

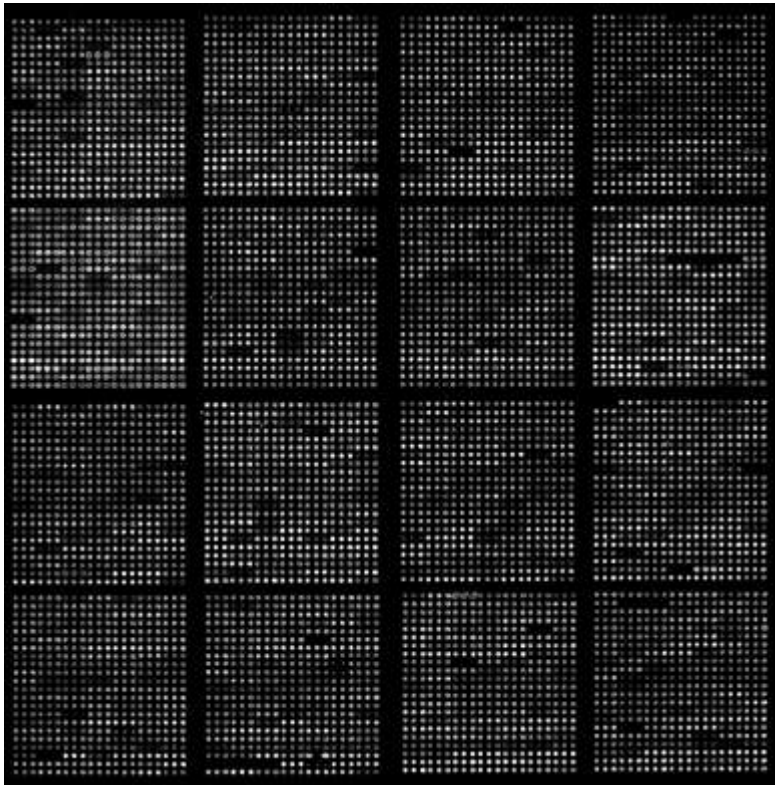
Principy microarrays

Pavla Gajdušková

Analytická cytometrie, 18. a 25. listopadu 2014

Microarrays

Kolekce DNA sond přichycených k pevnému podkladu



„Tištěná“ microarrays



Fotolitografie

Microarray technologie

- I. Výběr sond (probes):** cDNA vektory, BAC vektory, krátké nebo dlouhé oligonukleotidy, proteiny, tkáně
- II. Příprava microarray:** nanesení sond na sklo nebo membránu
- III. Design experimentu:** zvolení správné metody, použití referenčního vzorku, záměna fluorescenčních barev
- IV. Fluorescenční značení vzorků**
- V. Analýza microarray obrazů:** nalezení sond v obraze, korekce pozadí, výpočet intenzity v jednotlivých bodech
- IV. Analýza dat:** filtrování, normalizace, porovnání výsledků získaných z více microarray experimentů – klastrovací analýza

Oblasti použití microarrays v biologii

| Typ array | Sondy na microarray | Co se fluorescenčně značí a hybridizuje | ... analýza čeho |
|-----------------|------------------------------------|---|--|
| Expresní | DNA (cDNA, oligonucleotidy) | cDNA / mRNA | měření množství mRNA v bunkách, nádorech ... |
| miRNA | oligonukleotidy | miRNA | měření množství miRNA |
| CGH | DNA (BAC vektory, oligonukleotidy) | DNA | změny v genomu (zisk, ztráta chromozomů nebo jejich částí) |
| SNP | DNA (oligonukleotidy) | DNA | detekce „Single Nucleotid Polymorphisms“; změny v genomu |
| Metylace | DNA (CpG islands) | DNA (ovlivněná bisulfidem sodným) | míra metylace promotorových oblastí |
| Promoter | DNA (promotorové oblasti ~ 1kb) | DNA (ChiP obohacená) | místa vazby transkripčních faktorů, modifikace histonů |
| Tilling | DNA | všechno dříve zmíněné | všechno dříve zmíněné, sekvenování, anotace genů |
| Protein | protilátky | protein | exprese proteinů (ELISA) |

Oblasti použití microarrays v biologii

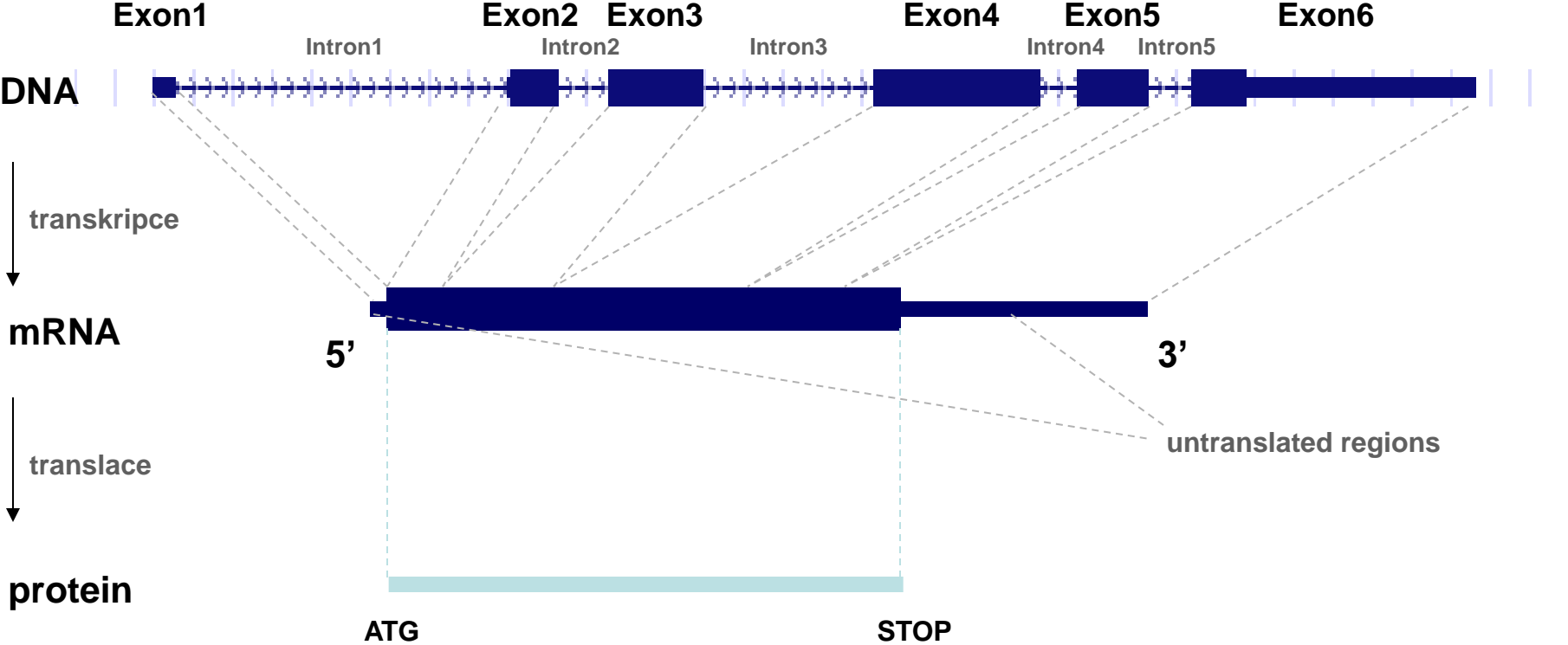


Oblasti použití microarrays v biologii

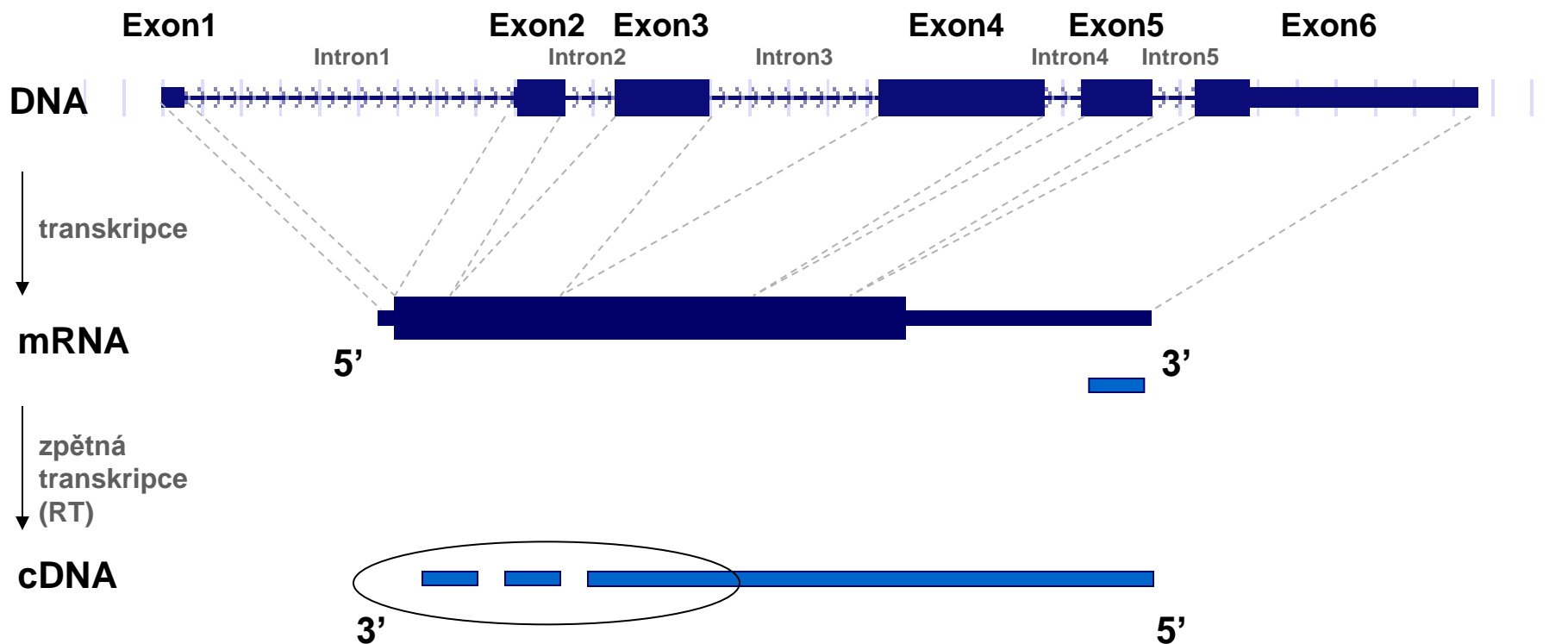


Schena M., Shalon D., Davis R. W., Brown P. O.
Quantitative monitoring of gene expression patterns with a
complementary DNA microarray. *Science* 270: 467-70, 1995.

Genová exprese




Genová exprese



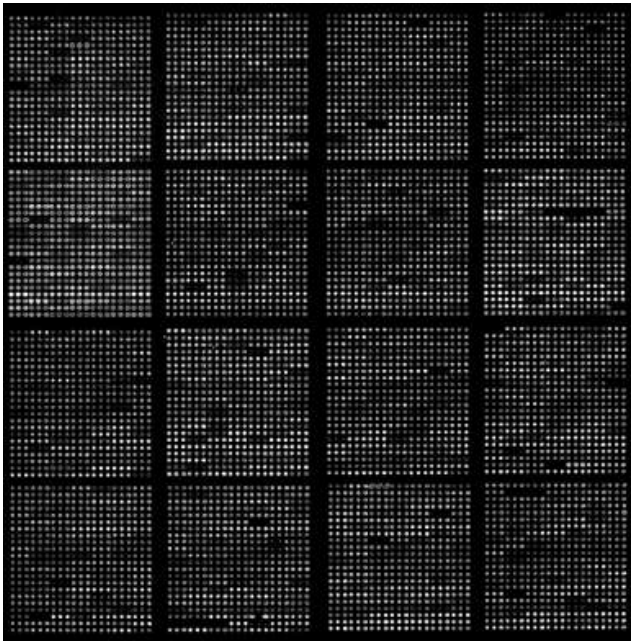
**cDNA: jednořetězcová DNA (v dalším kroku je možné syntetizovat druhý řetězec)
u genů s dlouhou mRNA nemusí vznikat vždy celá cDNA**

Metody měření množství mRNA

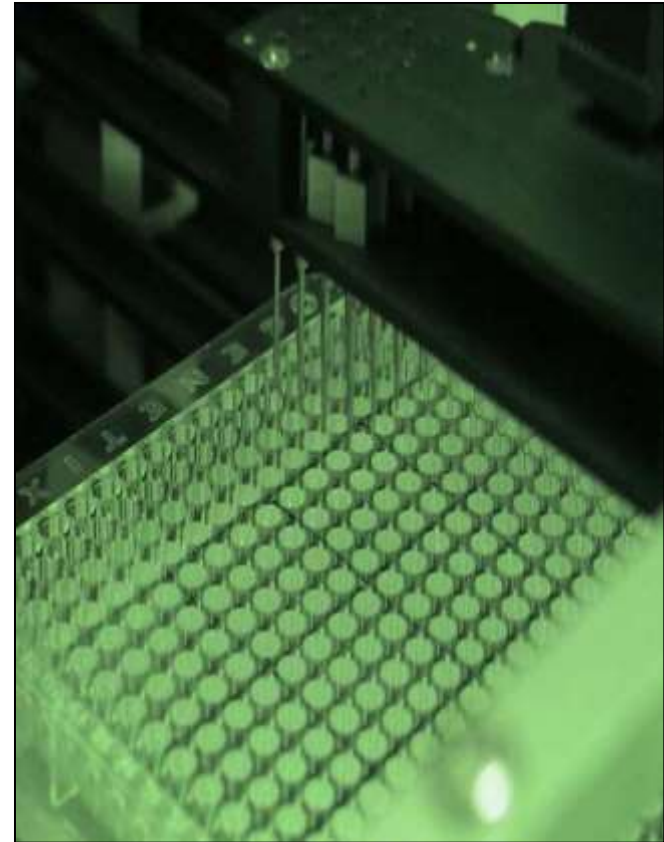
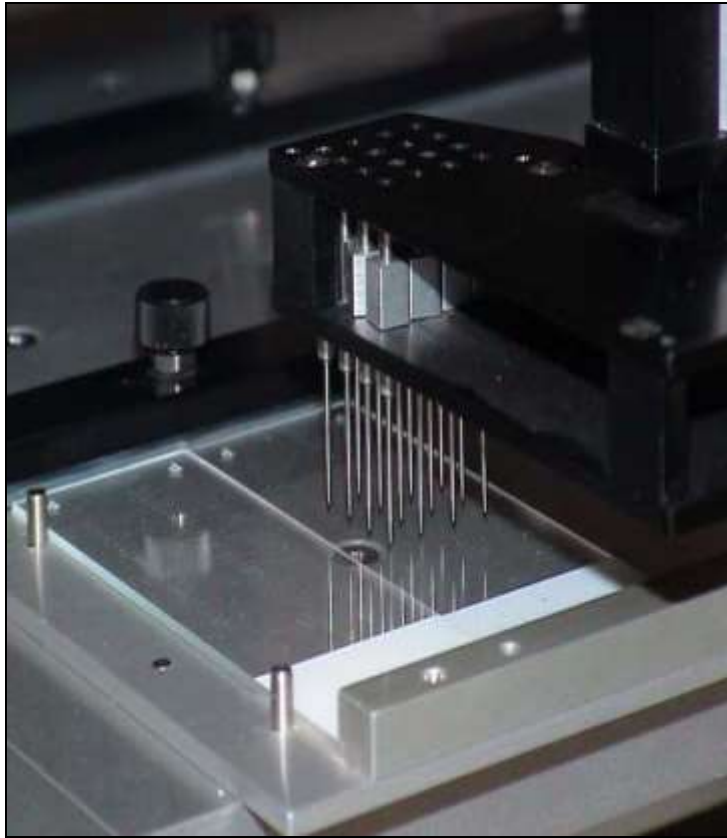
| <u>Method</u> | <u>Typical Throughput</u> | <u>Comments</u> |
|------------------------------|---|--|
| •Northern blot | 1 gene | •Standard procedure; “Gold standard”, Low throughput |
| •Subtractive cloning |  Increasing throughput | •Mid-1980's, Not comprehensive, High FP |
| •Differential display | | •1992, Follow up cloning required; Potential to identify rare mRNAs, High FP (Liang & Pardee, <i>Science</i> 257: 967-71, 1992) |
| •RT-PCR and Real-time RT-PCR | | •Sequence I.D. & semi-quantification, FP? |
| •2D protein gel/Mass Spec | | •2001, Sequence I.D. & quantification, FP? (Han et al., <i>Nature Biotech</i> 19: 946-951, 2001) |
| •ICAT/Tandem Mass Spec | | •1995, Prior sequence knowledge not mandatory, Moderate FP - depends on level of survey (Lee et al., <i>PNAS</i> 92: 8303-7, 1995; Velculescu et al. <i>Science</i> 270: 484-7, 1995) |
| •EST/SAGE | | •1995, Identification of differentially expressed genes dependent on arrayed elements, Low FP (Schemm et al. <i>Science</i> 270: 467-70, 1995) |
| •High density arrays | | 20,000-40,000 genes |

Měření množství mRNA

(microarrays tištěné pomocí skleněných kapilár)



Microarrays tištěná pomocí skleněných kapilár



Typ sond

I. dlouhé oligonukleotidy:

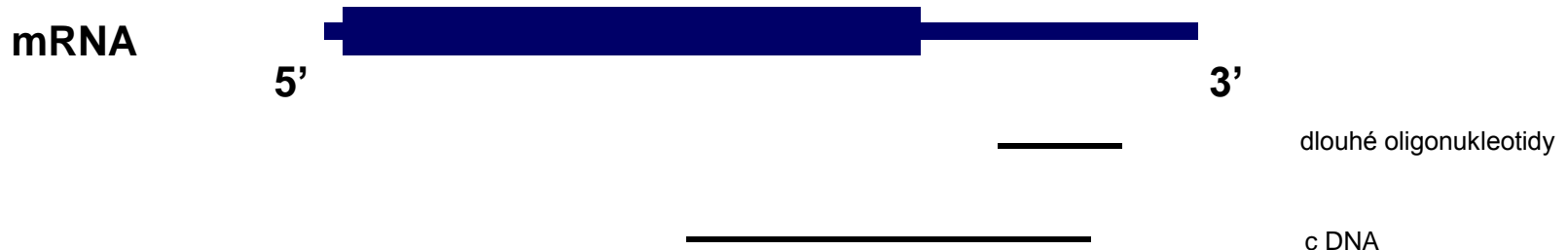
~ 60 - 70mers

komerčně dostupné (Operon, Agilent)

II. cDNA:

knihovny cDNA vektorů (IMAGE, MGC)

dostatečné množství DNA se vyprodukuje pomocí PCR
(univerzální primery pro daný typ vektorů)

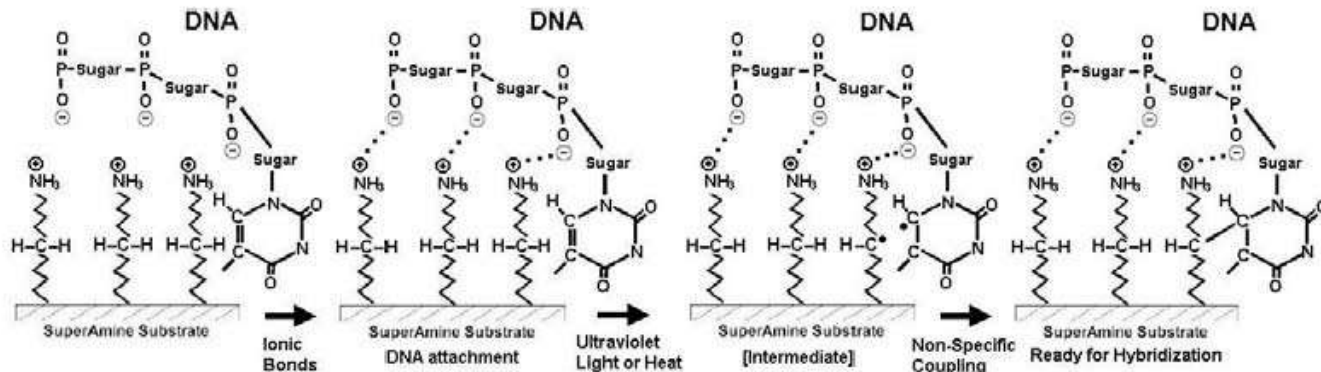
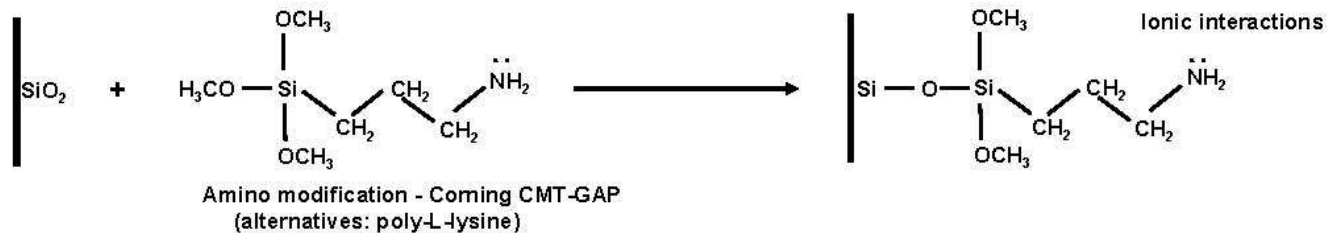


Úprava povrchu sklíček pro tisk arrays I

povrchová úprava skla: amino modifikace, poly-L-lysine

modifikace povrchu → natisknutí DNA sond → UV ozáření

Sklo



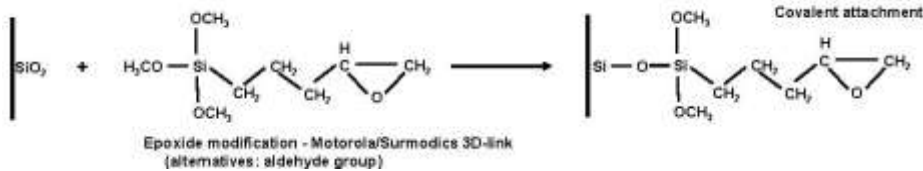
From Lee N. H. presentation: Introduction to High Density Microarrays

Úprava povrchu sklíček pro tisk arrays II

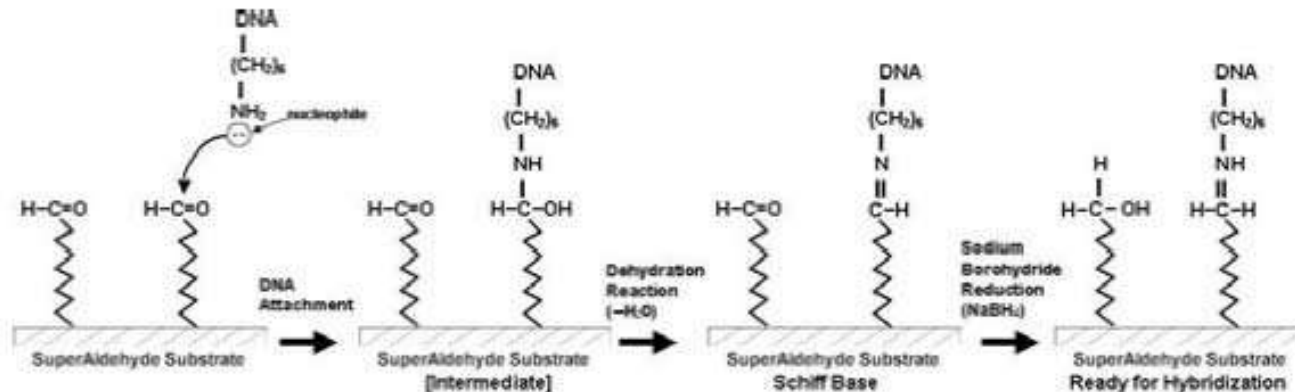
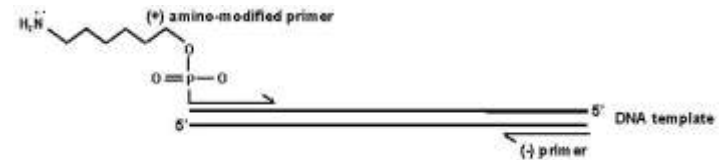
povrchová úprava skla: epoxidová modifikace

úprava DNA: amino-modifikace DNA

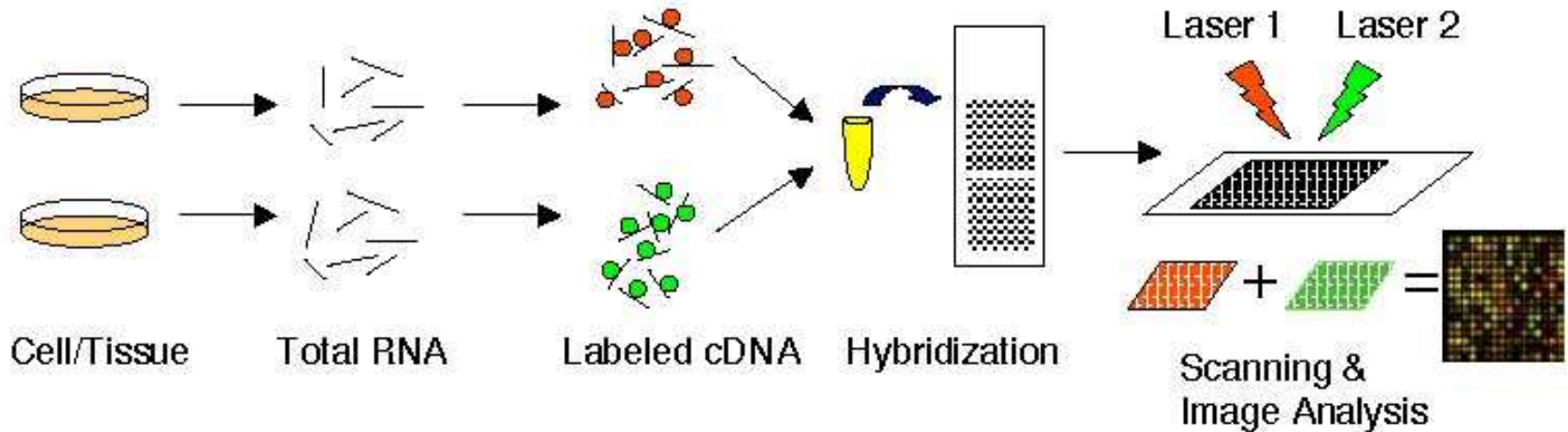
Sklo



DNA



Experimentální design

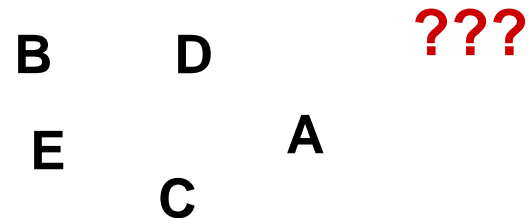
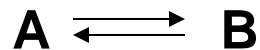


Příklady použití v molekulární biologii (na úrovni mRNA):

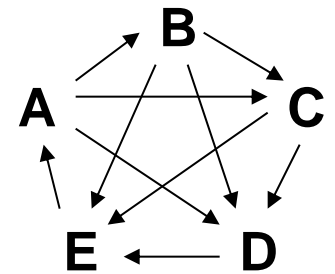
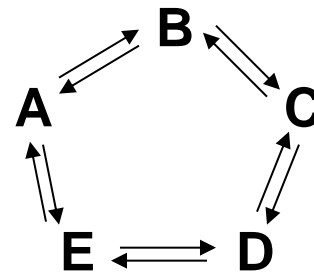
- aplikace chemické látky na buněčnou kulturu a její vliv na expresi různých genů (najít geny, které sníží nebo naopak zvýší expresi mRNA)
- zvýšení exprese mRNA zvoleného genu vnesením plasmidu → nalezení dalších genů se změněnou expresí
- snížení exprese mRNA zvoleného genu po vnesení specifické siRNA → nalezení dalších genů se změněnou expresí

Experimentální design

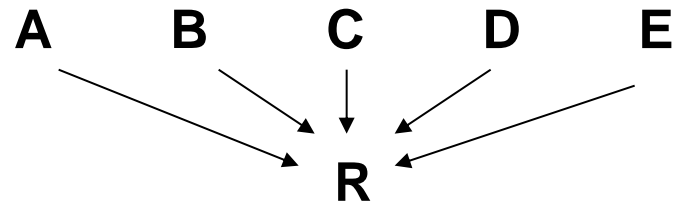
Porovnání exprese mezi vzorky:



1. **Loop design:** každé dva vzorky jsou hybridizovány na jedno sklo (plus vzájemná záměna fluorochromů)



2. **Reference design:** každý vzorek je hybridizován s referenčním vzorkem, který pak slouží jako převodník mezi různými vzorky



Experimentální design

Loop design

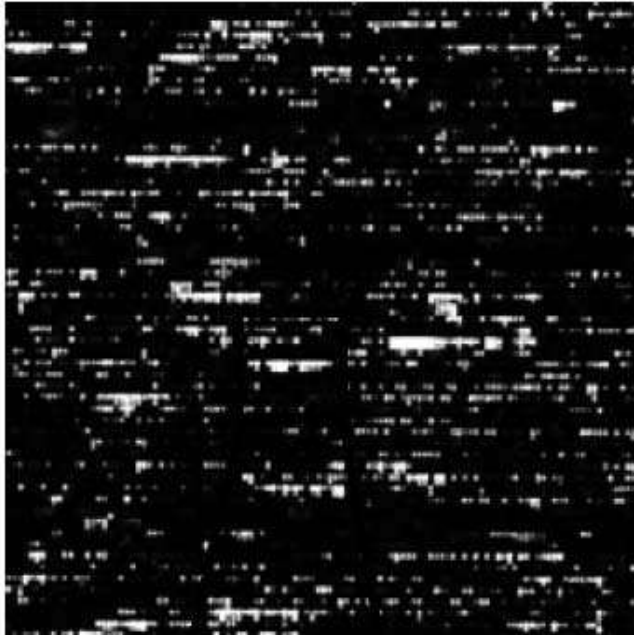
poskytuje přímé srovnání mezi vzorky
o každém vzorku získáme více informací - kontrola
vyžaduje větší množství RNA z každého vzorku
špatný vzorek více ovlivní celý experiment

Reference design

lze jednoduše rozšířit o nový vzorek
jednodušší interpretace výsledků
vyžaduje méně RNA ze vzorků
špatný vzorek méně ovlivní celý experiment

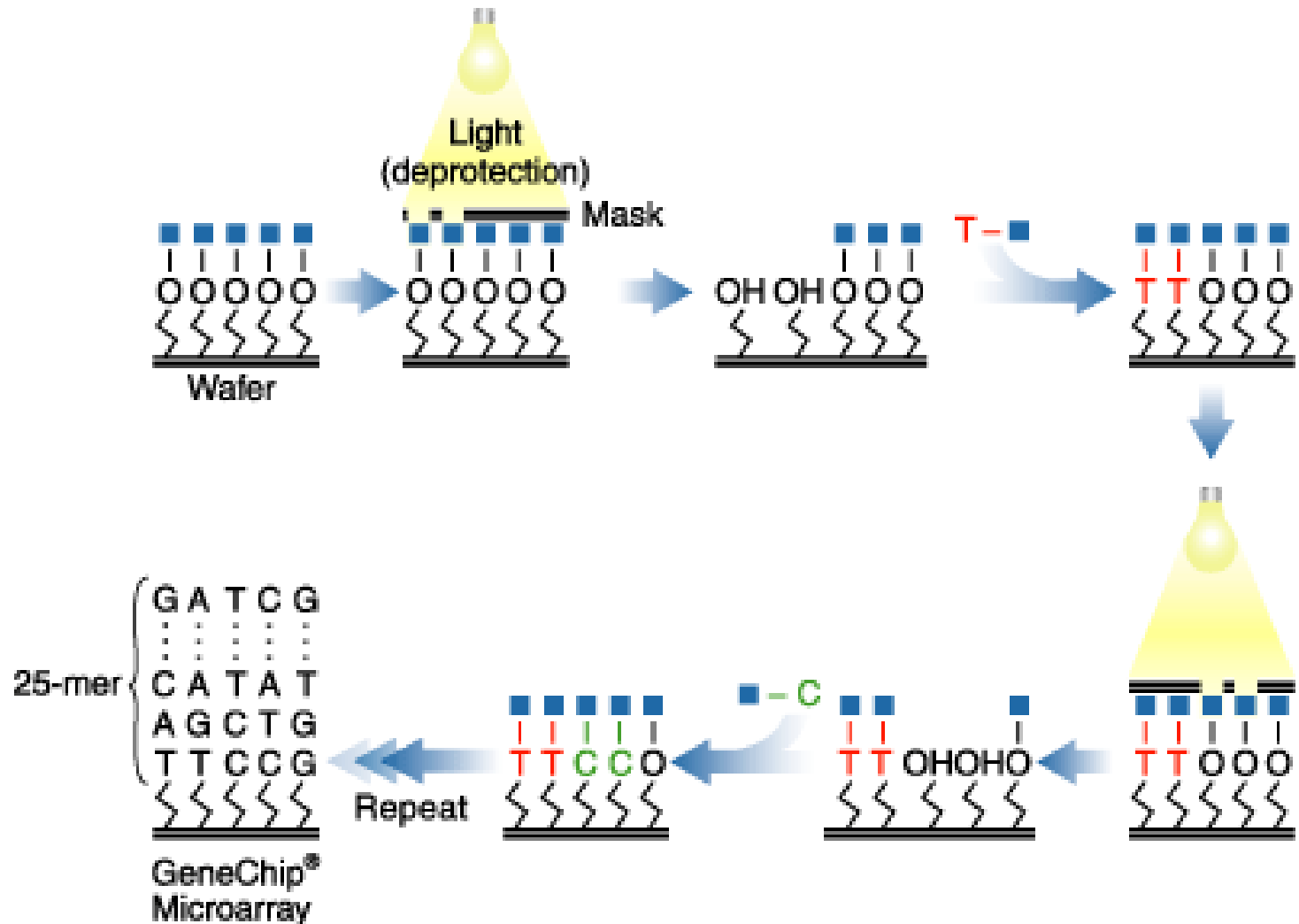
Měření množství mRNA

(fotolitograficky připravené microarrays)



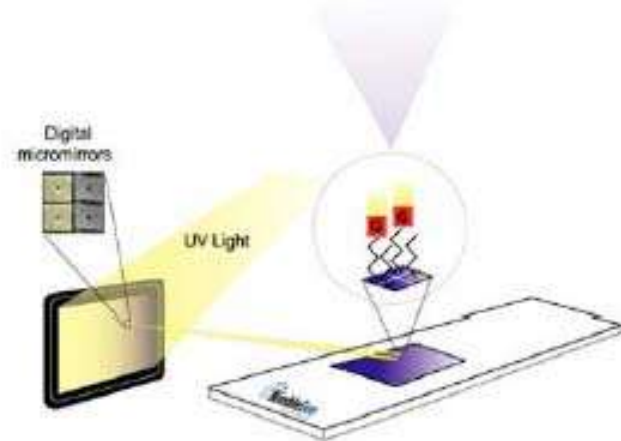
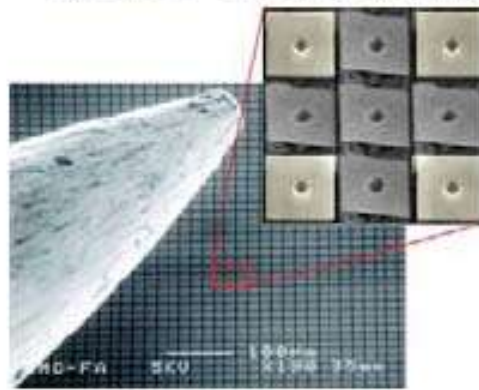
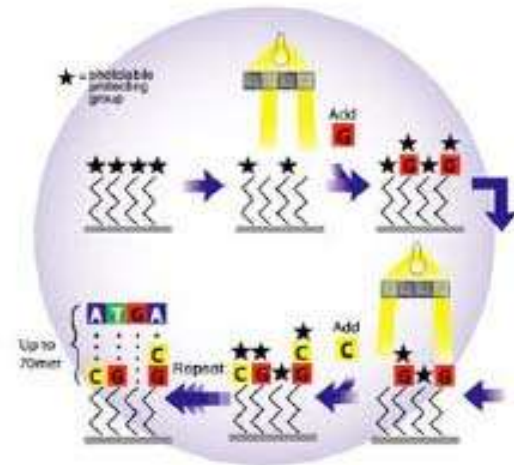
Fotolitografický způsob přípravy (Affymetrix)

sondy = oligonukleotidy délky 25 bází



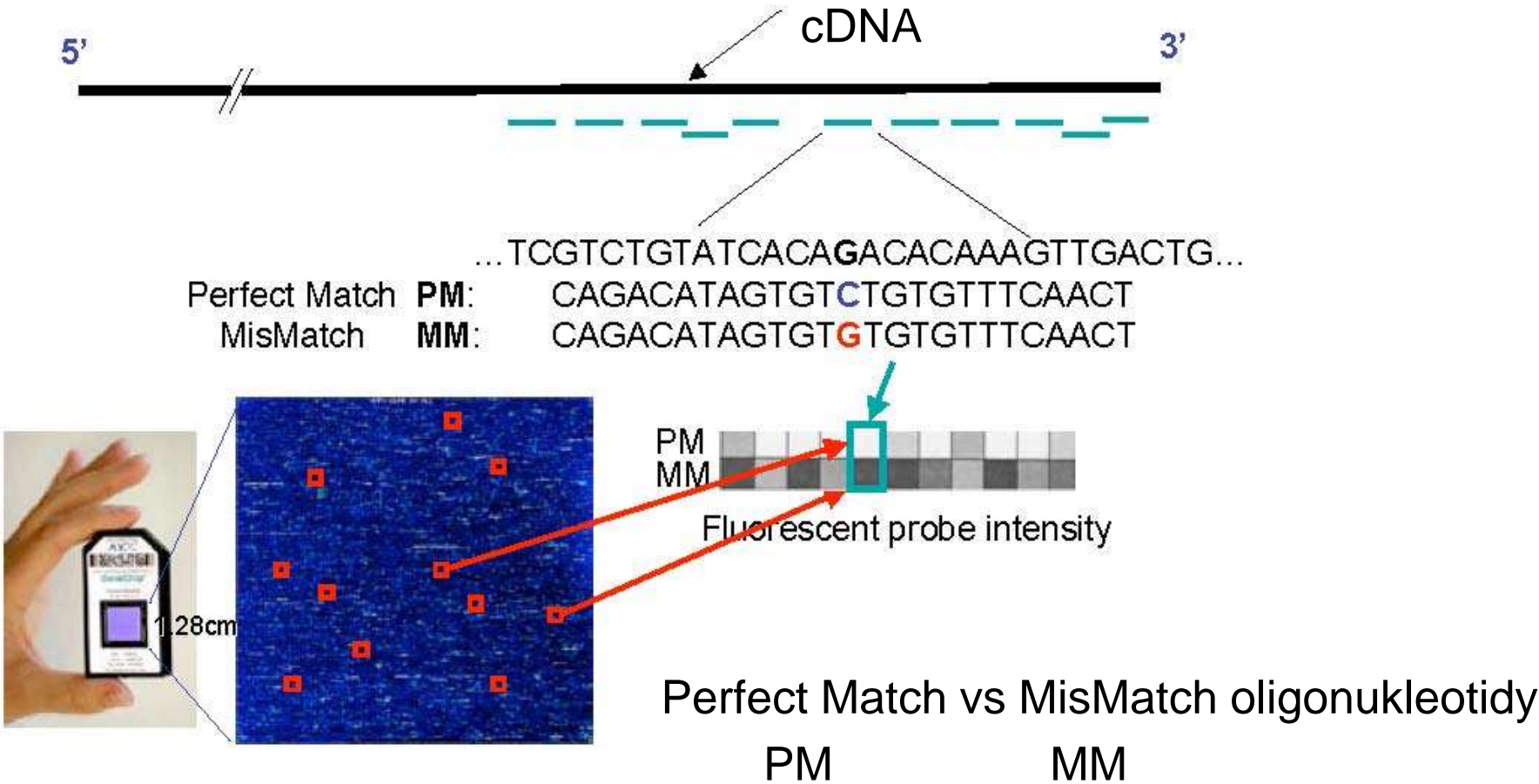
Fotolitografický způsob přípravy (II)

sondy = oligonucleotidy 45-85 bází
podobná teplota tání (T_m)



Typ sond

oligonukleotidy 25 bází



Typ sond

oligonukleotidy 25 bází

Dříve:

sondy blíže k 3' konci mRNA

11-16 na jeden gen

PM, MM sondy

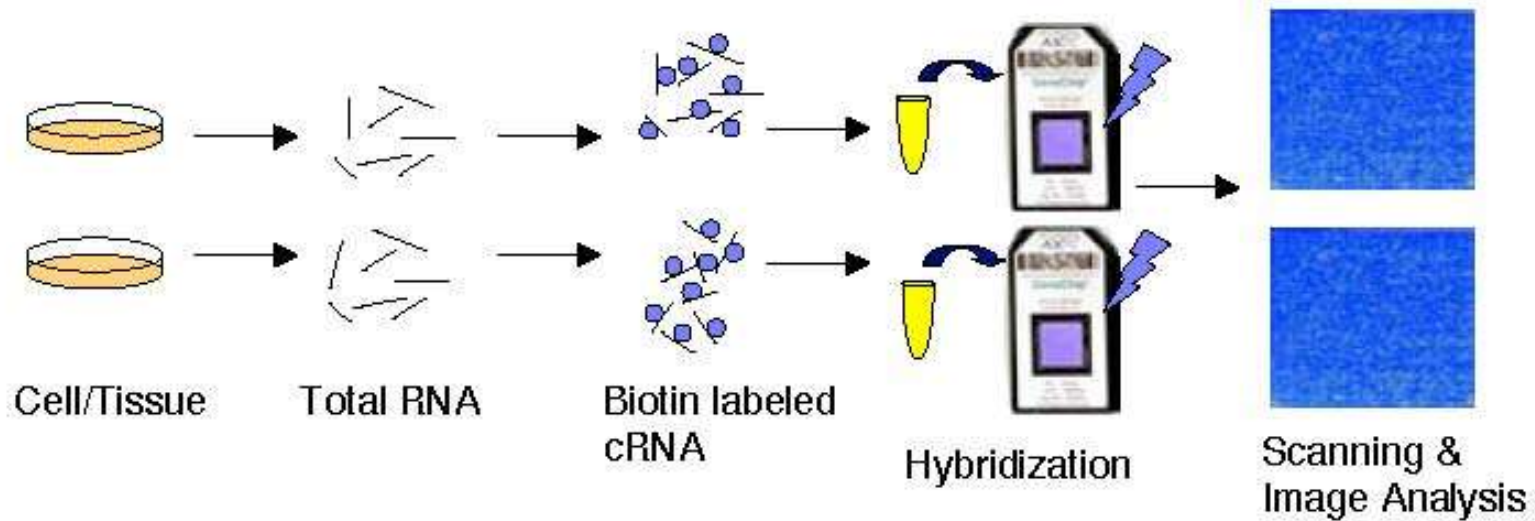
Nyní:

sondy v různých exonech genu (ideálně 4 sondy v každém exonu)

jenom PM sondy

umožňuje studovat alternativní sestřih

Experimentální design

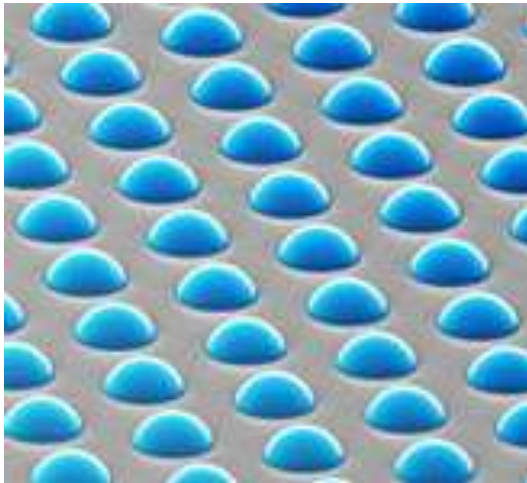


Příklady použití v molekulární biologii (na úrovni mRNA):

- aplikace chemické látky na buněčnou kulturu a její vliv na expresi různých genů (najít geny, které sníží nebo naopak zvýší expresi mRNA)
- zvýšení exprese mRNA zvoleného genu vnesením plasmidu → nalezení dalších genů se změněnou expresí
- snížení exprese mRNA zvoleného genu po vnesení specifické siRNA → nalezení dalších genů se změněnou expresí

Měření množství mRNA

(Alumina samosestavovací arrays)



Samosestavování silikonových kuliček

základní stavební jednotka: silikonová kulička (3uM)

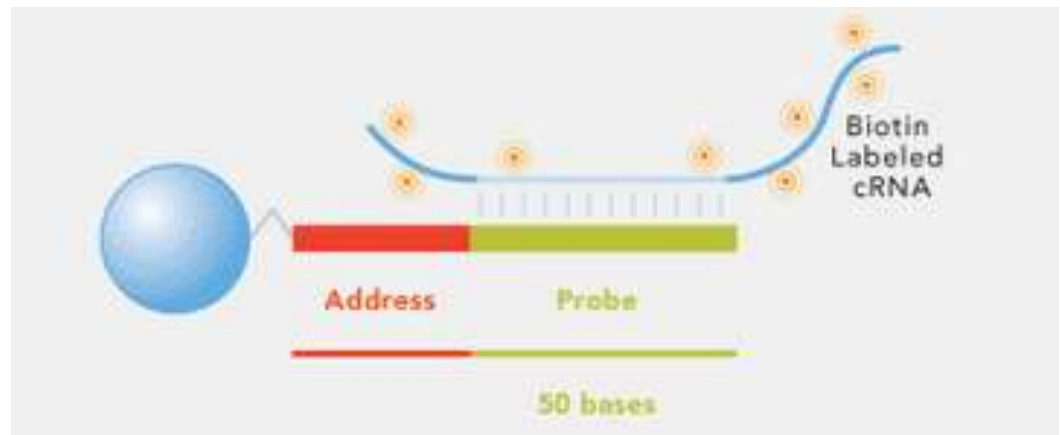
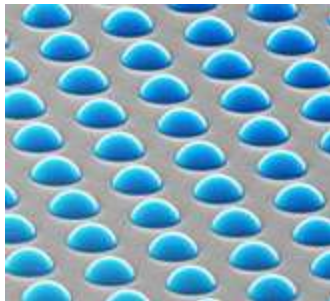
kulička nemá přesně dané místo na sklíčku, po fixaci na sklíčku je její typ identifikován díky sekvenci části oligonukleotidu

oligonukleotid:

I. **adresa** (definuje typ kuličky)

II. **vlastní sonda** - oligonucleotid (50 bp), který je specifický pro jednotlivé transkripty

míra exprese mRNA = intenzita fluorescence navázané cRNA



Objevování nových transkriptů

objevování nových transkriptů, které nejsou ještě ve veřejných databázích (např. SeqRef, Emsembl)

nebylo to možné pomocí výše zmíněných technologií, protože ty jsou založené na znalostech obsažených v databázích

Řešení:

tilling arrays (Affymetrix)

mRNA sequencing (Illumina, SOLID, Torrent, Roche 454)

„Tilling“ arrays

sondy na sklíčku pokrývají kompletně určitou oblast genomu popř. celý genom

repetitivní sekvence nejsou pokryty (před návrhem sond jsou odstraněny pomocí programu „RepeatMasker“)

sondy: oligonukleotidy

např: 14 arrays, každé obsahuje 2x 3 250 000 sond

25 bazí sonda, PM a MM, mezera mezi sondami 10 bazí

po hybridizaci s fluorescenčně značenou cRNA „svítí“ sondy, které představují transkribovaná místa ve studované oblasti (genomu)

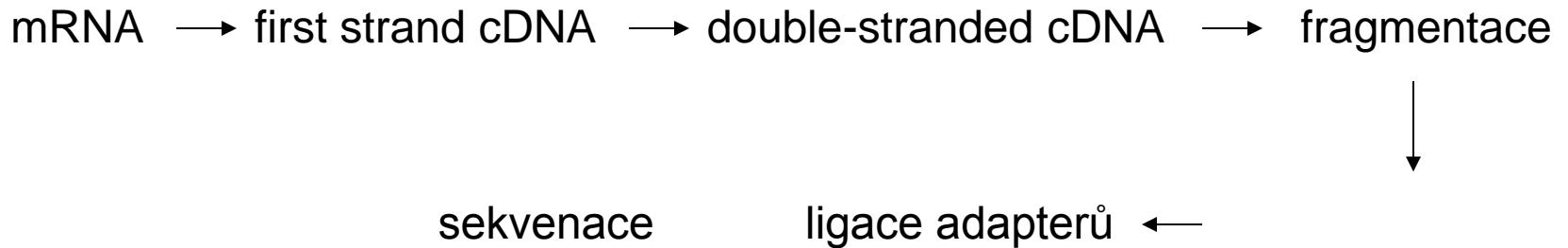
sondy v místech „bez transkripce“ mají intenzitu fluorescence na úrovni pozadí

lze detekovat nové exony, jejich alternativní sestřih

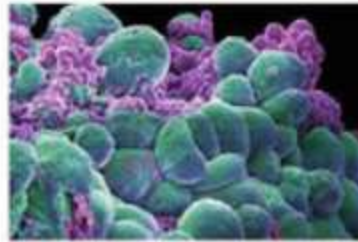
mRNA sequencing

objevování nových transkriptů pomocí sekvenační technologie

není potřeba navrhovat, tisknout nebo syntetizovat sondy



Library preparation



Attach Adapters, PCR

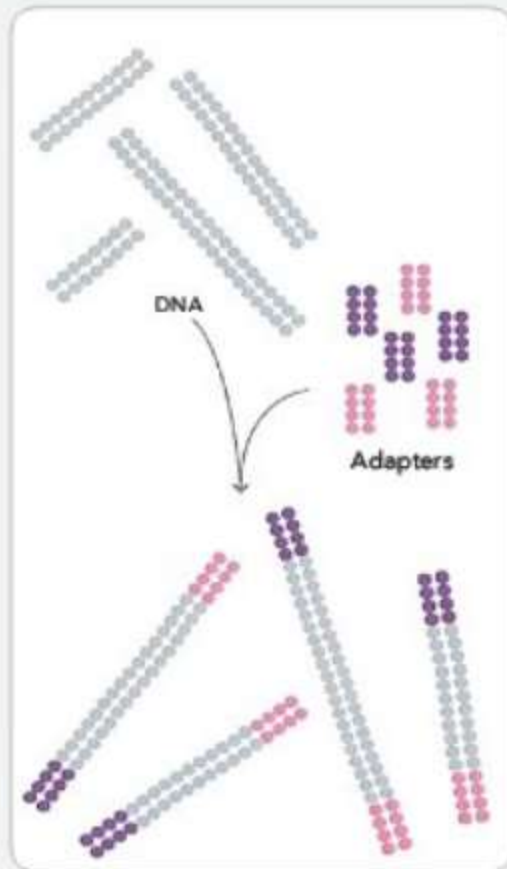


Read Alignment/Assembly

Sequencing platforms – select carefully!

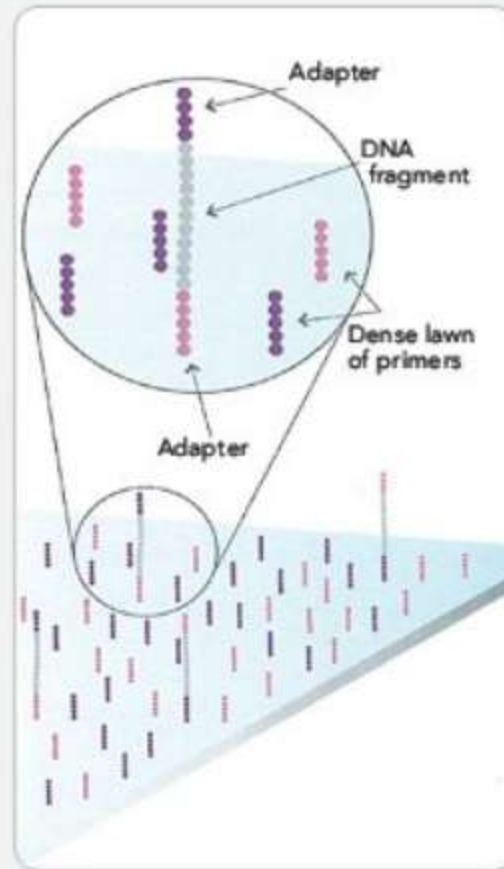
| | Run time | Millions of reads/run | Bases/read | Yield MB/run |
|-----------------------|-----------|-----------------------|------------|--------------|
| Illumina HiSeq 2000 | 11.5 days | <3,000 | 100+100 | <600,000 |
| PacBio | <2 hours | 0.03 | >3,000 | 100-150 |
| Oxford Nanopore | <6 hours | 0.1 | 9,000 | 1,000 |
| Ion Torrent Proton II | >4 hours | 250 | <200 | 50,000 |
| SOLiD 5500xl | 8 days | >1,410 | 75+35 | 155,100 |
| Illumina MiSeq | 39 hours | 15 | 250+250 | 7,500 |
| 454 FLX Titanium | 10 hours | 1 | 400 | 400 |

1. PREPARE GENOMIC DNA SAMPLE



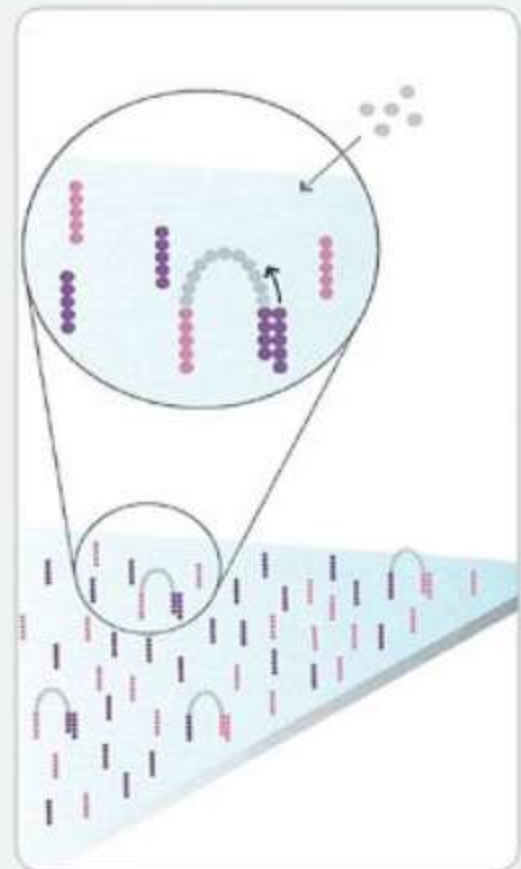
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



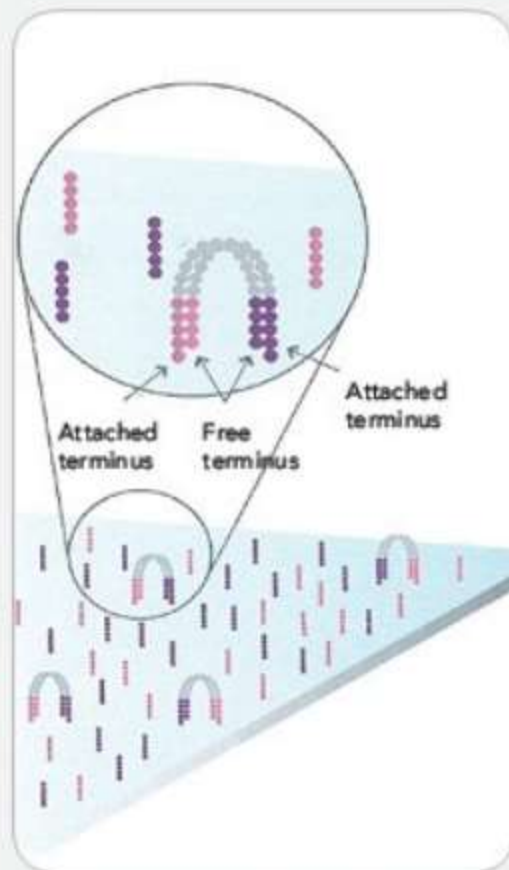
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



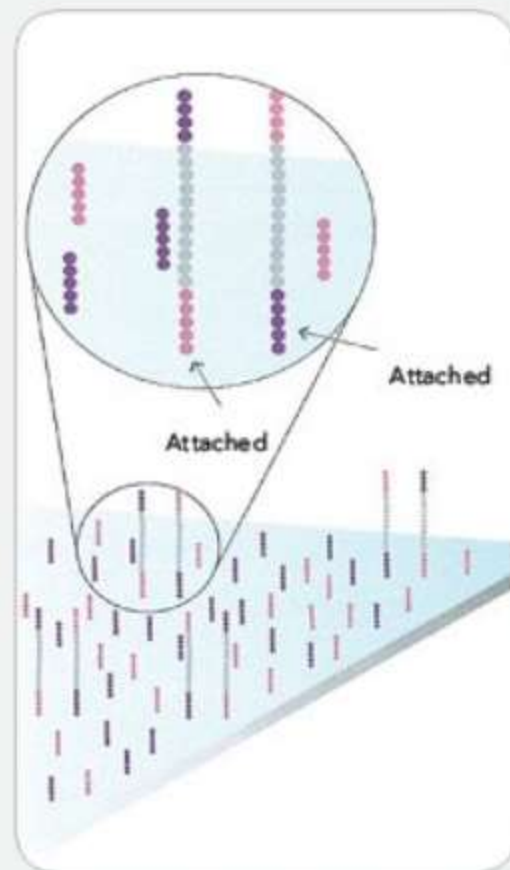
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE STRANDED



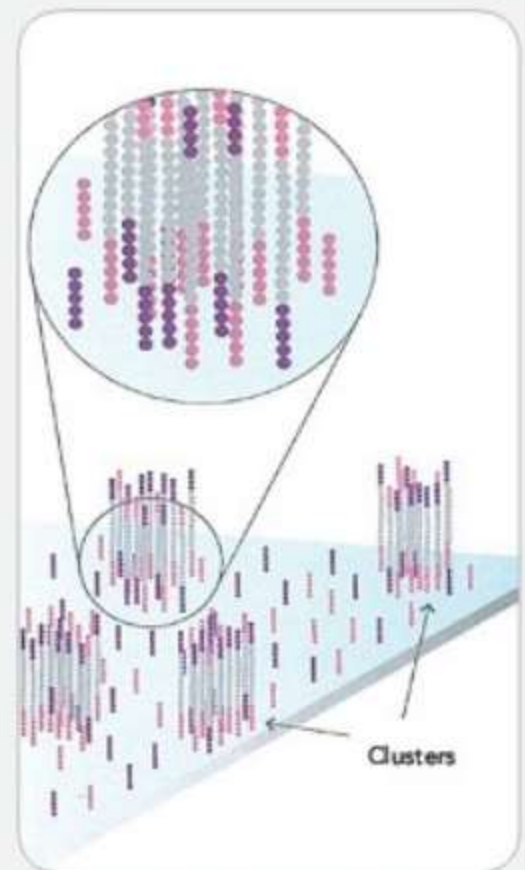
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



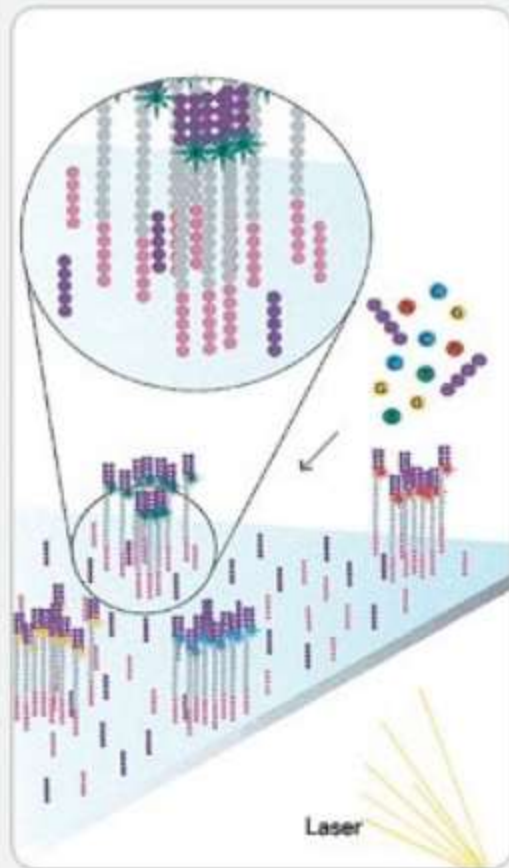
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

7. DETERMINE FIRST BASE



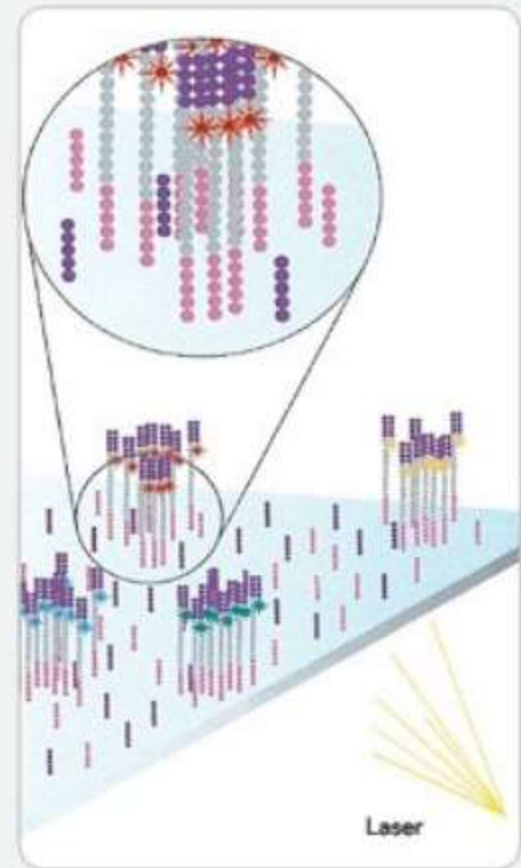
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE



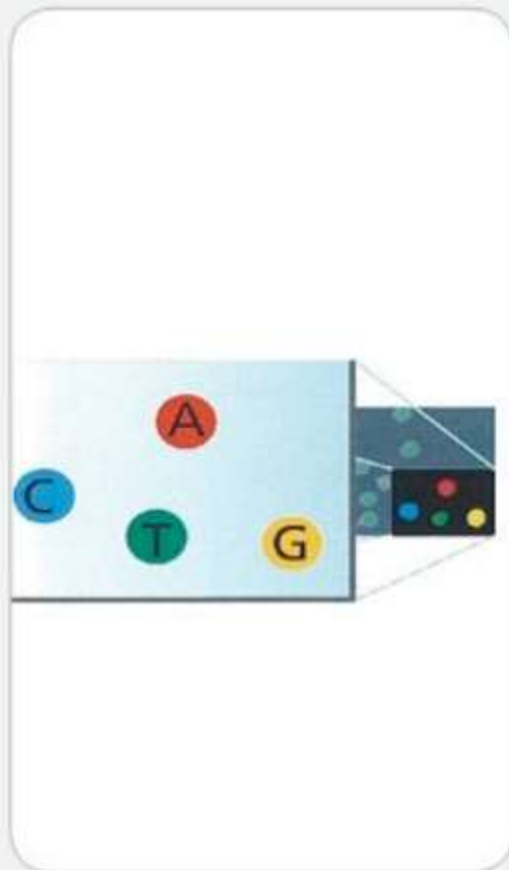
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE



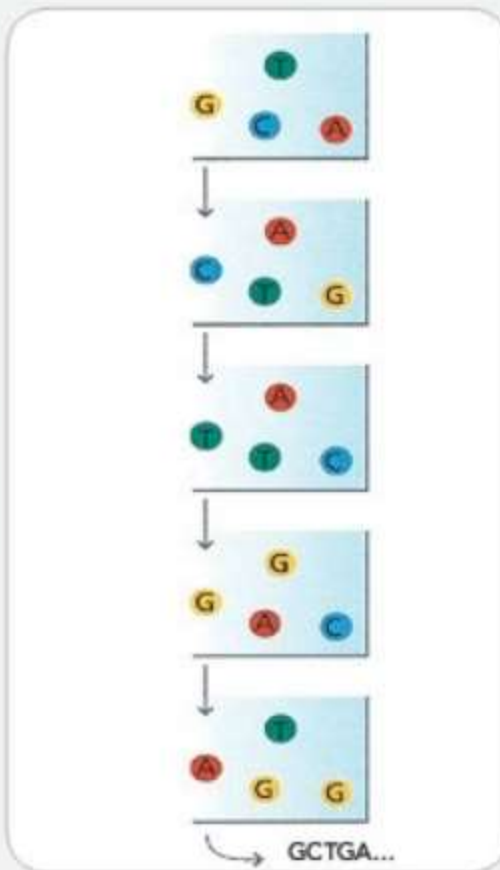
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE



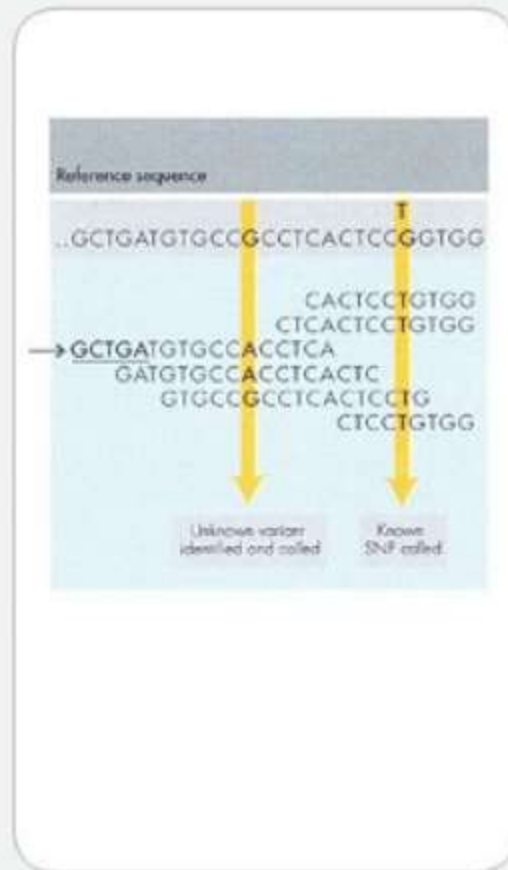
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES



Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

12. ALIGN DATA



Align data, compare to a reference, and identify sequence differences.

Úvod do statistického hodnocení dat

Předpříprava dat pro statistické hodnocení

analýza obrazu (měření intenzity bodů a pozadí)

normalizace (nalezení a odstranění systematických chyb, které nejsou způsobeny biologickým objektem)

filtrování dat (odstranění špatných bodů nebo hybridizací ze studie)

Nalezení rozdílně exprimovaných genů

výpočet zvolené statistiky a následné určení p hodnot

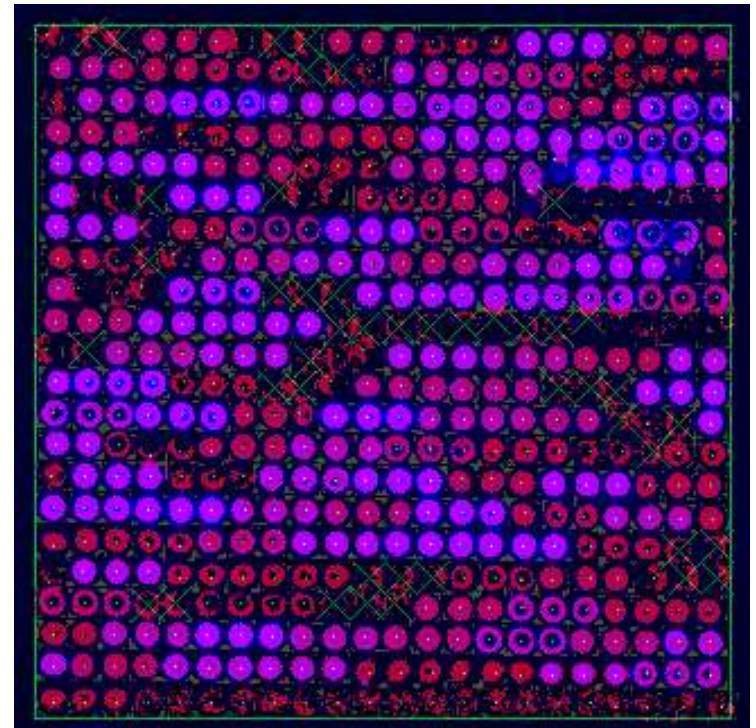
úprava p-hodnot

Analýza obrazu

rozdělení pixelů v nasnímaném obraze na ty, které nesou informaci o intenzitě bodů na sklíčku nebo pozadí

mnoho programů na analýzu microarray obrazů (GenePix, Spot, ...)

výsledek: txt soubor – každý řádek obsahuje informaci o jednom bodu na sklíčku (průměrná intenzita uvnitř bodu, intenzita okolí, variabilita mezi pixely uvnitř bodu, ...)



Subarray

Analýza obrazu

Nejdůležitější hodnota: poměr mezi intenzitami fluorescence R a G

R/G

Nejčastěji se vyjadřuje pomocí logaritmu o základu 2

$$M = \text{Log}_2 R/G$$

$$\text{Log}_2 R/G = 1$$

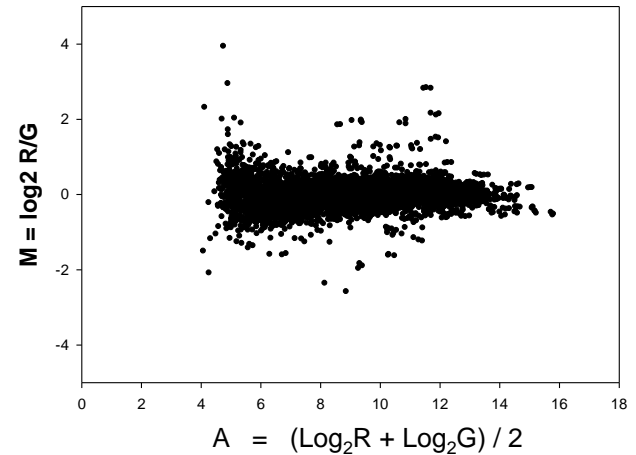
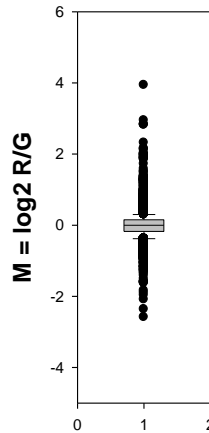
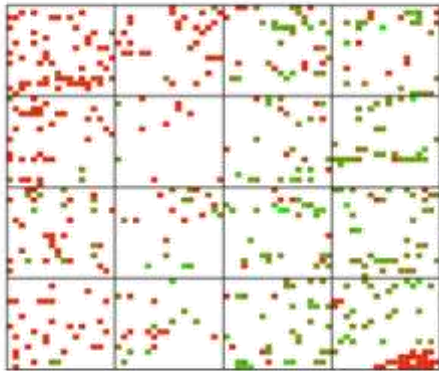
ve vzorku značeném prvním fl. je dvakrát více kopií specifické mRNA než ve vzorku značeném druhým fl.

$$\text{Log}_2 R/G = -1$$

ve vzorku značeném prvním fl. je poloviční množství kopií specifické mRNA než ve vzorku značeném druhým fl.

Analýza obrazu

$$M = \text{Log}_2 R/G$$



Další důležitá hodnota pro kontrolu kvality hybridizace je
průměrná intenzita bodu v obou snímaných kanálech

$$A = (\text{Log}_2 R + \text{Log}_2 G) / 2$$

Důležité předpoklady

Sondy na sklíčku jsou rozmístěny zcela náhodně

do stejné pozice na sklíčku neseskupujeme geny s podobnou funkcí; sekvenčně příbuzné; ležící na stejném chromosomu

Hybridizace byly prováděny v náhodném pořadí

kontroly byly hybridizovány dohromady se zkoumanými vzorky

Předpokládáme, že experiment ovlivní expresi pouze malého počtu genů v daném objektu (většina genů svoji expresi nemění)

průměr (medián) všech poměrů R/G je roven 1

průměr (medián) všech logaritmů poměrů R/G je roven 0

nestačí mít na sklíčku sondy pro geny, které nás zajímají nebo očekáváme, že jejich exprese se bude měnit

pro normalizaci jsou nutné i další geny, jejichž exprese se nemění (těch by měla být většina)

Odstranění „špatných“ bodů

**odstranění bodů: body s morfológickými abnormalitami
(problematický tisk)**

**s nízkou intenzitou (není exprese v
daném systému)**

s vysokým pozadím (negativní hybridizace)

Kontrolní body: prázdné body bez DNA (negativní kontrola)

„spiked“ body (pozitivní kontrola)

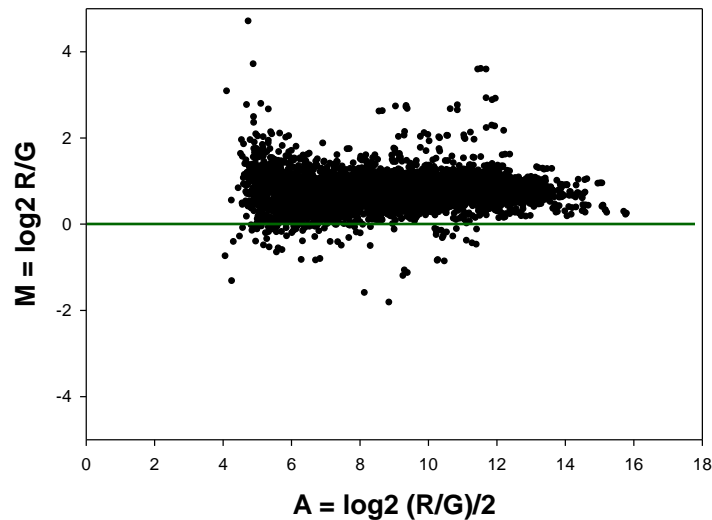
stejné sondy na různých místech sklíčka

Normalizace

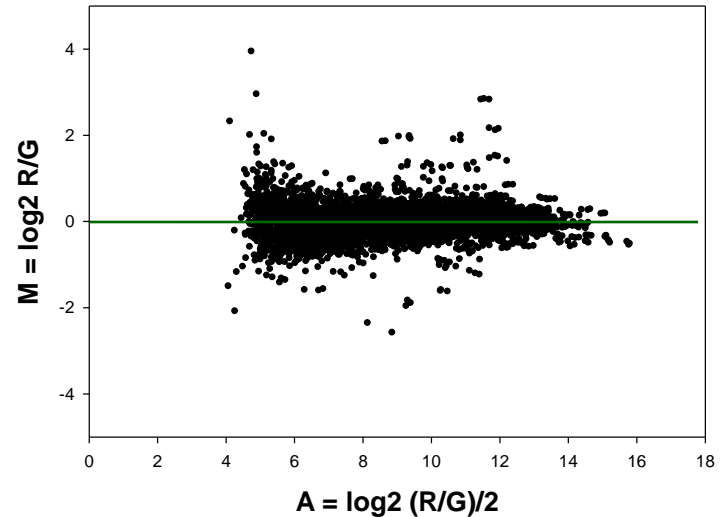
nalezení a odstranění systematických chyb, které nejsou způsobeny biologickým objektem

není splněná podmínka, že průměr (medián) všech logaritmů poměrů R/G je roven 0

Před normalizací:



Po normalizaci:



Nalezení rozdílně exprimovaných genů

Nulová hypotéza: medián exprese daného genu se statisticky neliší od teoretické hodnoty mediánu (v našem případě 0)

$$T = \frac{\bar{M}}{se(\bar{M})} \dots\dots \text{p hodnota}$$

riziko s jakou lze nulovou hypotézu odmítnout

rozdílně exprimované geny ... p hodnota < 0.01 (volitelný práh)

| : | Array 1 | Array 2 | Array 3 | Array 4 | p hodnota |
|---------|---------|---------|---------|---------|-----------|
| Gen 111 | 0.39 | | -0.39 | 0.06 | 0.78 |
| Gen 112 | -0.28 | 0.33 | 0.37 | 0.64 | 0.25 |
| Gen 113 | | 0.14 | 0.28 | 0.44 | 0.38 |
| Gen 114 | -0.19 | 0.13 | -0.13 | 0.38 | 0.99 |
| Gen 115 | 0.88 | 0.49 | 0.54 | 0.45 | 0.02 |
| : | | | | | |

Statistické problémy při studiu tisíců genů s malým počtem opakování experimentů

rozdílně exprimované geny ... p hodnota < 0.01

Příklad:

studujeme 20 000 genů na jednom sklíčku

během normalizace a kontroly kvality vyřadíme 12000 genů

testujeme 8 000 genů (pro každý vypočítáme p hodnotu)

p hodnota < 0.01 připouštíme, že 1% testovaných genů je označeno jako rozdílně exprimované pouze náhodnou variabilitou pokusů

$$8000 * 0.01 = 80 \text{ genů}$$

→ korekce p hodnot s ohledem k počtu testovaných genů

→ použití alternativních statistik

Klastrování

Klastrování (shluková analýza) je obecná metoda, kterou je možno použít ke spojování prvků (s podobnými vlastnostmi) do skupin (klastrů)

Microarray analýza:

Klastrování genů (řádků) → identifikace skupin genů, které mohou být společně regulované

Klastrování vzorků (sloupců) → nalezení skupin vzorků, které mají podobné změny v expresi genů (změny na úrovni DNA)

Příklad:

Sorlie et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. PNAS 98: 10869-10874, 2001.

Design experimentu

78 karcinomů prsu (71 duktálních, 5 lobulárních a 2 in-situ)

3 fibroadenomy

4 vzorky normální tkáně prsu

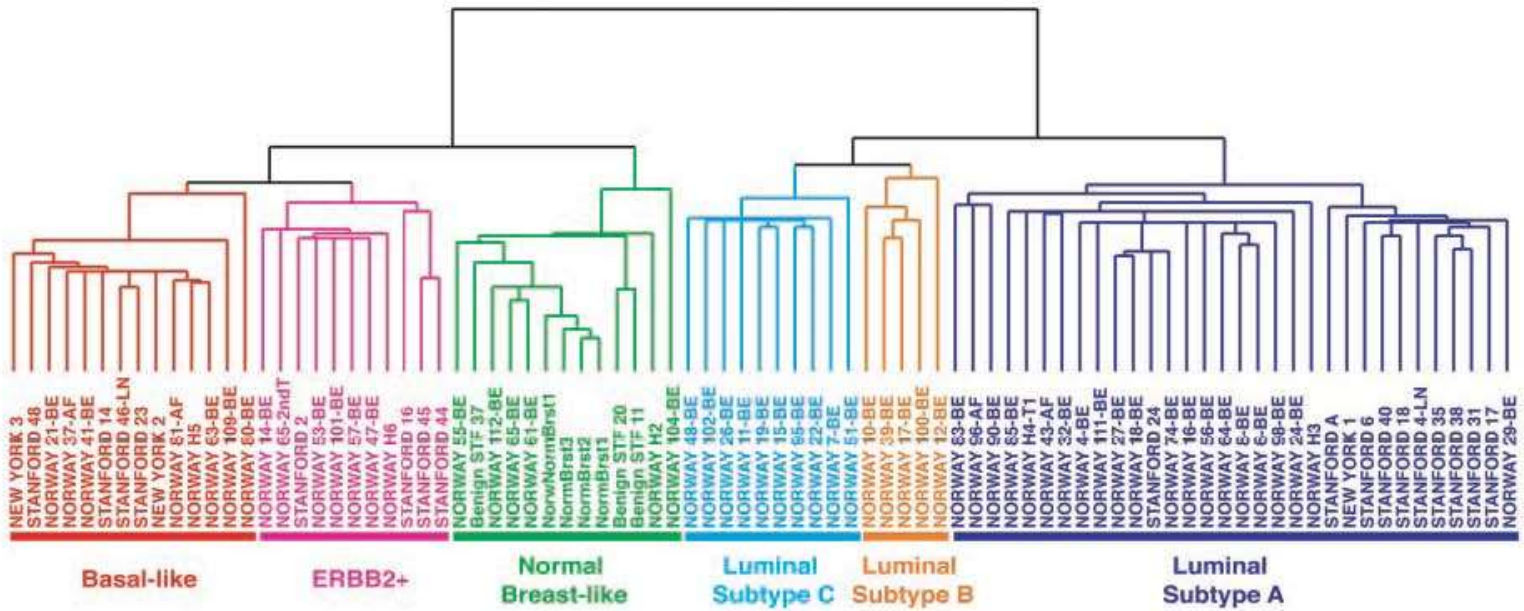
Microarrays: 8 102 cDNA klonů

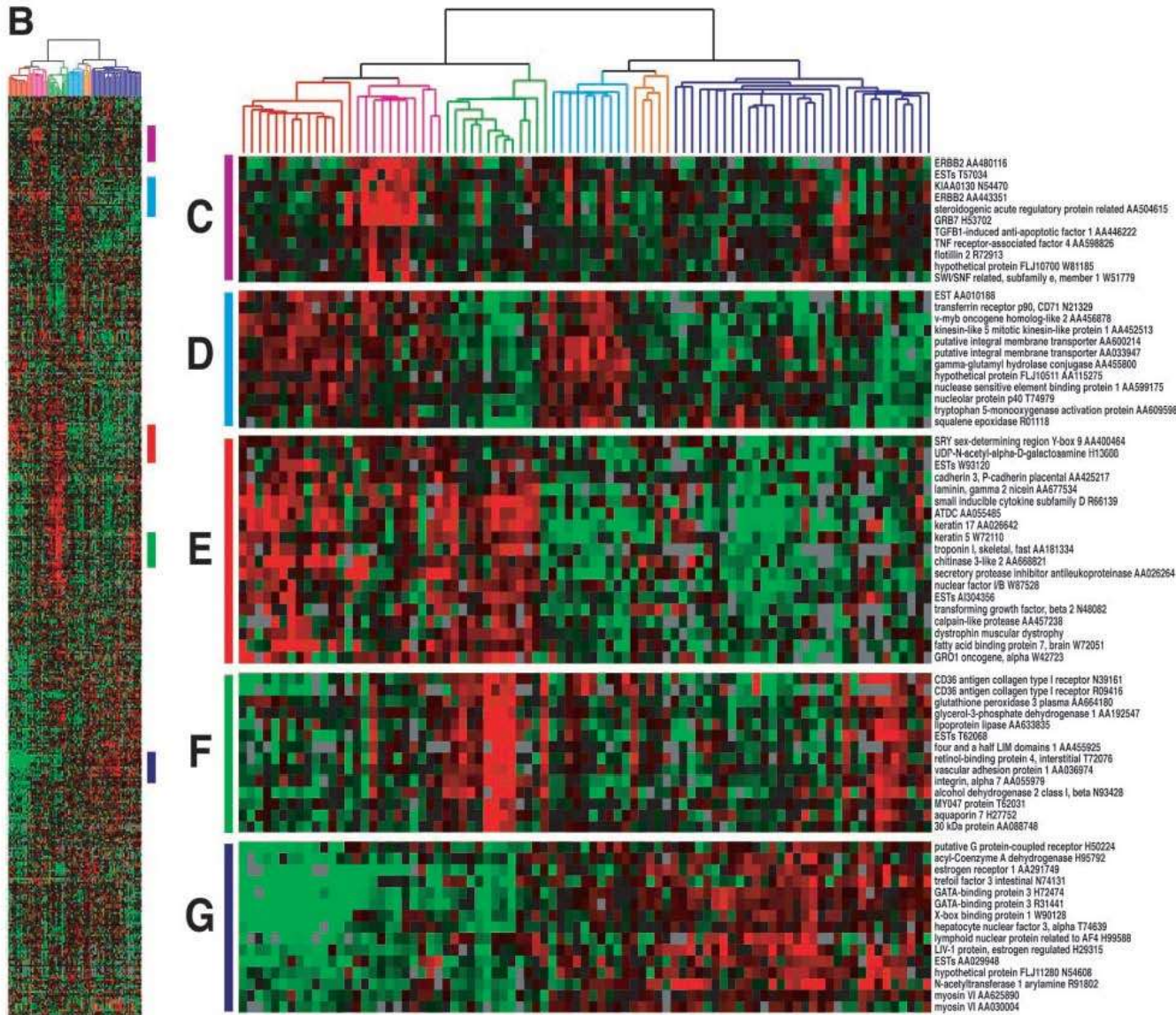
každý vzorek (Cy3) hybridizován s referenční RNA (Cy5)

Analýza: nalezeno 456 cDNA klonů (427 genů) s velkou variabilitou exprese mezi různými vzorky, ale podobnou expresí u příbuzných vzorků

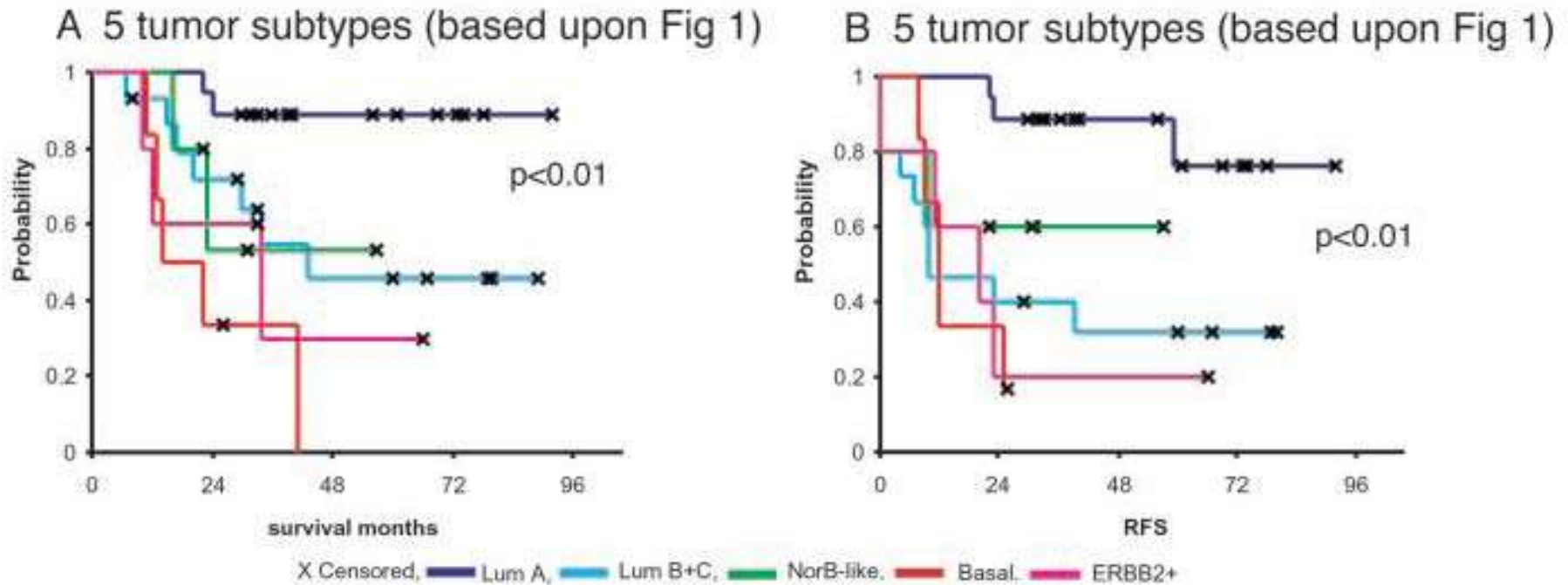
Otázka: Zda existuje rozdělení karcinomů do podskupin, které mají podobné změny v expresi genů?

Klastrování





Rozdělení do skupin a prognóza vývoje onemocnění



Podobné studie

- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536.
- Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, et al. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98: 262-272.
- Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, et al. (2004) Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2: E7.
- Paik S, Shak S, Tang G, Kim C, Baker J, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351: 2817-2826.
- Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, et al. (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436: 518-524.
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, et al. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365: 671-679.
- Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, et al. (2006) Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 3: e47.
- Liu R, Wang X, Chen GY, Dalerba P, Gurney A, et al. (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356: 217-226.
- Hu Z, Fan C, Oh DS, Marron JS, He X, et al. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7: 96.
- Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, et al. (2008) Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* 14: 518-527.
- van Vliet MH, Reyal F, Horlings HM, van de Vijver MJ, Reinders MJ, et al. (2008) Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genomics* 9: 375.

Veřejné databáze microarray dat

ArrayExpress

ChipDB

ExpressDB

Gene Expression Atlas

Gene Expression Database (GXD)

[Gene Expression Omnibus \(GEO\)](#)

GeneX

GermOnline

Human Gene Expression Index (HuGE Index)

List Of Lists Annotated (LOLA)

M-CHiPS (Multi-Conditional Hybridization Intensity Processing System)

MUSC DNA Microarray Database

NASCArrays

Oncomine

Public Expression Profiling Resource (PEPR)

READ (RIKEN cDNA Expression Array Database)

Rice Expression Database (RED)

RNA Abundance Database (RAD)

Saccharomyces Genome Database (SGD): Expression Connection

SGMD

[Stanford Microarray Database \(SMD\)](#)

Yale Microarray Database

yeast Microarray Global Viewer (yMGV)

Použití microarrays ke studiu DNA

Komparativní genomická hybridizace

BAC arrays

oligo arrays

SNP arrays

tilling arrays (BAC a oligonukleotidy)

exon-specific arrays

(dříve i cDNA arrays používané pro expresi)

Genotypování

SNP arrays

Sekvenování

Re-Sequencing arrays

ChIP-Chip experimenty

tilling arrays (oligonukleotidy)

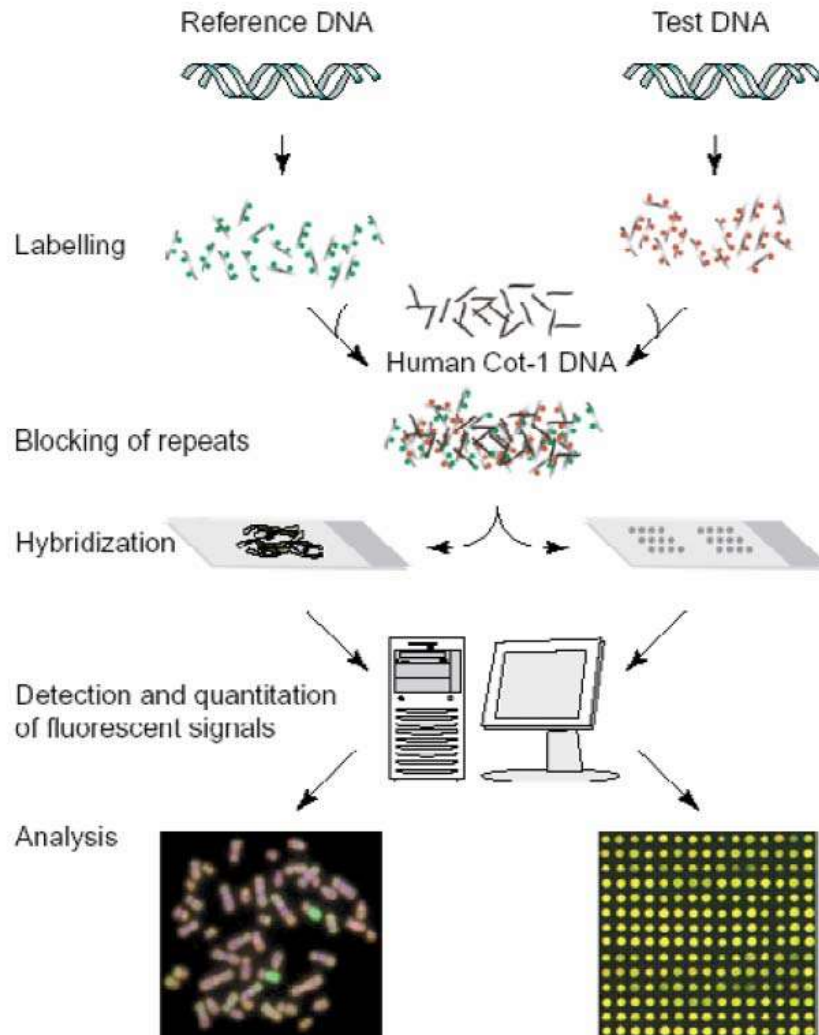
Komparativní genomická hybridizace (CGH)

molekulárně cytogenetická metoda, která slouží k analýze změn obsahu DNA v živých organismech

(delece, zisk, amplifikace různých oblastí genomu)

porovnávání intenzity fluorescence zkoumaného vzorku DNA a normálního diploidního vzorku DNA v různých místech genomu

Komparativní genomická hybridizace (CGH)



Komparativní genomická hybridizace (CGH)

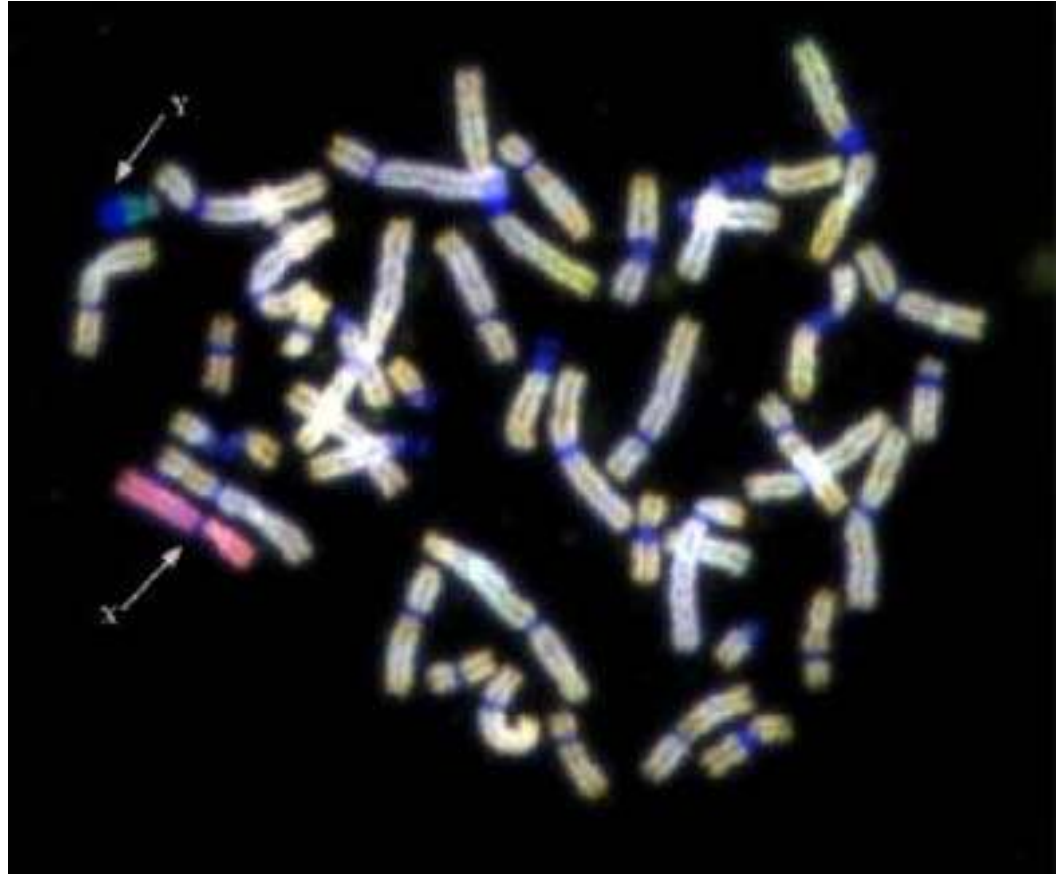
metafázní chromozomy
- dárce s normálním
diploidním karyotypem

DNA:

cy3
zkoumaný vzorek

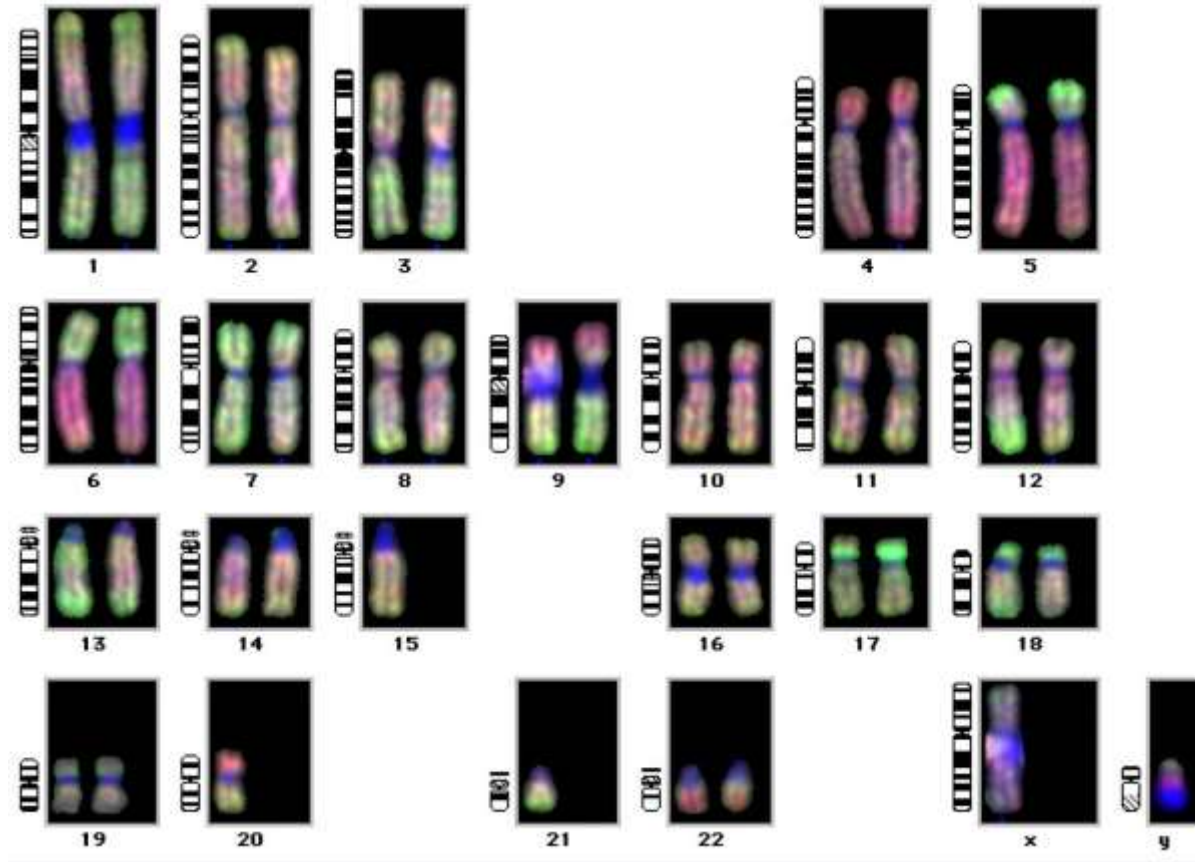
cy5
referenční DNA – 2n

rozlišení ~ 20MB



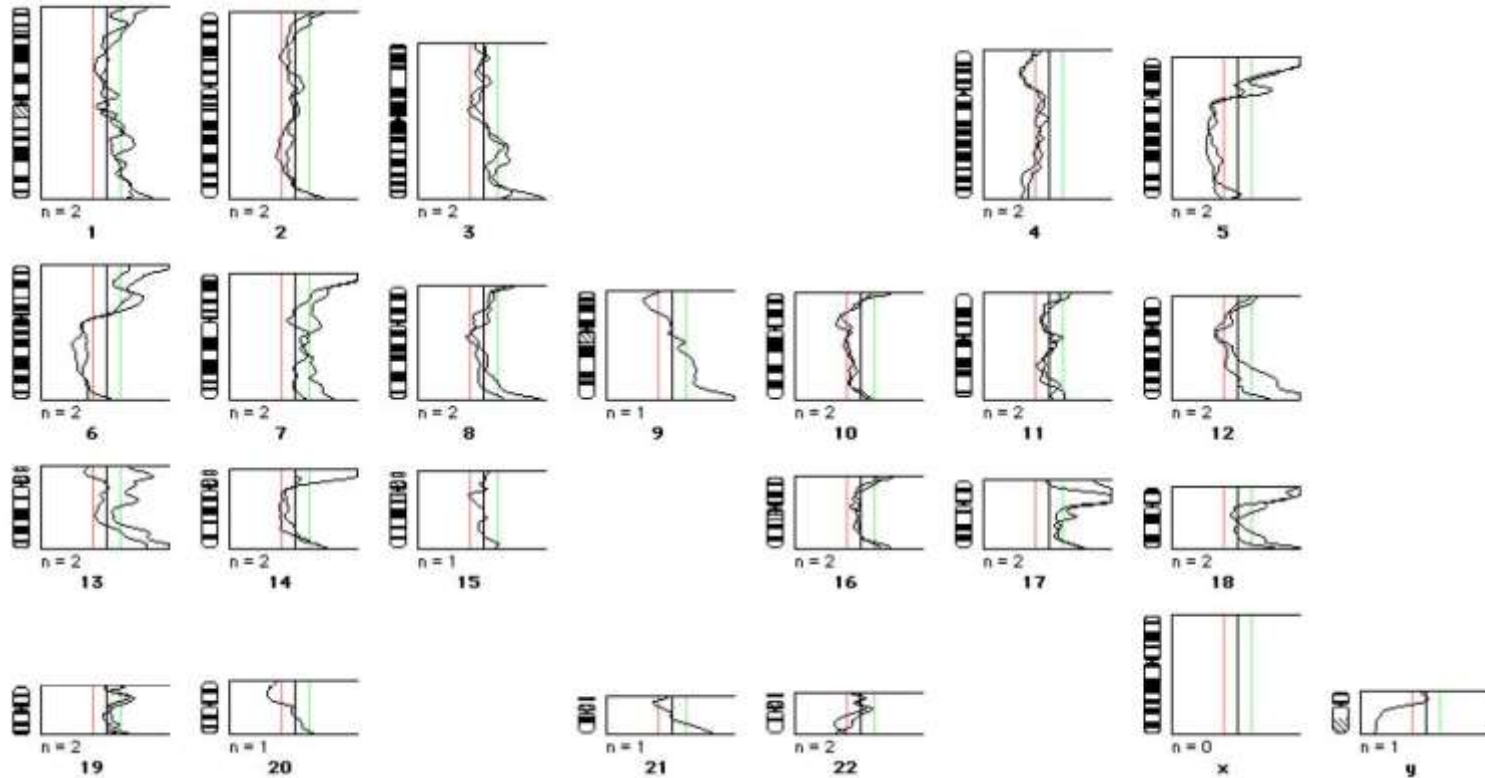
From Szuhai K. presentation: Determination of Genomic Imbalances by Genome-wide Screening Approaches

Komparativní genomická hybridizace (CGH)



From Szuhai K. presentation: Determination of Genomic Imbalances by Genome-wide Screening Approaches

Komparativní genomická hybridizace (CGH)



From Szuhai K. presentation: Determination of Genomic Imbalances by Genome-wide Screening Approaches

„Array“ komparativní genomická hybridizace (Array CGH)

**chromozomy nahrazeny body na mikroskopickém sklíčku,
které obsahují specifické DNA sekvence**

Typy sond natištěných na microarray sklíčku

BAC klony

až 32 000 BAC klonů na jednom sklíčku
~ 160 kb dlouhé úseky DNA

Oligonukleotidy

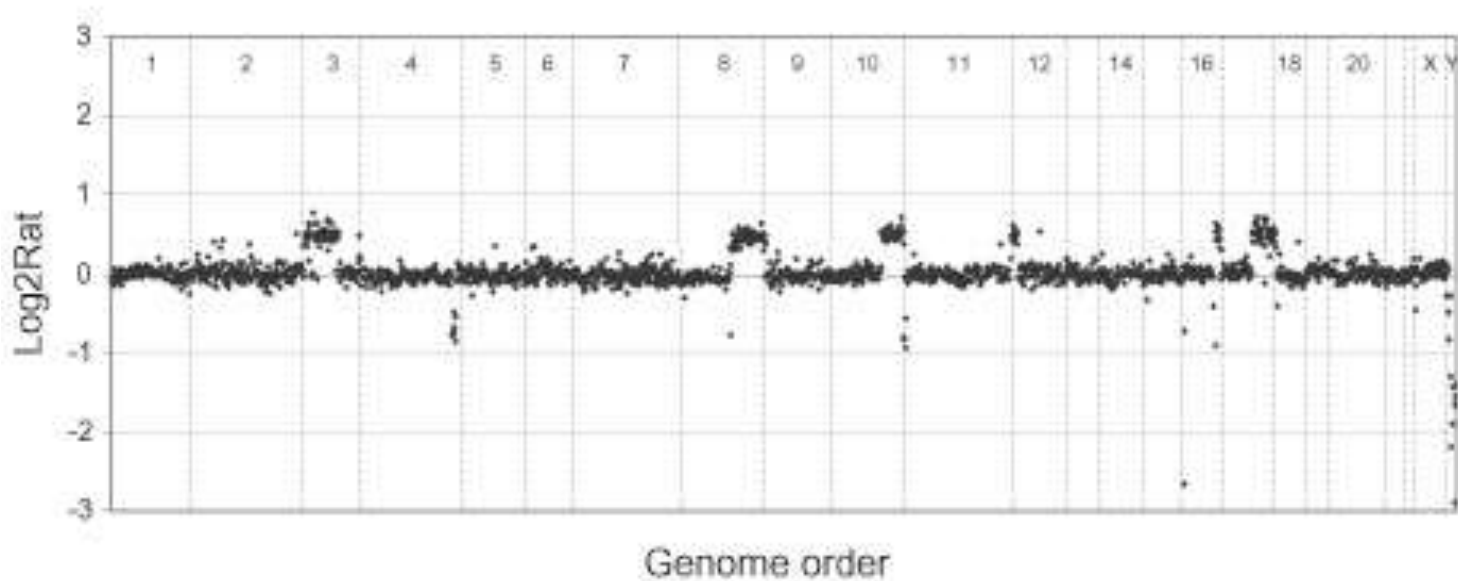
25 – 80 bazí dlouhé oligonukleotidy
mohou pokrývat i celý genom (repetitivní
sekvence jsou vynechány)

známe polohu a pořadí všech sond v lidském genomu

Array CGH s použitím BAC klonů

$$\text{Log}_2\text{Rat} = \text{Log}_2 R/G$$

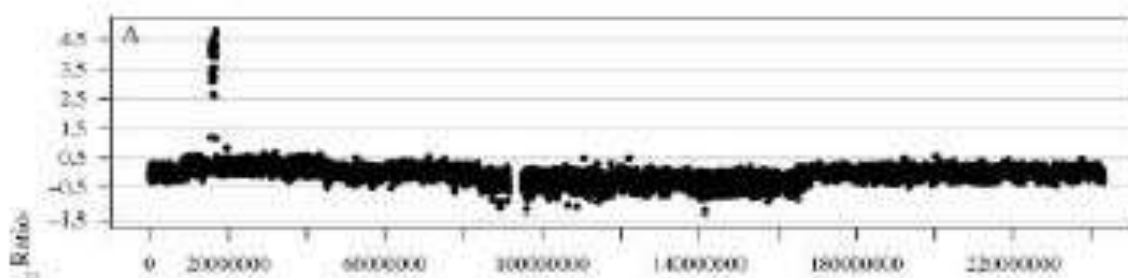
| | | | |
|--------------------------------|---------------------------|-------------------------------|--------------------|
| $\text{Log}_2\text{Rat} = 0$ | 2 kopie | $\text{Log}_2\text{Rat} = -1$ | 1 kopie (“loss”) |
| $\text{Log}_2\text{Rat} = 0.5$ | 3 kopie (“gain”) | $\text{Log}_2\text{Rat} < -1$ | homozygotní delece |
| $\text{Log}_2\text{Rat} = 1$ | 4 kopie (“gain”) | | |
| $\text{Log}_2\text{Rat} = 2$ | 8 kopií (“amplification”) | | |



2464 BAC klonů

UCSF HumArray3.1

Array CGH - oligonukleotidy (NimbleGen)



6-kb median
probe spacing

SNPs

SNP = single nucleotide polymorphism

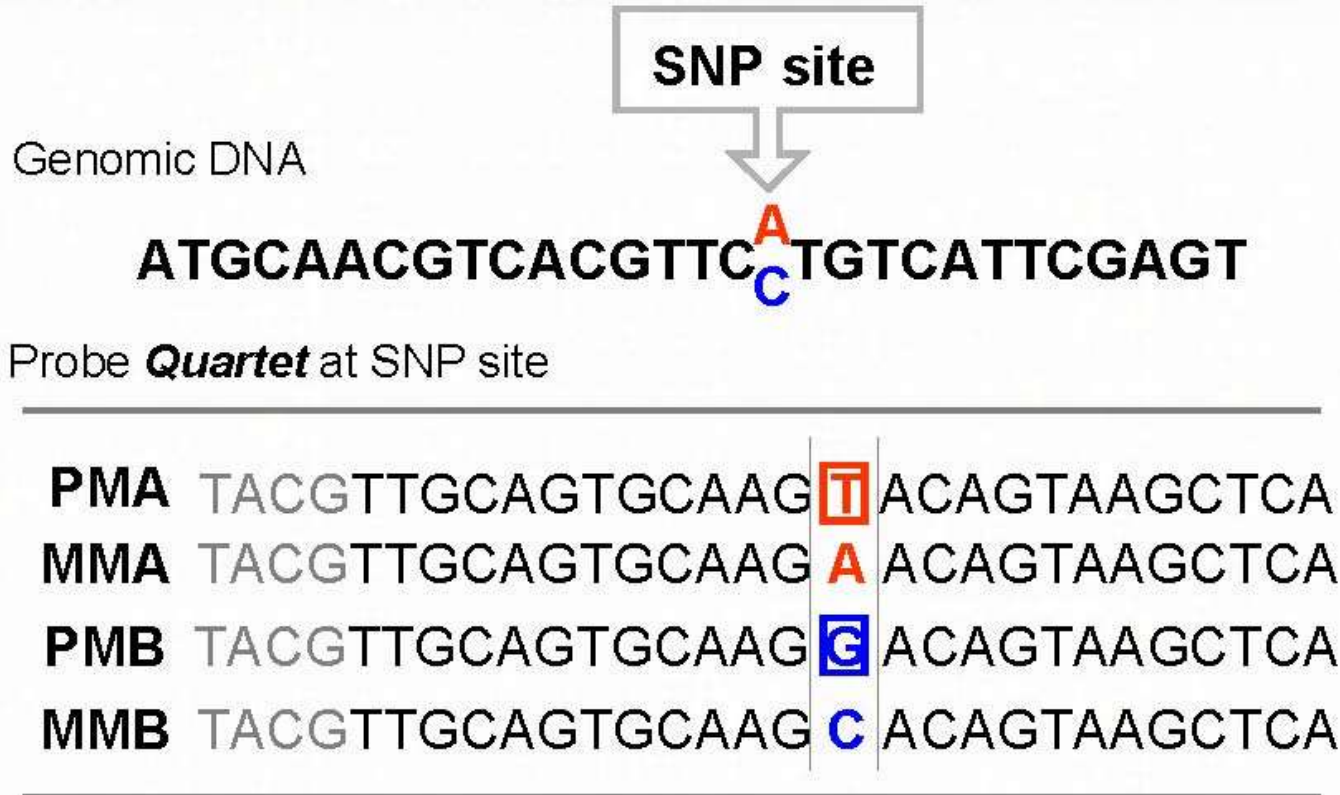
jednonukleotidové variace, které jsou náhodně rozmístěny v genomu (bodové mutace rozšířené v populaci)

nukleotidová variace, která se vyskytuje alespoň u 1% jedinců v populaci

předpokládaný počet SNPs: 10 milionů

výskyt specifických SNP spojen s predispozicí k určitým chorobám

SNP Arrays – probe design (Affymetrix)

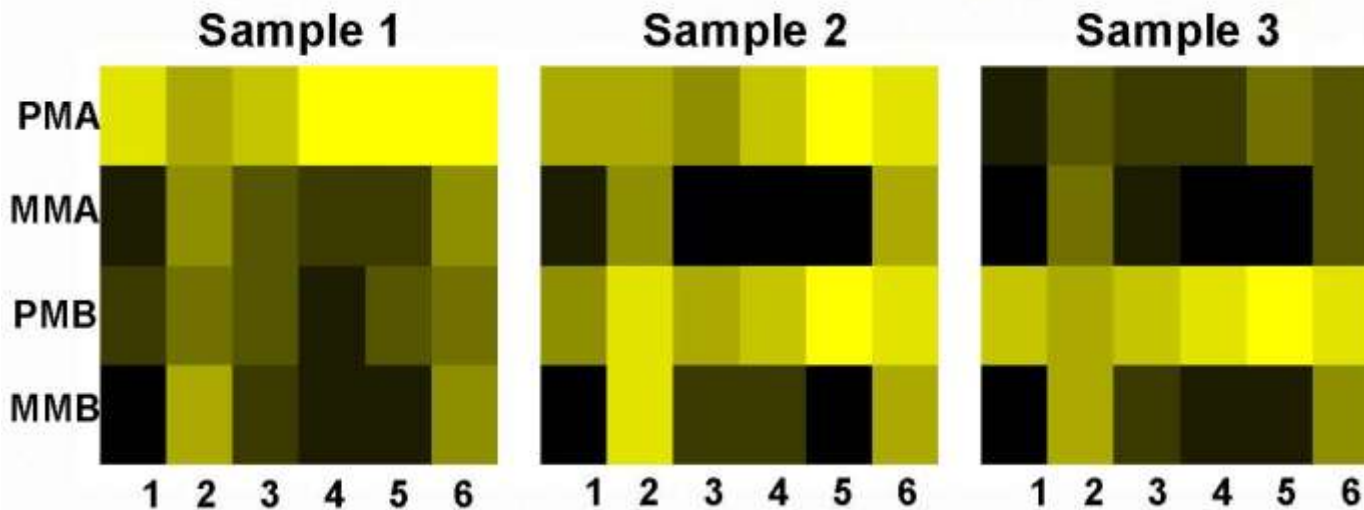


From Xiao Y. presentation: Exploration and Analysis of Affymetrix SNP Arrays. Center for Bioinformatics & Molecular Biostatistics, UCSF Division of Biostatistics

SNP Arrays – probe design

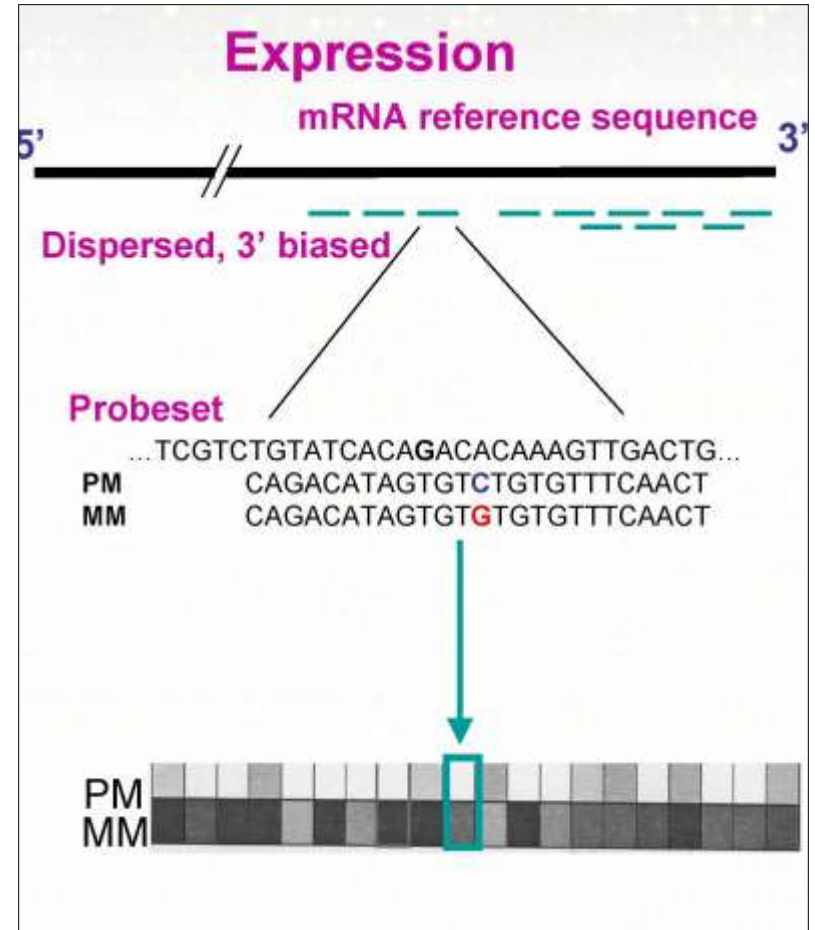
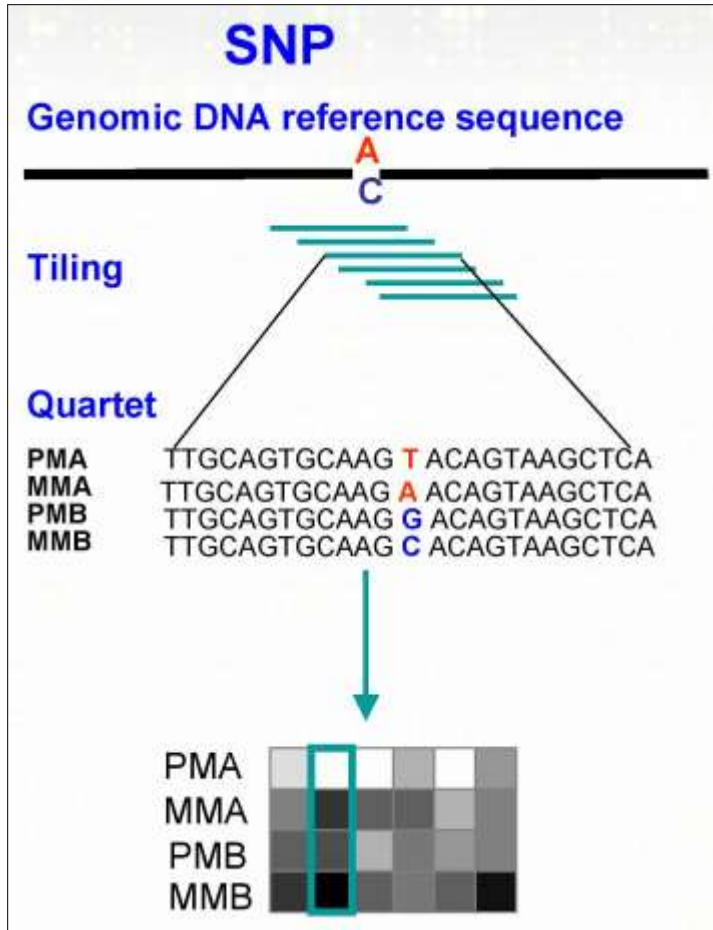
| -4 | -2 | -1 | 0 | 1 | 4 |
|------------------|------------------|------------------|------------------|------------------|------------------|
| PMA ₁ | PMA ₂ | PMA ₃ | PMA ₄ | PMA ₅ | PMA ₆ |
| MMA ₁ | MMA ₂ | MMA ₃ | MMA ₄ | MMA ₅ | MMA ₆ |
| PMB ₁ | PMB ₂ | PMB ₃ | PMB ₄ | PMB ₅ | PMB ₆ |
| MMB ₁ | MMB ₂ | MMB ₃ | MMB ₄ | MMB ₅ | MMB ₆ |

| Arrays | No. Quartets |
|--------|--------------|
| 10K | 14 |
| 100K | 10 |
| 500K | 6 |



From Xiao Y. presentation: Exploration and Analysis of Affymetrix SNP Arrays. Center for Bioinformatics & Molecular Biostatistics, UCSF Division of Biostatistics

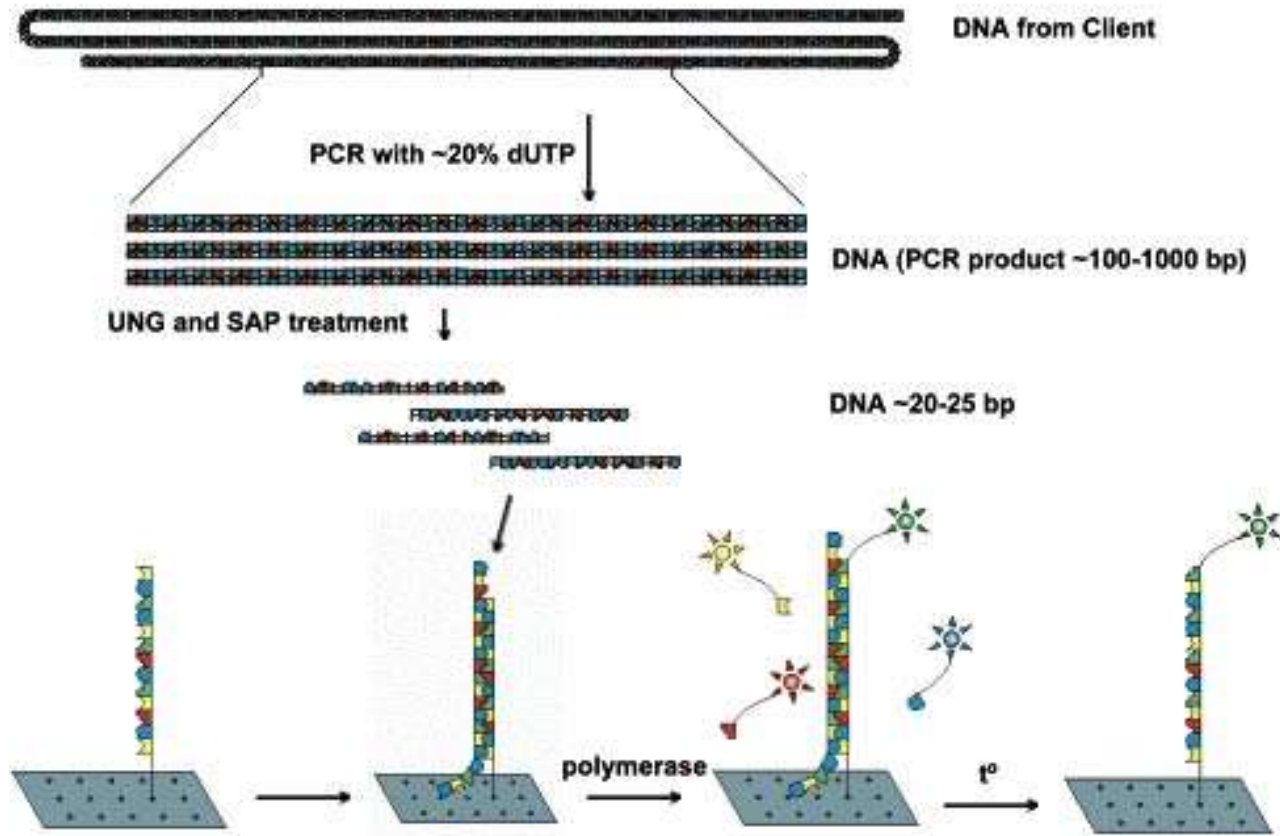
SNP arrays x expression arrays



From Xiao Y. presentation: Exploration and Analysis of Affymetrix SNP Arrays. Center for Bioinformatics & Molecular Biostatistics, UCSF Division of Biostatistics

SNP Arrays - APEX technologie

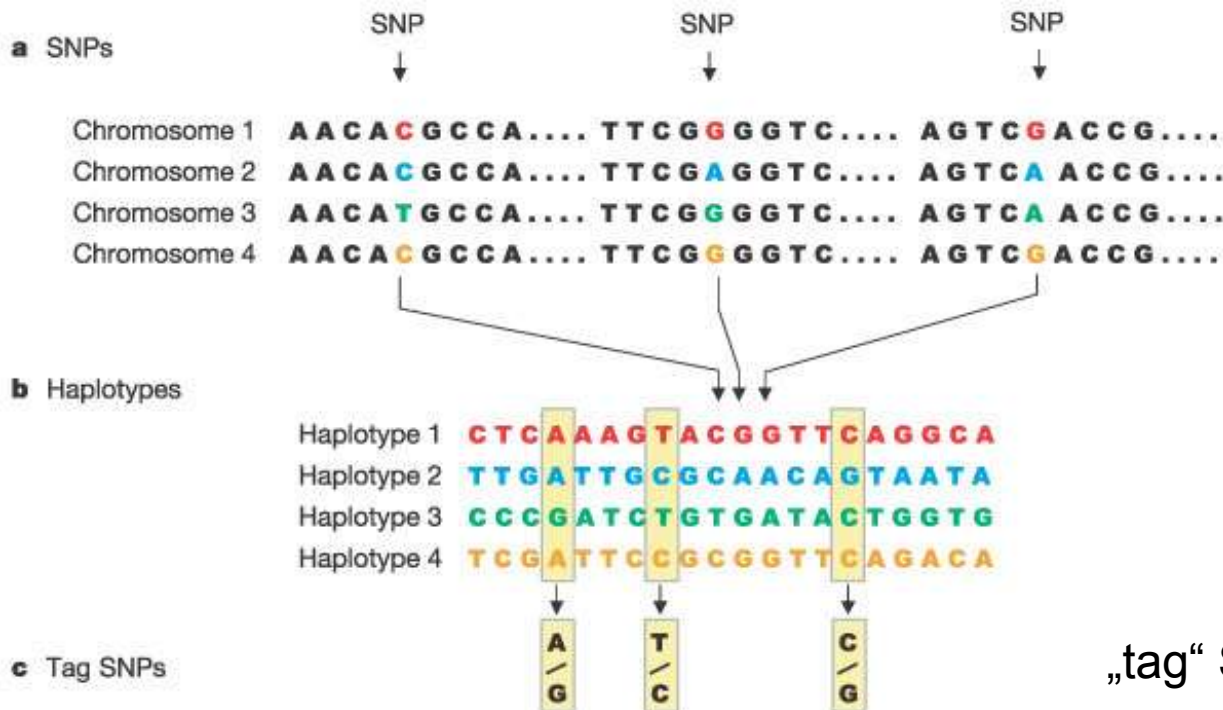
APEX = Arrayed Primer Extension



Kurg A. et al., Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet Test* 4:1-7, 2000.

Velké studie SNP

HapMap projekt: mezinárodní projekt, jehož cílem je identifikovat a katalogizovat SNPs v lidské populaci a vybrat z nich „tag“ SNPs, kterými se skupiny lidí odlišují



SNPs, které jsou na DNA blízko sebe se také společně dědí a určují **haplotyp** dané skupiny lidí

„tag“ SNPs odlišují dané haplotypy

HapMap projekt

<http://www.hapmap.org/index.html.en>

HapMap kolekce lidské DNA 270 vzorků DNA

| | | |
|-----------|----------|--------------------------------------|
| populace: | Nigerie | 30 trojic vzorků (matka, otec, dítě) |
| | Japonsko | 45 nepříbuzných vzorků |
| | Čína | 45 nepříbuzných vzorků |
| | USA | 30 trojic vzorků (matka, otec, dítě) |

The International HapMap Consortium. **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 449, 851-861. 2007.

The International HapMap Consortium. **A Haplotype Map of the Human Genome.** *Nature* 437, 1299-1320. 2005.

Velké studie SNP

3000 zdravých jedinců

| | | |
|---------------|-------------------------|----------|
| 2000 pacientů | bipolar disorder | (1 SNP) |
| ” | coronary artery disease | (1 SNP) |
| ” | Crohn’s disease | (9 SNPs) |
| ” | hypertension | |
| ” | rheumatoid arthritis | (3 SNPs) |
| ” | type 1 diabetes | (1 SNP) |
| ” | type 2 diabetes | (3 SNPs) |

P value < 5×10^{-7}

Studovali 500 000 SNPs pomocí Affymetrix microarrays

Wellcome Trust Case control Consortium. **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** Nature. 2007 Jun 7;447(7145):661-78.

Odchytky od referenčního genomu větší než 1kb

ještě v roce 2003 se myslelo, že většina „zdravých“ lidí se od referenčního genomu liší velmi nepatrně (SNPs, mikrosatelity)

array komparativní genomická hybridizace odhalila mnoho větších oblastí DNA, které se u zdravých lidí vyskytují v různém počtu

Copy number variation

DNA segment (většinou větší než 1 kb), který se u daného jedince vyskytuje v jiném počtu kopií než v referenčním lidském genomu

existuje mnoho takových oblastí v genomu (řádově tisíce)

“Database of Genomic Variants”

<http://projects.tcag.ca/variation/>

Copy number polymorphism – výskyt u více než 1% jedinců dané populace

Využití HapMap kolekce ke studiu copy number variant
všichni jedinci v této kolekci byli zdraví, přesto se našlo velké množství oblastí DNA (12% genomu), které se u těchto lidí nacházejí v různém počtu kopií

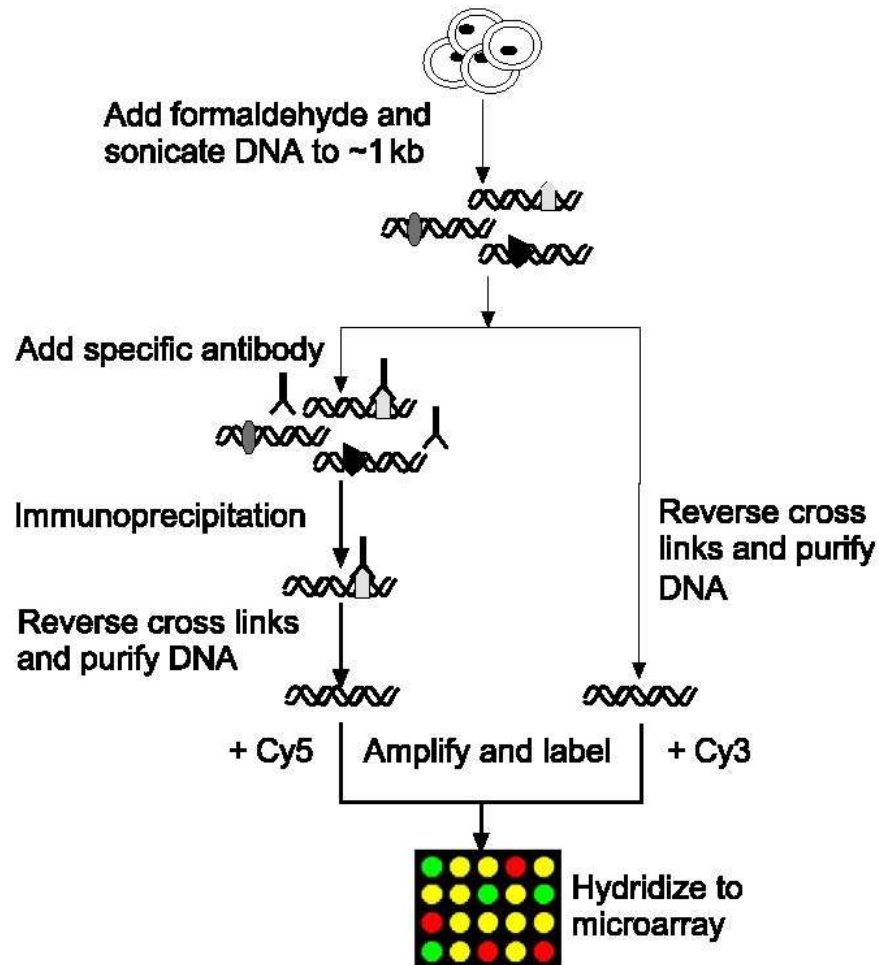
Copy number variation

hledání fenotypových projevů CNV („neškodná“ genomová varianta nebo příčina nemoci???)

CNV: pathogenic x benign x unknown clinical significance

vnáší „zmatek“ do experimentů, které např. hledají příčinu vrozených genetických poruch (mentální opožděnost, vývojové odchylky)

Chromatin ImmunoPrecipitation on chip ChIP-Chip



Nalezení vazebného místa

256 kb oblast 1p32 pokrytá překrývajícími se PCR produkty (~400 bp)

protilátka: trimethylace histonu H3 Lys4

