

## Rozdíly v genomové DNA v lidské populaci

Pavla Gajdušková  
Molekulární fyziologie genomu, 15. dubna 2008

## Referenční lidský genom

Zdroj DNA:

60% sekvencí pochází ze sekvenování DNA od jednoho dárce (sekvenování a sestavování BAC klonů)

některá místa genomu se nepodařilo touto prvotní metodou sekvenovat a postupně se doplňují nebo opravují chybně sestavená místa

dnešní sekvence lidského genomu: doplněna o sekvence z dalších 708 zdrojů

pokrývá většinu euchromatických oblastí  
stále obsahuje 302 nedosekvenovaných mezer  
..... a pravděpodobně i mnoho dalších chyb

## Referenční lidský genom

Sekvence lidského genomu zveřejněná na internetu

NCBI US National Center for Biotechnology Information

Aktualizace: „Build“ v současnosti Build 36.1  
<http://www.ncbi.nlm.nih.gov>

Další databáze: UCSC Genome Browser (March 2006 freeze)  
<http://genome.ucsc.edu>

Ensembl (hg18)  
<http://www.ensembl.org>

## Odchytky od referenčního genomu

jednonukleotidové změny  
(bodové mutace, indels, SNPs)

2 – 1000 bp  
(mikrosatelity, minisatelity, indels, inverze)

1kb až submikroskopické  
(copy number variants, segmentální duplikace, inverze, translokace)

mikroskopické až subchromozomální  
(chromozomální delece, inverze, intrachromozomální translokace)

celochromozomové až celogenomové  
(intrachromozomální translokace, kruhové chromozomy, isochromozomy, marker chromozomy, aneuploidie, aneuzomie)

## 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

## Metody detekce SNPs

RFLP - restriction fragment length polymorphism

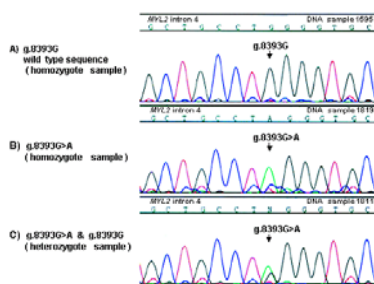
TaqMan analysis, FRET sondy

Sekvenování

APEX = arrayed primer extension

SNP microarrays

## Sekvenování



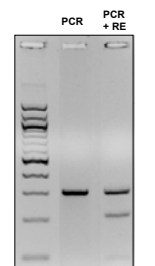
## RFLP – restriction fragment length polymorphism

PCR produkt – SNP místo leží někde uvnitř

sekvence jedné varianty se **štěpí** pomocí RE

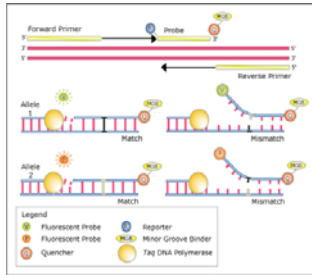
sekvence druhé varianty se **neštěpí** pomocí stejného RE

elektroforéza PCR produktu před a po štěpení RE



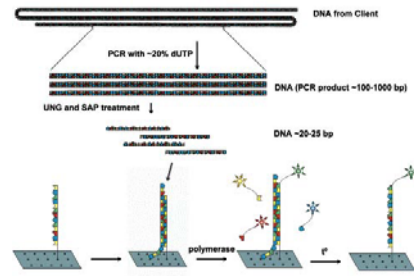
## SNPs - TaqMan analysis

modifikace real-time PCR  
2 primery + 2 TaqMan sondy (obsahují sekvenci obou variant SNP)



## APEX technologie

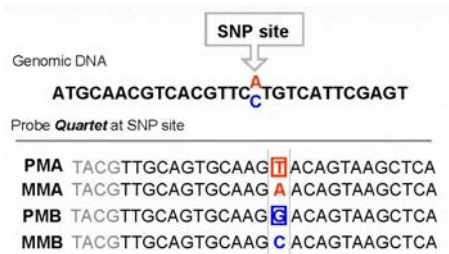
APEX = Arrayed Primer Extension



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

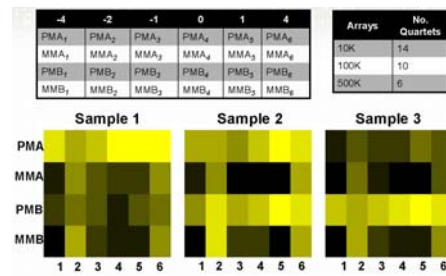
## SNP Arrays – probe design

sondy: 25 nukleotidů dlouhé, fotolitografický tisk (Affymetrix)



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

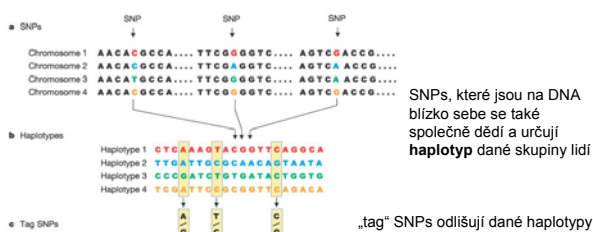
## SNP Arrays – probe design



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

## 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



## 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 nepřibuzných vzorků
	Čína	45 nepřibuzných 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 < 5x10 <sup>-7</sup>

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.

## Mikrosatelity a minisatelity

mikrosatelity: 1-4 nukleotidy, 10-100x opakované  
používají se ke studiu příbuznosti, populační studie  
časté mutace - způsobené chybami polymerázy při replikaci DNA (slippage)  
markery nestability genomu

minisatelity: 10-100 nukleotidů, 1000x na různých místech genomu  
GC-bohaté  
puriny častější na jednom vláknu, pyrimidiny na druhém  
patří mezi nejvíce nestabilní oblasti genomu  
využití pro fingerprinting

## Odchyly 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

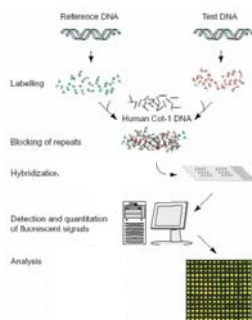
## Array komparativní genomická hybridizace (aCGH)

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

## Array komparativní genomická hybridizace (aCGH)



## Typy sond natištěných na microarray sklíčku

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

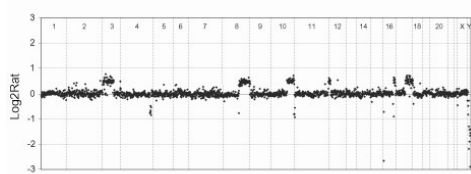
**Oligonukleotidy** 25 – 80 párů dlouhé oligonukleotidy  
mohou pokrývat i celý genom (repetitivní  
sekvence jsou vynechány)

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

## 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

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