pochopitelnou roli v patologii onemocnění nebo jiném vyšetřovaném biologickém procesu.

Příklady:

Testování karcinomu prsu lidským epidermálním růstovým faktorem 2 (HER2)

Měření hladiny cholesterolu lipoproteinů s nízkou hustotou (LDL) pro hodnocení srdečního rizika

Absence
jasného
biologického
odůvodnění
testů omics
biomarkerů –
proč je to
problém

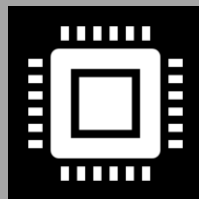
Když se nedá test založený na omicsových biomarkerech biologicky odůvodnit, je o to důležitější ho správně VYTVOŘIT a poté správně VALIDOVAT, aby byla zajištěna vědecká spolehlivost!

Z důvodů vyššího rizika „přetrénování“ těchto testů je potřeba přísných kritérií, validace a odpovědnosti ještě vyšší než u samostatných testů založených na biomarkerech.

Problém (ne) sdílení komplexních datových souborů a výpočetních modelů



K dispozici jsou databázové úložiště pro soubory omicových dat, ale sdílení dat není rutinní a bez přístupu k datům a přesně definovanému výpočetnímu modelu je replikace a ověření obtížnější než pro biomarkery založené na jednotlivých analytech.

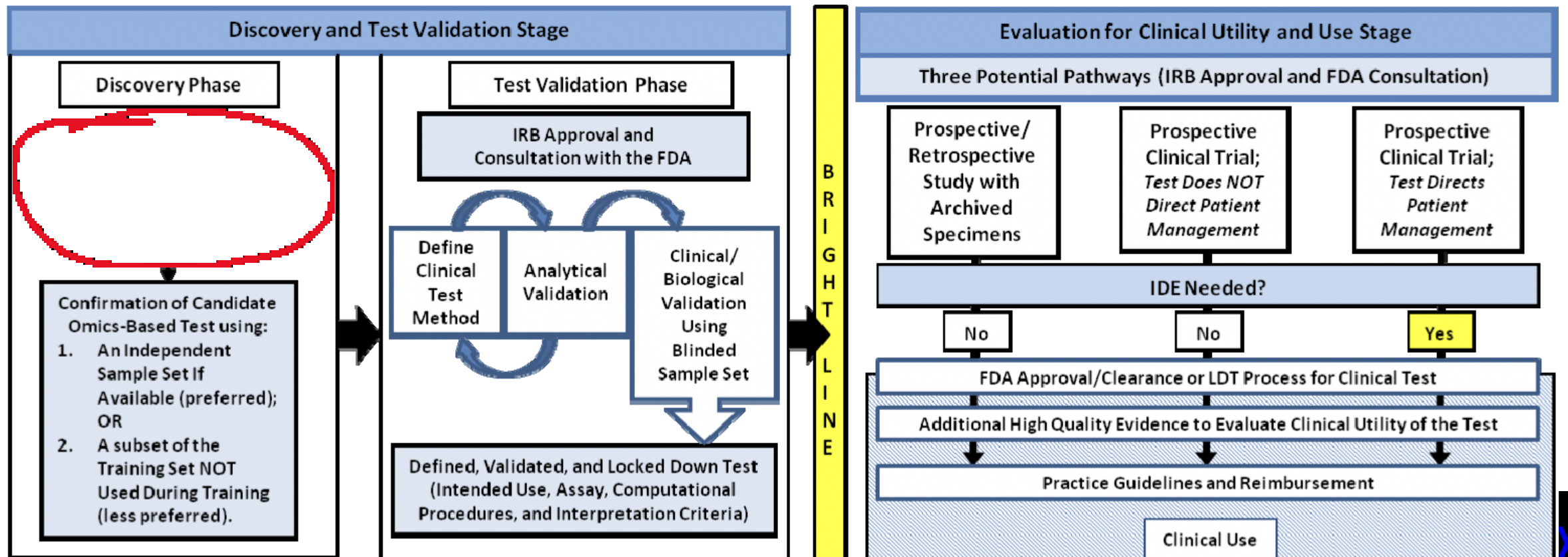


I když nezávislé validační studie jsou drahé, potřeba replikace v omicových studiích je nutná vzhledem ke složitosti dat, které mohou vést k chybám (od jednoduchých chyb správy dat až po nesprávně navržené výpočetní modely).



Tato úroveň složitosti neexistuje pro výzkum, vývoj a validaci testů s jedním biomarkerem.

Doporučení IOM komise pro vývoj testů založených na omicsových datech





ANIL POTTI

CASE PROGRESSION

**JULY
2010**

Potti is accused of falsifying information on his resume, and Duke launches an investigation into his work

**NOV
2010**

Potti resigns

**OCT
2011**

Patients in Potti's clinical trials file a lawsuit against the University

**JAN 29 THUR
2015**

**LAWSUIT SET
TO START**

Více info: <https://ori.hhs.gov/content/case-summary-potti-anil>

MUNI | RECETOX



Published Mar. 26, 2019

RALEIGH, N.C. — Duke University will pay \$112 million to settle a whistleblower lawsuit after federal prosecutors said a research technician's fake data landed millions of dollars in federal grants, the school and the government said Monday.

**JULY
2010**

Potti is accused of falsifying information on his resume, and Duke launches an investigation into his work

**NOV
2010**

Potti resigns

**OCT
2011**

Patients in Potti's clinical trials file a lawsuit against the University

**JAN 29 THUR
2015**

**LAWSUIT SET
TO START**

Více info: <https://ori.hhs.gov/content/case-summary-potti-anil>

MUNI | RECETOX



Published Mar. 26, 2019

RALEIGH, N.C. — Duke University will pay \$112 million to settle a whistleblower lawsuit after federal prosecutors said a research technician's fake data landed millions of dollars in federal grants, the school and the government said Monday.

**JULY
2010**

Potti is accused of falsifying information on his resume, and Duke launches an investigation into his work

**NOV
2010**

Potti resigns

**OCT
2011**

Patients in Potti's clinical trials file a lawsuit against the University

**JAN 29 THUR
2015**

11 Sept 2015 – Potti uznán vinným z podvádění ve výzkumu

Zatím 14 článků ztaženo

Více info: <https://ori.hhs.gov/content/case-summary-potti-anil>

MUNI | RECETOX

Více info o kauze

[Retraction Watch Database \(retractiondatabase.org\)](https://retractiondatabase.org)

Které články A. Potti byly ztaženy z tisku?

[Anil Potti - Wikipedia](#)

[Detailed Case Histories - Fostering Integrity in Research - NCBI Bookshelf \(nih.gov\)](#)

[41-01 Duke Officials Silenced Med Student Who Reported Trouble in Anil Potti's Lab - The Cancer Letter](#)

[Federal Register :: Findings of Research Misconduct](#)

Jak se
hledá
potenciální
biomarker v
omics
datech

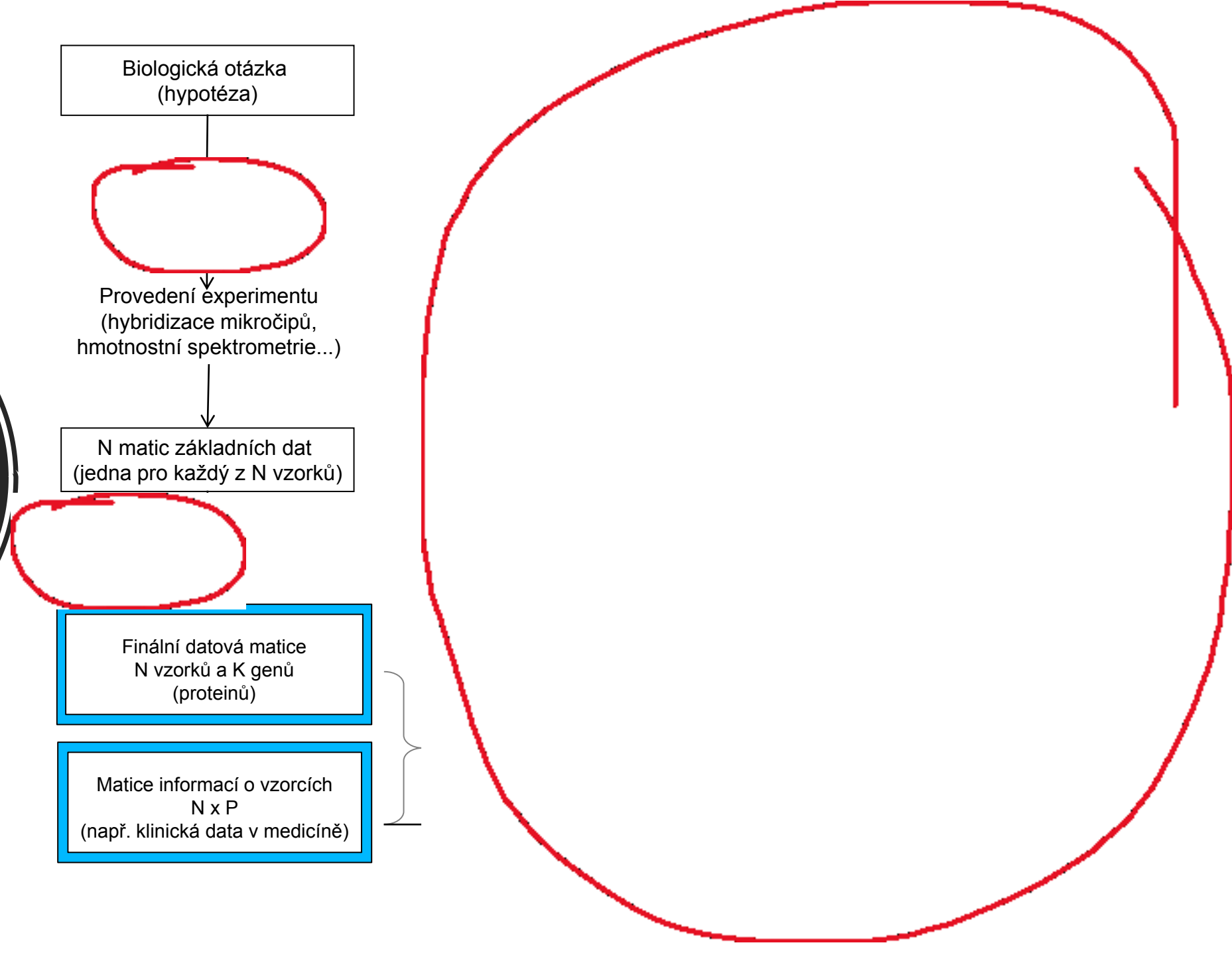
Biologická otázka
(hypotéza)

Provedení experimentu
(hybridizace mikročipů,
hmotnostní spektrometrie...)

N matic základních dat
(jedna pro každý z N vzorků)

Finální datová matice
N vzorků a K genů
(proteinů)

Matice informací o vzorcích
N x P
(např. klinická data v medicíně)



Úprava
omicsových
dat do
podoby, kdy
je možná
derivace
biomarkerů
trvá
podstatně
déle než u
jiných dat

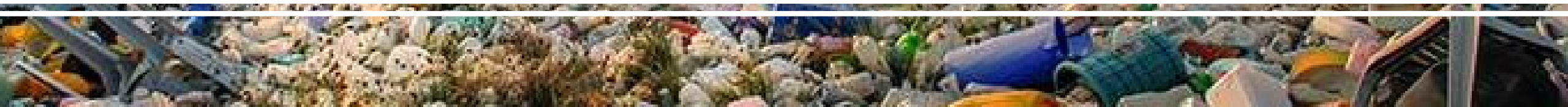
Data obsahují velké množství technického i biologického šumu, který je nutné odstranit

Protože jedno spuštění přístroje obvykle není schopno analyzovat všechny vzorky, vytváří se nežádoucí matoucí efekty (efekty dávky), které je nutno odstranit

Technologie jsou velice nové (a vznikají stále!) a algoritmy pro optimální zpracování jejich dat se vytvářejí a testují i 5-10 let - neexistují zlaté standardy a mnohé implementace jsou plné chyb

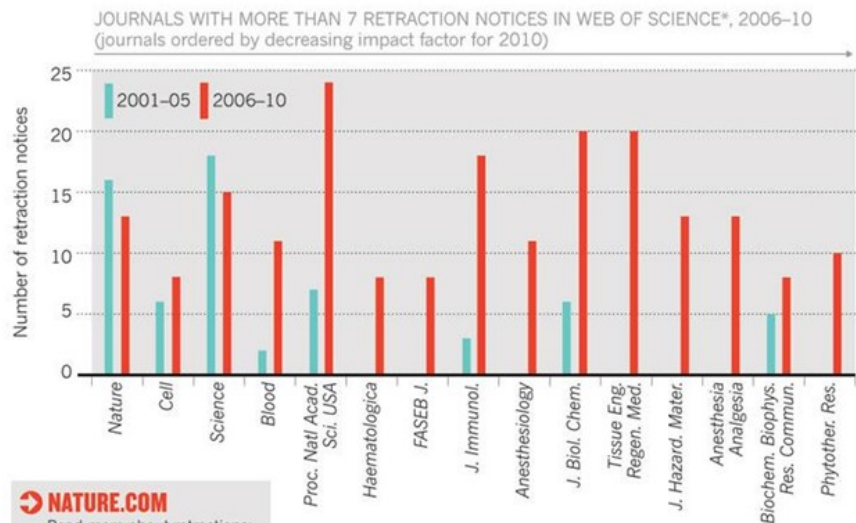
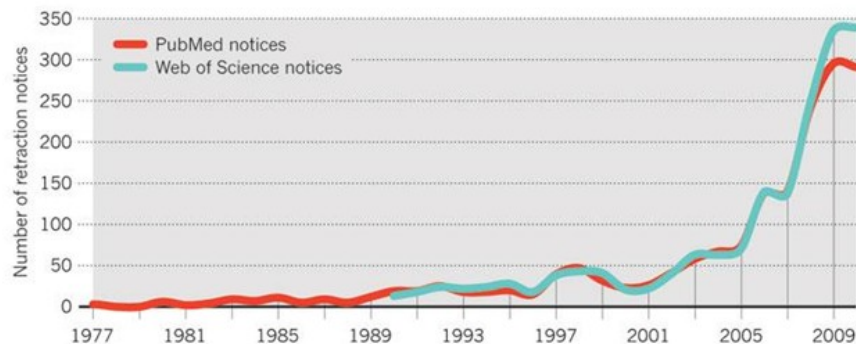


Vědecké časopisy jsou (přesto) plné odpadu



RISE OF THE RETRACTIONS

In the past decade, the number of retraction notices has shot up 10-fold (top), even as the literature has expanded by only 44%. It is likely that only about half of all retractions are for researcher misconduct (middle). Higher-impact journals have logged more retraction notices over the past decade, but much of the increase during 2006–10 came from lower-impact journals (bottom).



NATURE.COM
Read more about retractions:
go.nature.com/2uweek

*Not shown: *Acta Crystallographica E* saw 81 retractions during 2006–10.

Podíl článků stažených z tisku se zvyšuje

Za analyzovanou dekádu vzrostl počet článků pouze o 44%, počet retrakcí článků se zvýšil desetinásobně!

Pouze **0.02%** článků je staženo z tisku!

Van Noorden, R. (2011) Science publishing: The trouble with retractions. *Nature*. 2011 Oct 5;478(7367):26-8.

Důvody stažení publikací

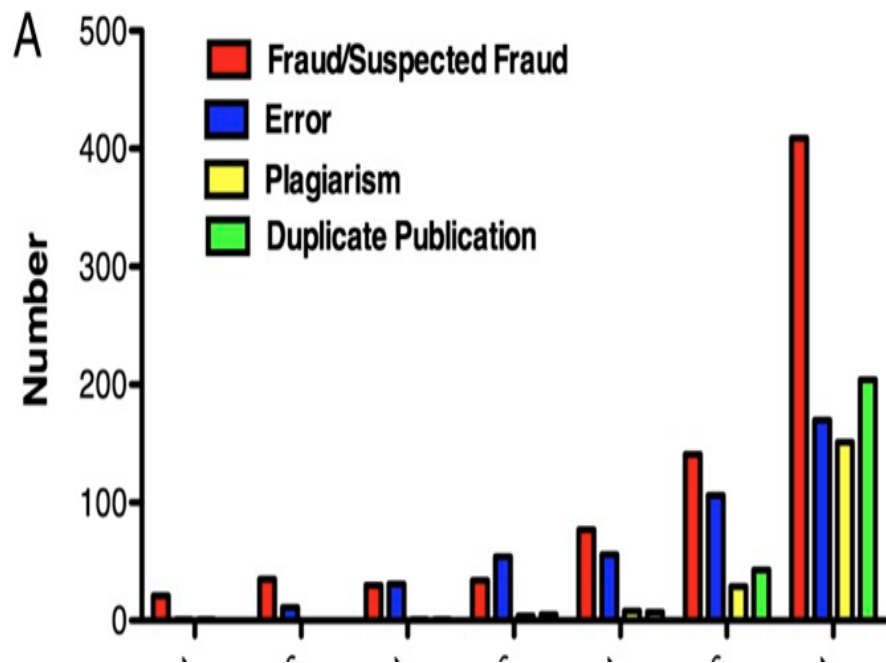
Analýza **2,047** stažených biomedicínských a přírodovědných vědeckých článků

Nejčastější důvod: **podvod** (nebo podezření z podvodu): **43.4%**

21.3% článků bylo staženo kvůli **čestné chybě (honest error)**

Až **31.8%** článků, které byly staženy zůstali neoznačeny

Většina autorů své stažené články stále cituje...



Adapted Figure 1 from Fang et al. (2012) Misconduct accounts for the majority of retracted scientific publications. PNAS 2012 Oct 16; 109(42):17028-1703

Retractions in the medical literature: how many patients are put at risk by flawed research?

R Grant Steen

Bylo analyzováno **180 primárních a 851 odvozených klinických studií**, které byly provedeny na základě výzkumu ze stažených publikací.

U 180 primárních studií bylo léčeno **9189** pacientů (z více než 28 tisíc)

U 851 odvozených studií bylo léčeno **70 501** pacientů (z více než 400 tisíc)

Studie, které byly staženy pro **podvod**, léčily statisticky významně více pacientů, než studie, které byly staženy pro **chybu**.

Biomarkery z omicsových dat

Jsou často komplexní:

- **Složené z více charakteristik** (více genů, proteinů...)
- **Bez jasně definovaného biologického zdůvodnění**

Pocházejí z dat:

- zatížených **významným technickým šumem** z různých zdrojů
- analyzovaných **metodami, které nejsou standardizované**
- které jsou pouze **korelované** s měřenou proměnnou (např. nejsou koncentrace ani počty molekul)
- které jsou **komplexní a obtížně se sdílejí**

Bez bioinformatika není biomarker!

- Svět biomedicínského výzkumu se podstatně změnil
- Hledání molekulárních biomarkerů se provádí ve vysoce specializovaných experimentech které produkují velká data
- nikdo se neobejde bez bioinformatika!