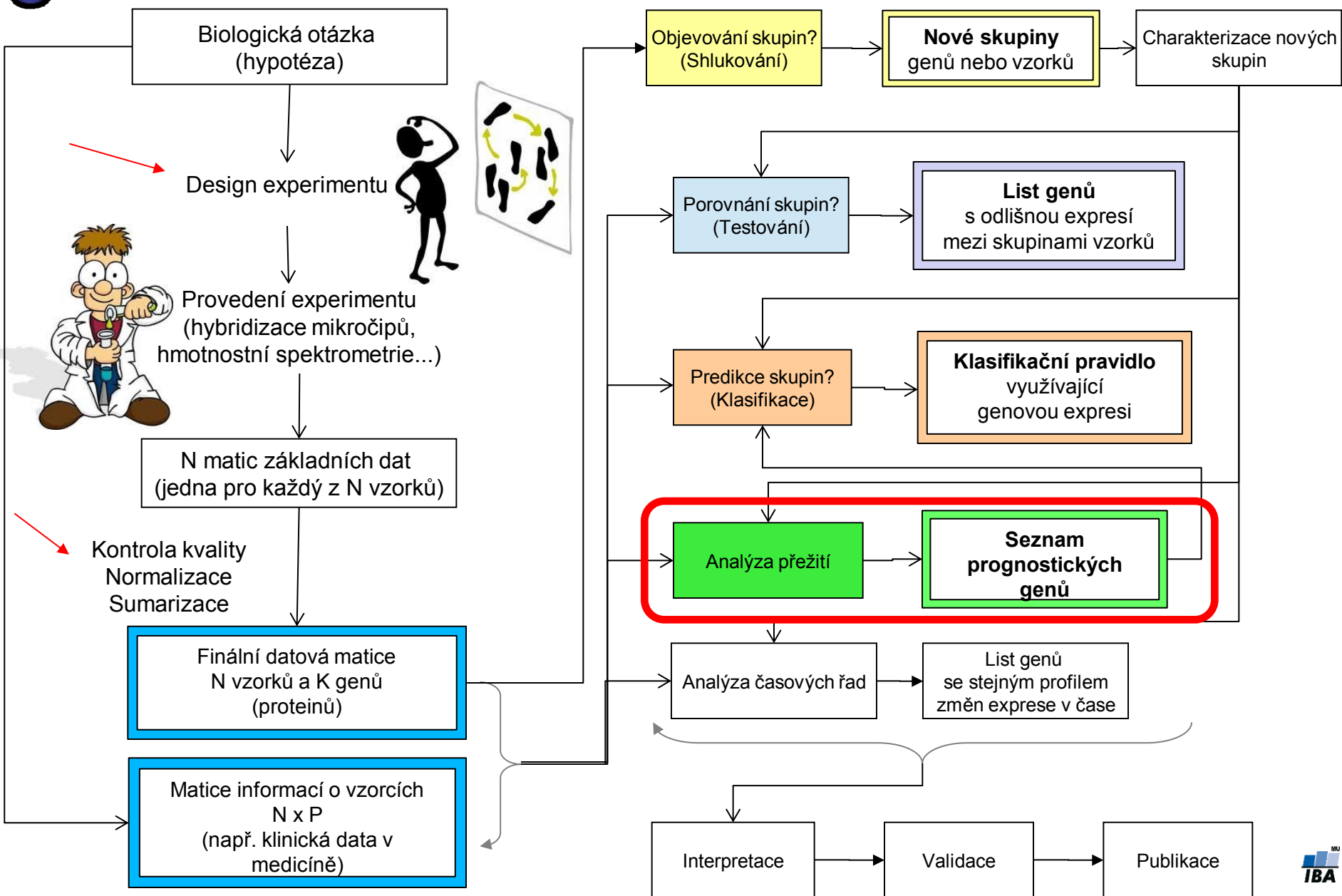




Analýza přežití



Společná schéma analýzy dat



Analýza přežití a genomická data

- Dva základní cíle s přesně definovanými otázkami:
 1. **Predikce rizika události u molekulárních podtypů** – Mají skupiny pacientů definované pomocí genové exprese rozdílné přežití?
 2. **Predikce času přežití na základě genové exprese** – Má exprese genu vliv na přežití?
- **Přežití** je čas do nějaké námi sledované události
 - Úmrtí (overall survival)
 - Relaps (relaps-free survival)
 - Návrat onemocnění (disease-free survival)
 - ...

Data analýzy přežití

- Sbírání dat od zadaného času (začátek studie, diagnóza)
- Dvě proměnné:
 - Výsledek – Nastala událost?
1 = událost nastala
0 = událost nenastala
 - Čas přežití
- Časy přežití s výsledkem 0 představují tzv. **cenzorované hodnoty** – je to čas do konce pozorování, a nebo posledního záznamu

Pacient	Výsledek	Čas přežití (měsíce)
1	1	4
2	1	11
3	0	56*
4	1	8
5	0	44*
6	0	48*
7	0	57*
8	1	3

Metody analýzy přežití

- V závislosti od otázky sa používajú dve základní funkce:
 - **Kaplan – Meierův odhad přežití** – Mají skupiny definované pomocí genové exprese rozdílné přežití?
 - **Coxův model proporcionálních rizik** – Má exprese genu vliv na přežití?

Kaplan-Meierův odhad prežití

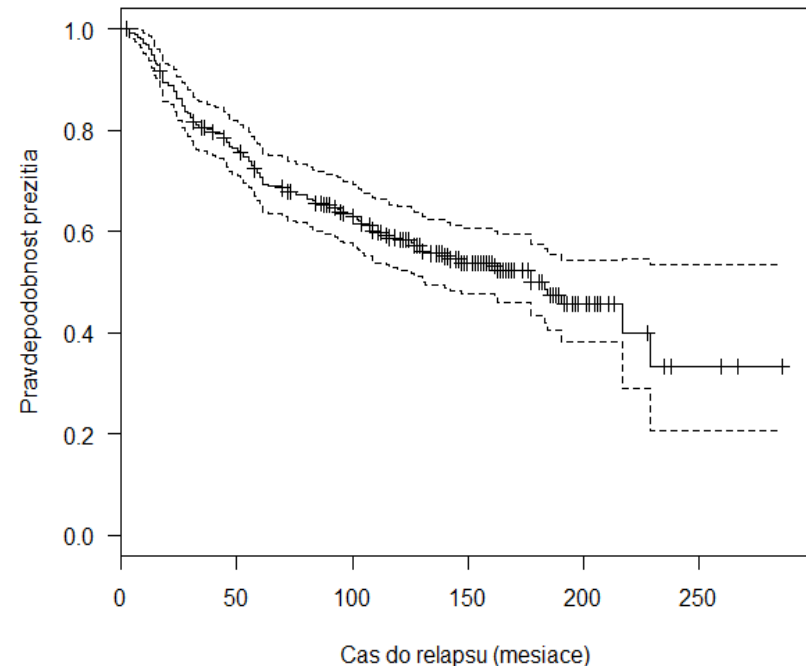
- Definovaný jako pravděpodobnost prežití do času t
- Pro každý časový interval t odhadne podíl přežívajících, za použití pravděpodobnosti
- Jedná se o neparametrický odhad

$$S(t) = p_1 * p_2 * p_3 * \dots * p_t$$

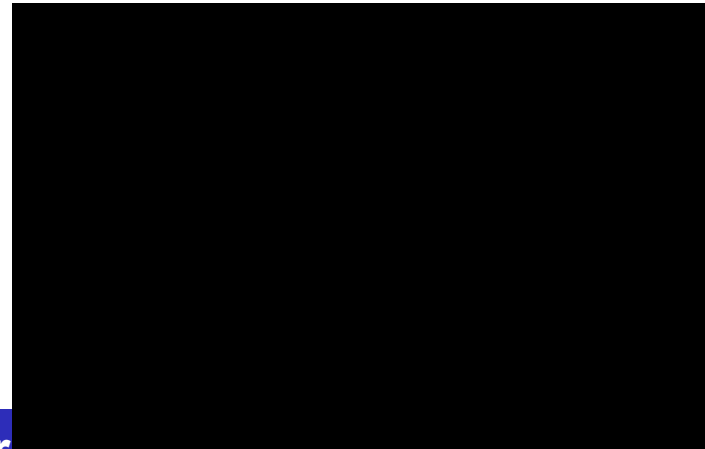
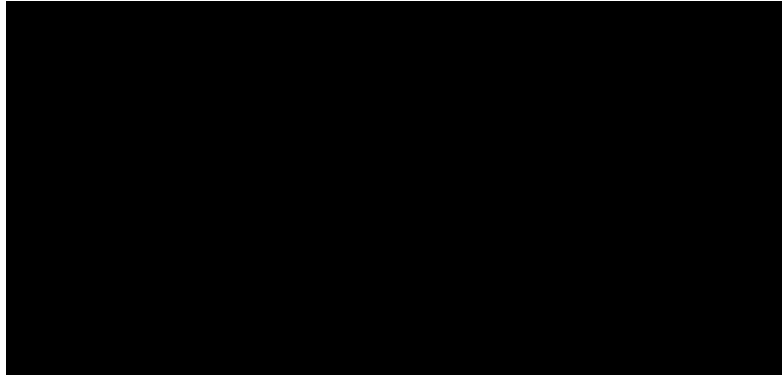
$$p_i = \frac{r_i - d_i}{r_i}$$

r_i – počet živých pacientů na začátku časového intervalu t_i

d_i – počet úmrtí za časový interval t_i



Kaplan-Meierův odhad přežití - příklad



t(i)	N	# úmrtí	# censor.	RIZIKO
0	21	0	0	1
6	21	3	1	$1 \cdot (18/21) = \mathbf{0.8571}$
7	17	1	1	$\mathbf{0.8571} \cdot (16/17) = 0.8067$
10	15	1	2	$0.8067 \cdot (14/15) = 0.7529$
13	12	1	0	$0.7529 \cdot (11/12) = 0.6902$
16	11	1	3	$0.6902 \cdot (10/11) = 0.6275$
22	7	1	0	$0.6275 \cdot (6/7) = 0.5378$
23	6	1	5	$0.5378 \cdot (5/6) = 0.4482$

Porovnání křivek přežití

- Dva testy pro zjištění párových rozdílů v analýze přežití:
 - **Gehanův-Breslowův-Wilcoxonův test**
 - Přiřazuje větší váhy úmrtím v dřívějších časových bodech
 - Může být zavádějící, pokud je velké procento pacientů cenzorováno v dřívějších časových bodech
 - **Mantelův-Haenszův log-rank test**
 - Standardně používaný
 - Předpokládá nezávislost cenzorování a výskytu jednotlivých událostí
 - Přiřazuje stejné váhy úmrtím ve všech časových bodech
 - Silný, pokud je předpoklad proporcionality rizik splněný

Log-rank test

- Umožňuje statistické zhodnocení rozdílu v přežití, ale neposkytuje kvantifikaci tohoto rozdílu a nebere v úvahu vliv dalších proměnných
- j – čas, Z – testová statistika, hypotéza je zamítnutá $Z > z_\alpha$

$$Z = \frac{\sum_{j=1}^J (O_{1j} - E_{1j})}{\sqrt{\sum_{j=1}^J V_j}}$$

$$V_j = \frac{O_j \left(\frac{N_{1j}}{N_j}\right) \left(1 - \frac{N_{1j}}{N_j}\right) (N_j - O_j)}{N_j - 1}$$

$$O_j = O_{1j} + O_{2j}$$

$$N_j = N_{1j} + N_{2j}$$

$$E_{1j} = \frac{O_j}{N_j} N_{1j}$$

O_j – počet pozorovaných událostí v čase j
 N_j – počet subjektů v riziku

Coxův model proporcionálních rizik

- Měří okamžité riziko události
- Můžeme testovat více proměnných (nejen genovou expresi, ale i další, jako je věk, ...)

$$h(t, \mathbf{x}) = h_0(t) * e^{(b_1 x_1 + \dots + b_p x_p)}$$

$h(t)$ je riziko v čase t

$h_0(t)$ je základní riziková funkce společná pro všechny subjekty ve studii

$e^{(b_1 x_1 + \dots + b_p x_p)}$ vyjadřuje relativní riziko daného subjektu vzhledem k subjektu se základním rizikem

$x_1, x_2, x_3, \dots, x_p$ jsou vysvětlující proměnné

$b_1, b_2, b_3, \dots, b_p$ jsou odhadnuté regresní koeficienty

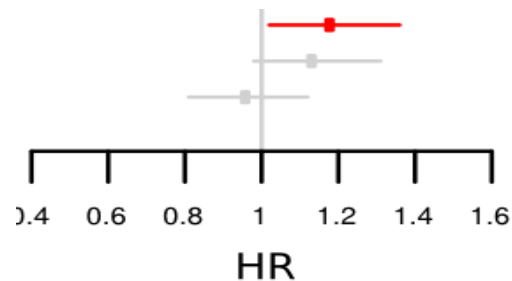
Coxův model proporcionálních rizik II.

- Předpoklad: proporcionalita rizika
- Interpretace parametrů: změna o jednotku znamená změnu rizika $h(t)$ o hodnotu odhadnutého koeficientu
- Změna genové exprese o jednotku se nedá přímo porovnat mezi experimenty (čísla jsou relativní, mají jinou škálu) => škálování expresních hodnot genu \mathbf{x}_g před analýzou:

$$y_g = \frac{\mathbf{x}_g - \text{median}(\mathbf{x}_g)}{\text{IQR}(\mathbf{x}_g)}$$

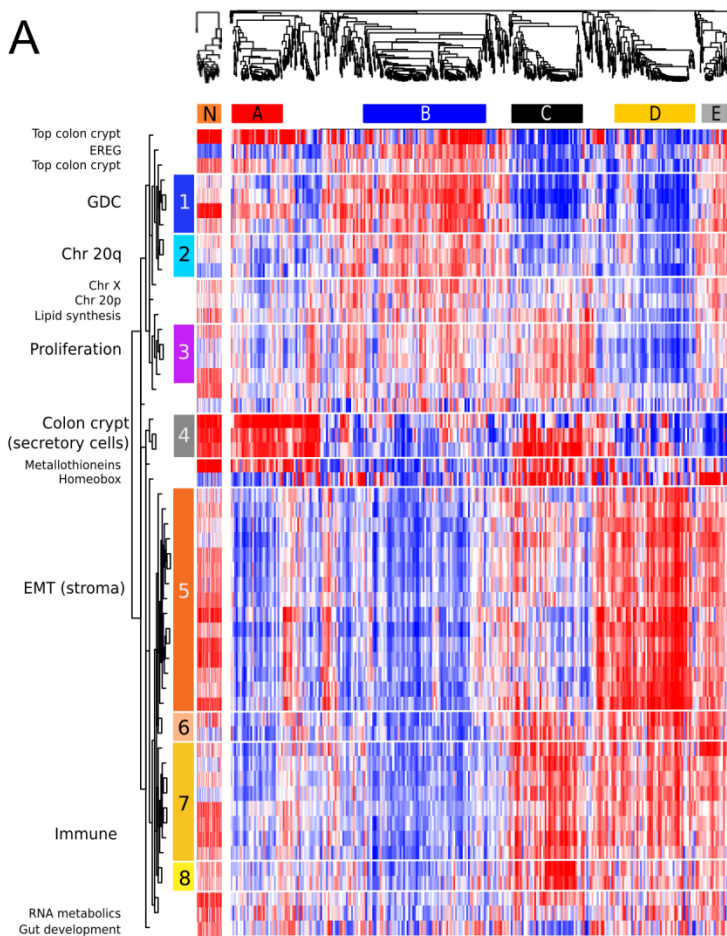
Coxův model proporcionálních rizik III

- Předpoklad: proporcionalita rizik:
 - Podíl rizikových funkcí libovolných dvou jedinců je proporcionální, nezávislý na čase (třeba otestovat)
- Jinak použít parametrické metody
- Zobrazení (tzv. forestplot)

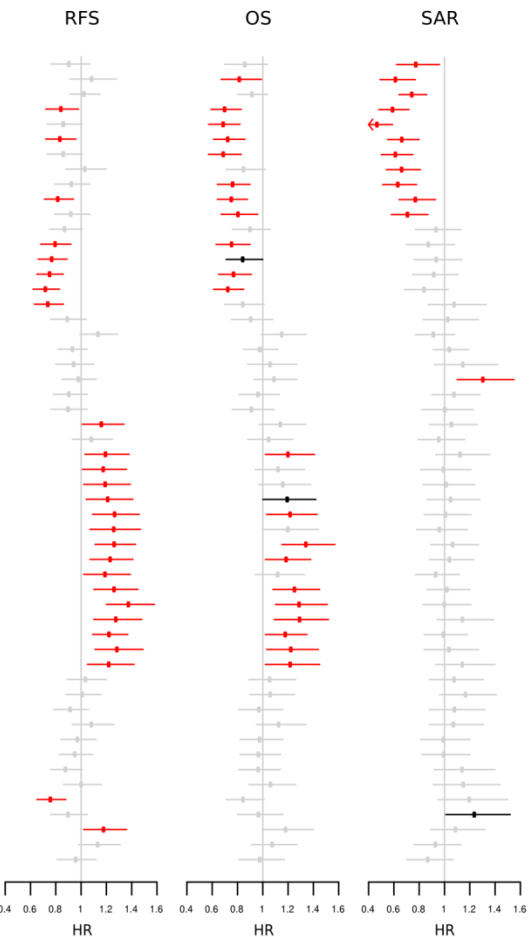


● HR
_____ 95% IS

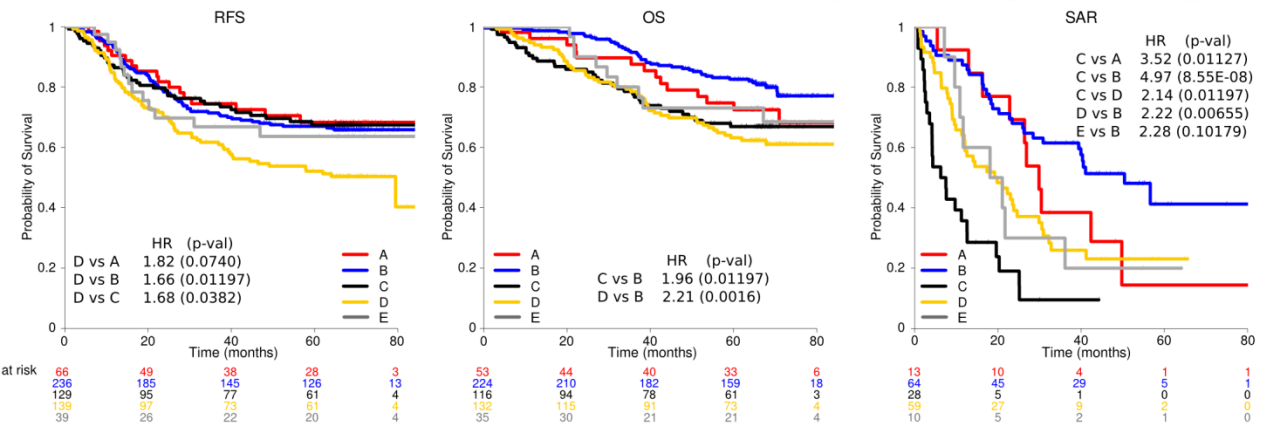
A



B



C



*Zavolat statistika po vykonání experimentu
je asi jako požádat doktora o posmrtné
vyšetření:
pravděpodobně bude schopný říct, na co
experiment zemřel.*

Ronald Fisher