



LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERSITY
Interní hematoonkologická klinika LF MU a FN Brno
Centrum molekulární biologie a genové terapie



Moderní metody analýzy genomu

Aplikace

25.11. 2011

Boris Tichý



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována
Evropským sociálním fondem
a státním rozpočtem České republiky

Aplikace nových technologií

Analýza DNA

Celogenomový screening

Sekvence

SNP

Strukturní aberace, početní aberace

Cílený screening

Sekvence

SNP

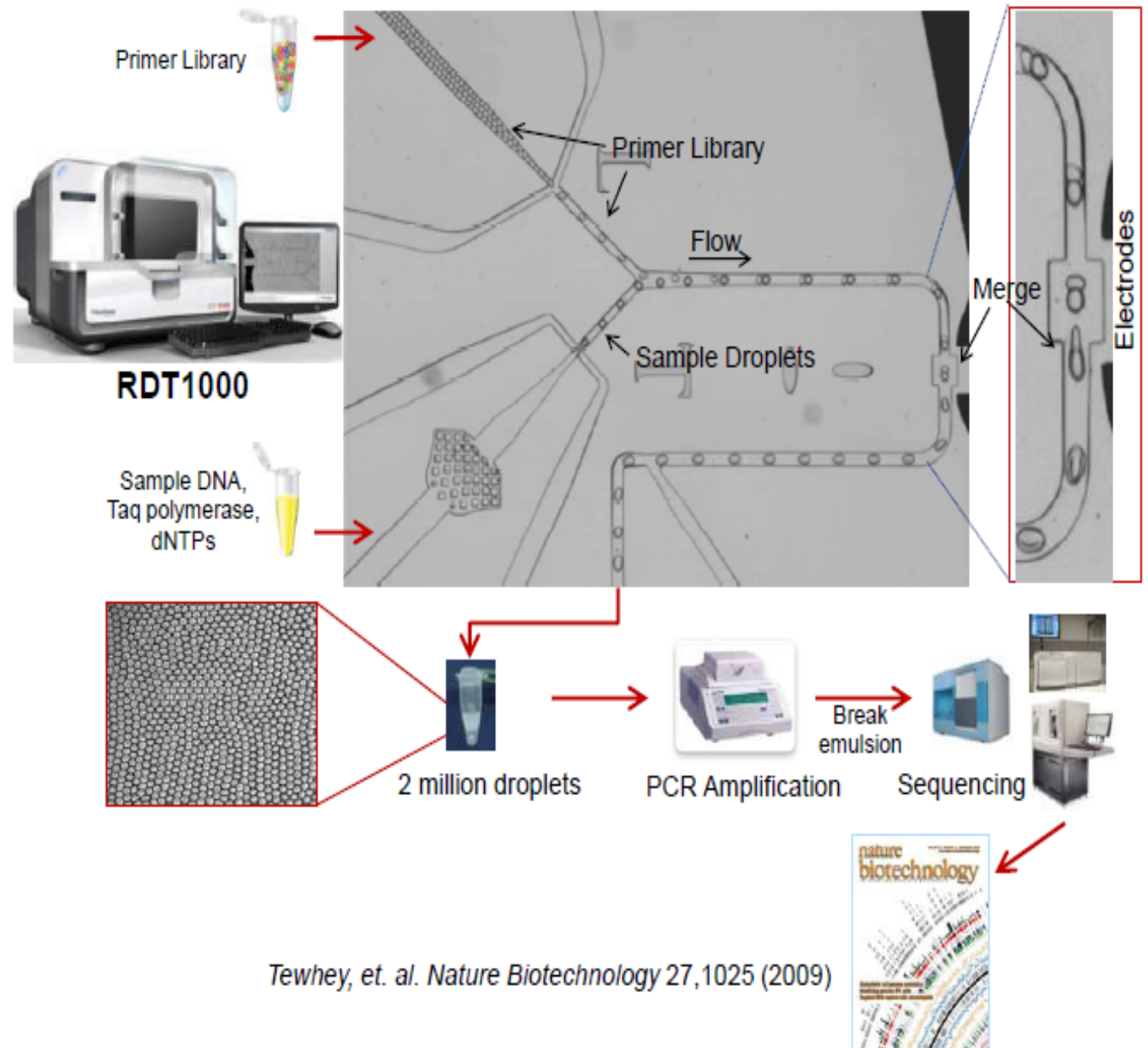
Cílený screening

Target enrichment

Hybridizace
na čipu
v roztoku

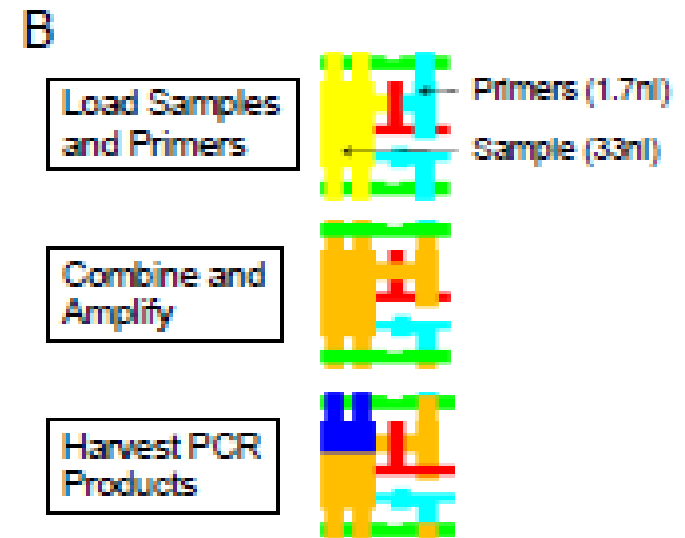
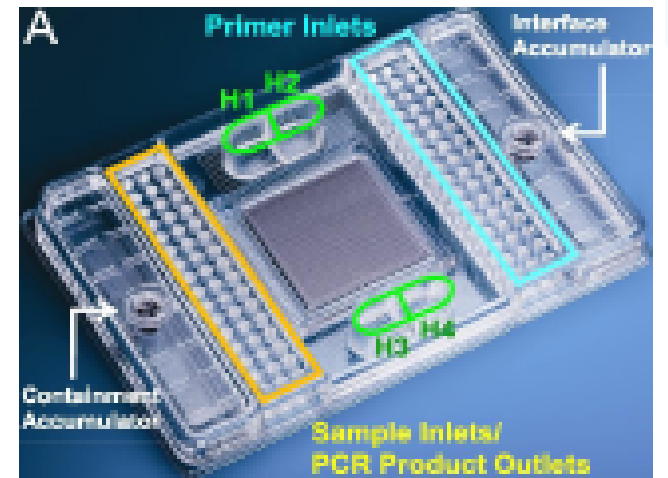
PCR
běžná
mikrofluidní

Targeted Sequencing Workflow Using the RDT1000



Tewhey, et. al. Nature Biotechnology 27,1025 (2009)

Figure 1: The Access Array System



Exome sequencing

- Všechny exprimované geny
- Většinou včetně nekódujících
- Hybridizace (v roztoku)

Gene enrichment

- Jeden gen – např. dědičné poruchy
 - PCR, hybridizace
 - multiplexing
- Skupiny genů – např. multifaktoriální nemoci, nádory
 - PCR, hybridizace
- Úseky genomu – strukturní aberace
 - hybridizace

Deep sequencing

Germinální mutace/SNP

Heterozygot – 1:1

Mozaika – variabilní

Somatické

Různý poměr zdravých/postižených buněk

Jednotlivé subpopulace postižených buněk (např. nádorových)

=> Deep sequencing

Sekvenování s vysokým pokrytím

! max. senzitivita dána chybovostí enzymů

CMBGT - plány

a) dědičné poruchy

Často velmi dlouhé geny

(např. Duchenova muskulární dystrofie – dystrofin – 79 exonů, 2,4Mb, 14kb mRNA)

=> >150 sekvencí (forward + reverse)

=> 150*3€/pacient

GS Junior

75.000 reads => 500 reads/amplikon => 4-5 vzorků/běh

1.500€/běh

MiSeq

Kapacita ~1 mld. bazí =~ 3 mil. reads

Dystrofin – 14.000 bazí => ~150 reads po 100 bazích

=> 100x pokrytí → 15.000 reads

=> 200 vzorků/běh

1-2.000€/běh + náklady na enrichment !(150 – 1.000€/vzorek)!

Více genů – stejná situace

CMBGT - plány

b) leukemie/lymfomy

“kontaminace” vzorku nenádorovými (zdravými) buňkami

klonální evoluce nádoru

somatické mutace

=> hledáme metodiku pro *de novo* detekci mutací s citlivostí pod 1% buněk s aberací

=> resequencing microarray

=> deep sequencing

CMBGT - plány

b) leukemie/lymfomy

resequencing microarray

krátké oligonukleotidy (25 bazí)

udávaná citlivost až 1 %

9 genů, 120 miRNA (~ 40 kb sekvence)

110 long-range PCR

potřeba analyzovat desítky vzorků ← optimalizace algoritmu

neanalyzovatelné úseky (GC, homopolymery,...)

variabilní sezitivita

falešné pozitivní/negativní nálezy – problematické ověření

CMBGT - plány

b) leukemie/lymfomy

Deep sequencing

Sekvenování s vysokým pokrytím

Enrichment?

Hybridizace

PCR

Long-range PCR, multiplex PCR

Vyhodnocení

Vhodný SW

Statistický model? - rozdíly tranzice/transverze, poziční rozdíly

RNA-Seq

Transcriptome sequencing

Single-end – kvantifikace

Paired-end – struktura transkriptů

Tag sequencing

3' tagy, kvantifikace, bez informace o struktuře

Degradome sequencing

5' tagy, identifikace cílů microRNA

Small RNA sequencing

MicroRNA kvantifikace i de-novo identifikace

SAGE (Serial Analysis of Gene Expression)

Konkatemery tagů, původně Sanger sekvenace

RIP (RNA ImmunoPrecipitation)

Imunoprecipitace, RNA vázající proteiny

RNA-Seq

Oligo-dT vs. hexamer priming reverzní transkripce

3', 5' end reprezentace

PolyA selekce vs. rRNA deplece

Zastoupení non-polyA (non-coding) RNA

Vysoce abundantní geny

Až 75% transkriptů ← 5% exprimovaných genů

Možnost detekce fúzních genů

Speciální algoritmy

Mapování readů na exon-exon rozhraní

Aplikace nových technologií

Epigenomika/epigenetika

In biology, and specifically genetics, epigenetics is the study of heritable changes in phenotype (appearance) or gene expression caused by mechanisms other than changes in the underlying DNA sequence, hence the name epi- (Greek: επί- over, above) -genetics. These changes may remain through cell divisions for the remainder of the cell's life and may also last for multiple generations. However, there is no change in the underlying DNA sequence of the organism;^[1] instead, non-genetic factors cause the organism's genes to behave (or "express themselves") differently.

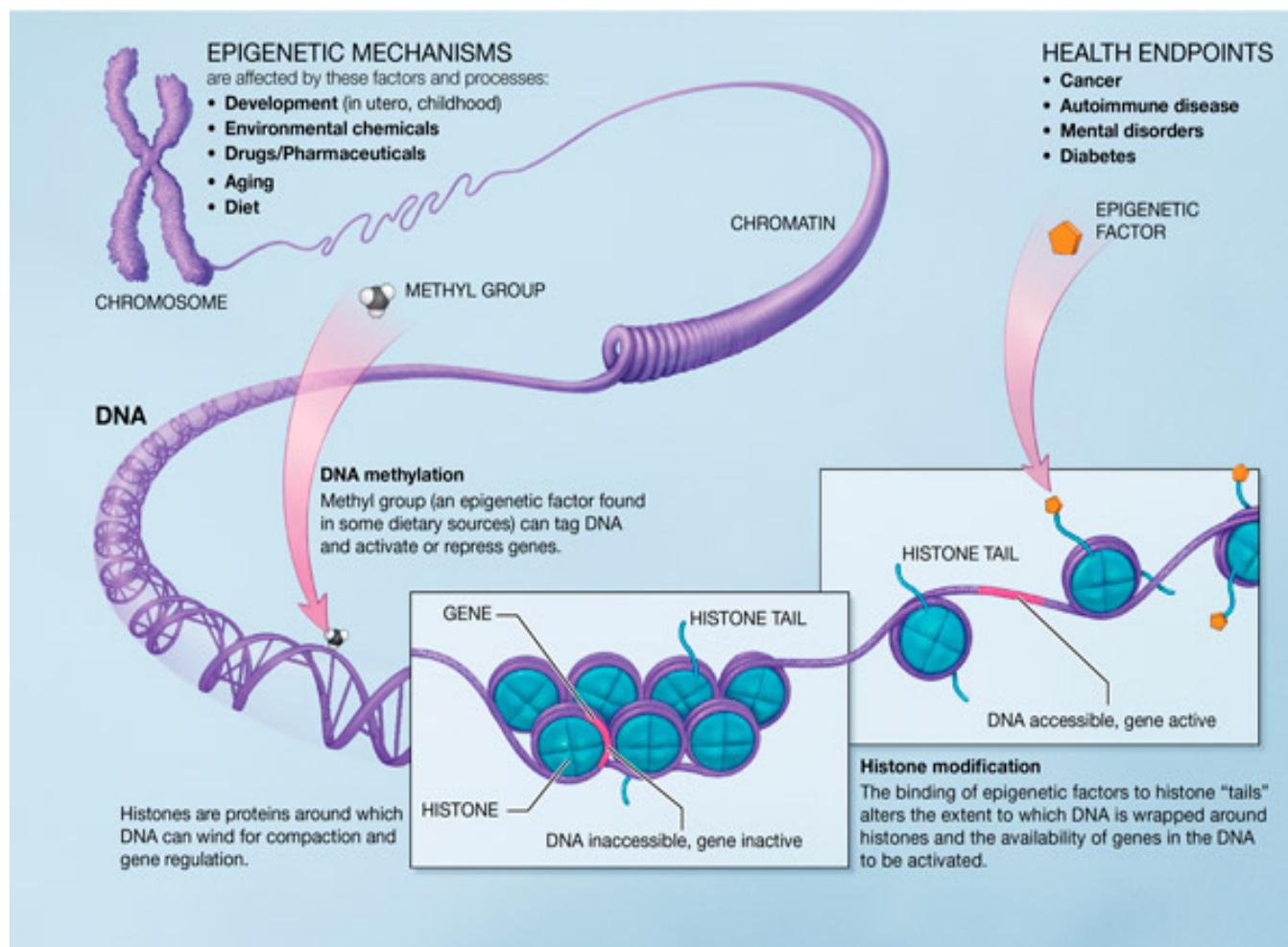
Epigenetika

DNA metylace

C → Met-C, snížená exprese

Modifikace histonů

Aktivní I neaktivní chromatin



Chromatinová imunoprecipitace

Modifikované histony

Acetylované, metylované

Další DNA vázající proteiny

Transkripční faktory

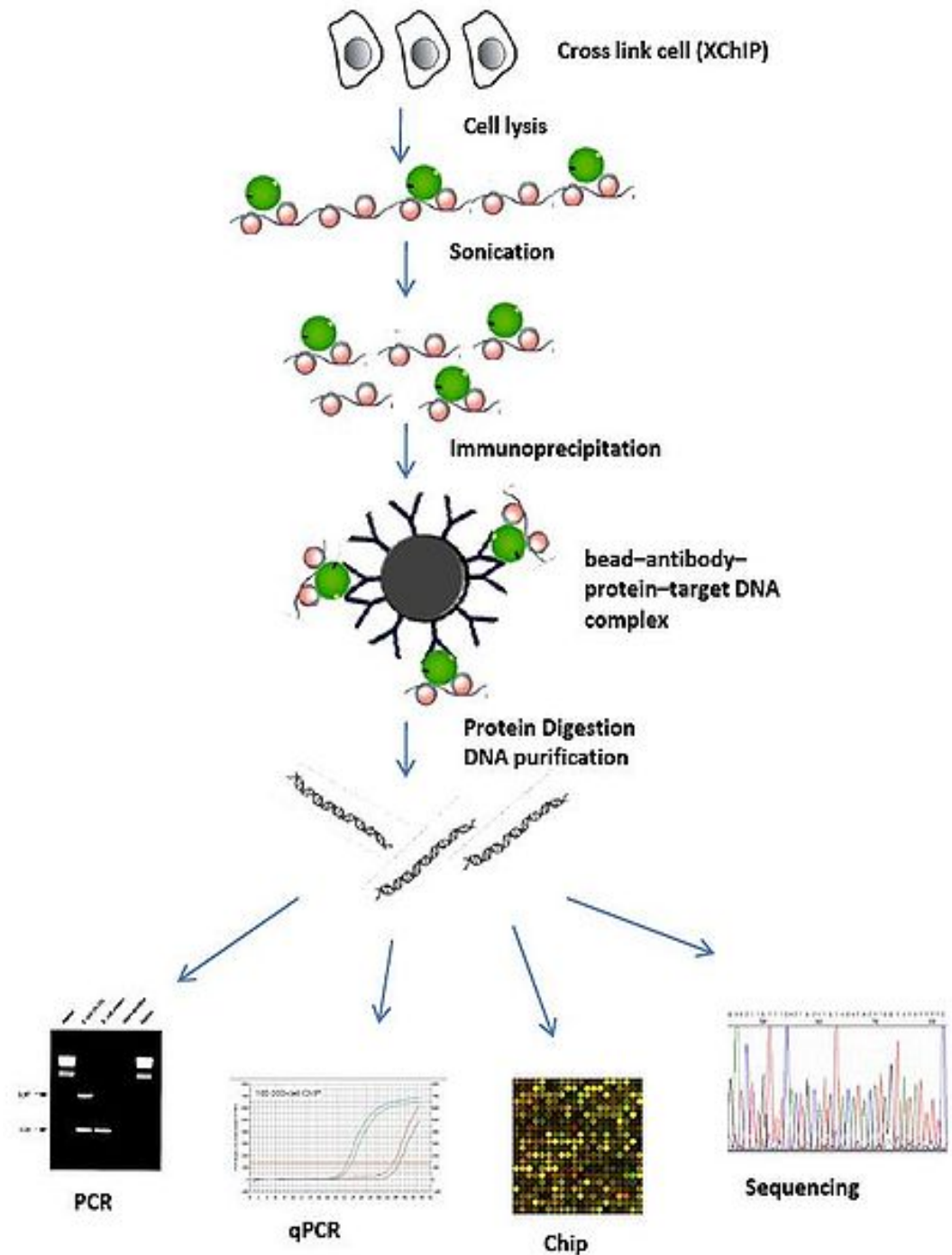
RNA polymerázy

Metylovaná DNA

Kvalita protilátky

Změna epitopu

formaldehyd



Metylace DNA

MeDIP - Immunoprecipitace metylované DNA

Protilátka rozpoznávající Met-C

Pozice metylovaných úseků DNA

Bisulfite treatment

Konverze C \rightarrow U, Met-C se nemění

Přesná identifikace jednotlivých metylovaných bazí

Některé dědičné choroby, nádory

Metagenomika

Sekvenace mikrobiálních populací

Informace I o nekultivovatelných organismech

Identifikace nových genů, nových vlastností

Půdní, vodní, střevní bakterie apod.

Analýza NGS dat

Assembling – vytvoření kontigů

De-novo

Mapování na referenční sekvenci

Identifikace variant

SNP (SNV)

In/del

Strukturní aberace – chromosomy, transkripty

Kvantifikace

DNA – amplifikace/delece, místa vazby transkripčních faktorů

RNA – tagy, exony, transkripty

Analýza NGS dat

Integrace dat

Databáze

Protein-protein interakce

Ontologie – GO (GeneOntology), MeSH terms

Transkripční faktory, microRNA targets

Expresní profily – GSEA, profily chemických látek

Různé typy dat

ChIP + RNA + metylace + DNA copy number

Analýza NGS dat

Statistická analýza

“klasické” statistické metody

T-test, ANOVA, neparametrické testy

Vícerozměrné analýzy

Clustering (HCL, k-means), PCA

Klasifikační algoritmy

Lineární (LDA), nelineární (KNN)

Vazba s dalšími informacemi (analýza anotací)

Over-reprezentace

Sítě

Aplikace nových technologií

Analýza RNA

Protein kódující RNA

Nekódující RNA – miRNA, snoRNA, lincRNA

Počet = digitální genová exprese

Sekvence, struktura – mutace, sestřihové varianty

Microarrays vs. NGS

Pro NGS není bezpodmínečně nutné znát sekvenci RNA (microRNA)

Expresní profily

Soubor exprese všech/vybraných genů v určité tkáni (orgán, nádor,...)

Porovnání vzorků navzájem

Posouzení vlivu externího faktoru

Reálné využití expresních profilů v medicíně

Omezené

Technické i interpretační problémy

Breast Cancer

MammaPrint (70 genů)

OncotypeDX (21 genů) – už > 150.000 pacientek

Lymfomy

GC vs. ABC DLBCL

Využití výzkum

Integrace s dalšími metodami

Chromatinová imunoprecipitace (ChIP)

Bioinformatika (motivy, protein-protein interakce)

Interpretace dat

UP/DOWN geny

Asociace mezi geny (transkripční faktory, miRNA, protein-protein,...)

Příklad

Gen TP53 → protein p53

Transkripční faktor, buněčný cyklus, apoptóza, DNA repair

Otázka:

Rozdíl mezi wt a mut TP53