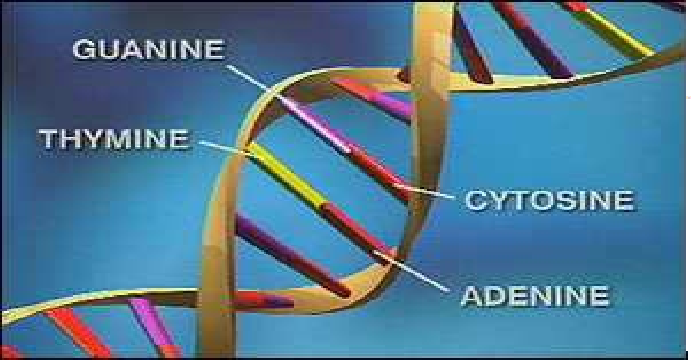
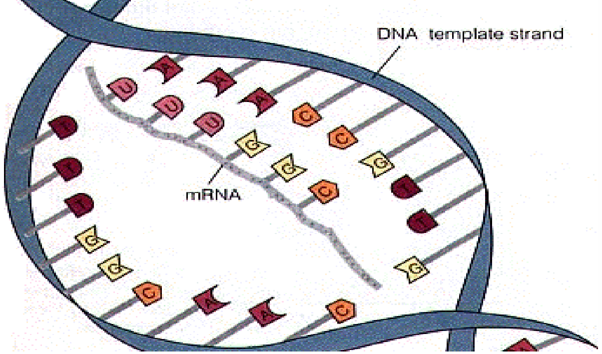
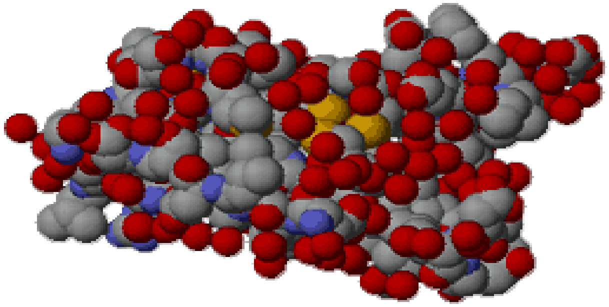
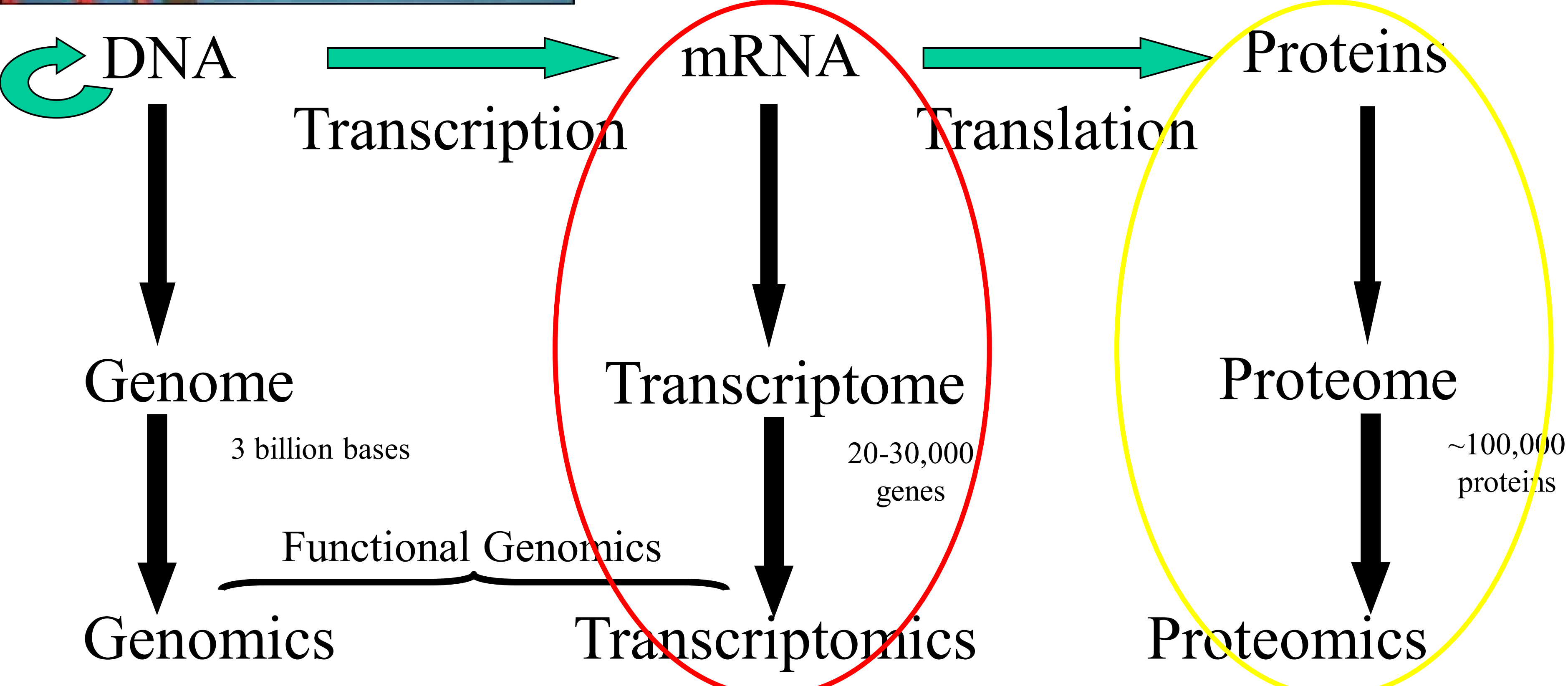
Analýza genové exprese









DNA

Transcription

mRNA

Translation

Proteins

Genome

3 billion bases

Transcriptome

20-30,000 genes

Proteome

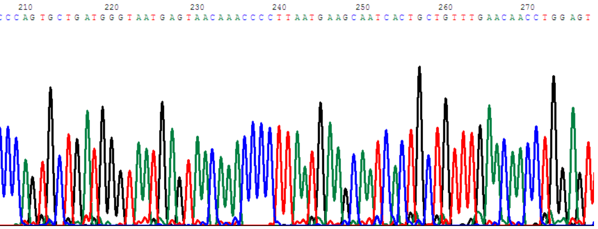
~100,000 proteins

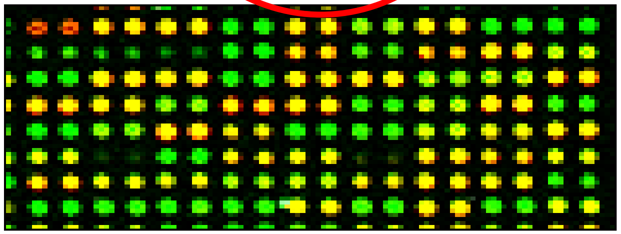
Functional Genomics

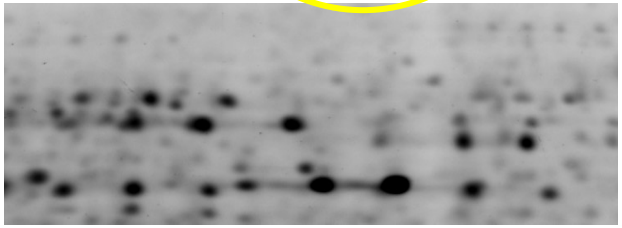
Genomics

Transcriptomics

Proteomics







Rozdíly v expresi genů mohou mít velký význam

- Today known that humans and chimpanzees are 98.7% identical in their genome sequences and that large differences tissue-specific differences exist in expression – particular in the brain (Enard et al. 2002)

Analýza genové exprese

(1) Kvantifikace kandidátních transkriptů

(2) Microarrays

(3) RNAseq (transkriptom)

Kandidátní geny - relativní kvantifikace pomocí standardů

- Měření úrovně exprese (např. v různých typech tkání nebo treatement vs. non-treatement atd.
- mRNA → reverzní transkripce → cDNA → PCR
- **housekeeping geny** – slouží jako standard pro měření
- stejný počet kopií ve všech buňkách
- exprimované ve všech buňkách, nezávislé na experimentu

Srovnání množství PCR produktu na elektroforéze

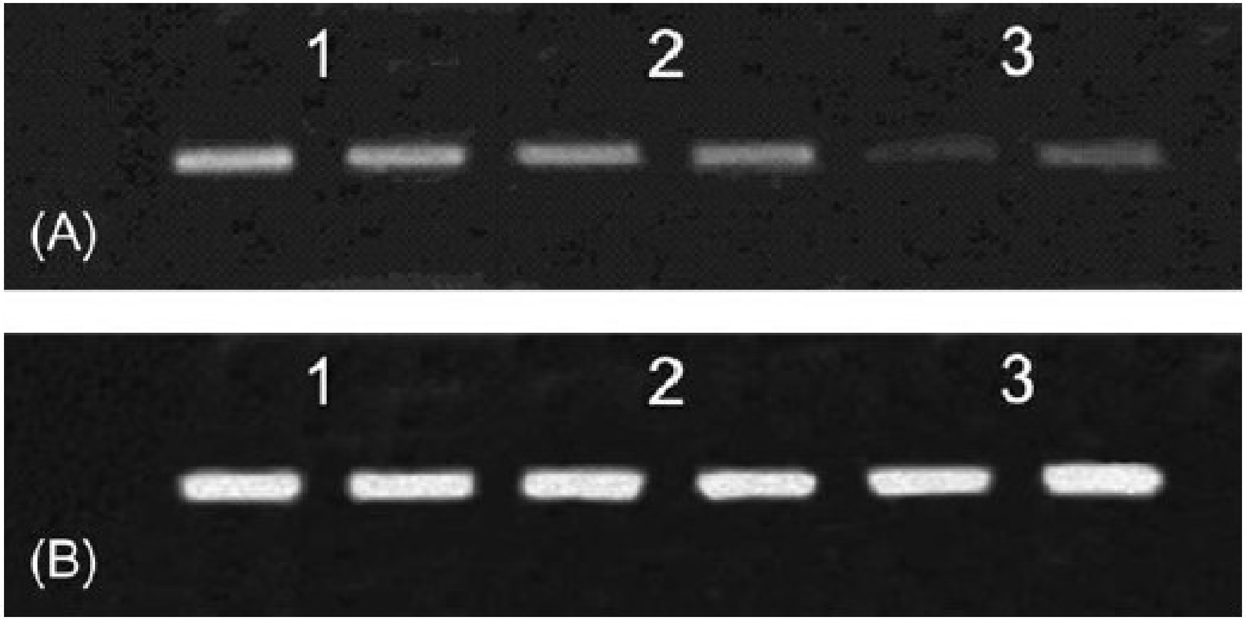
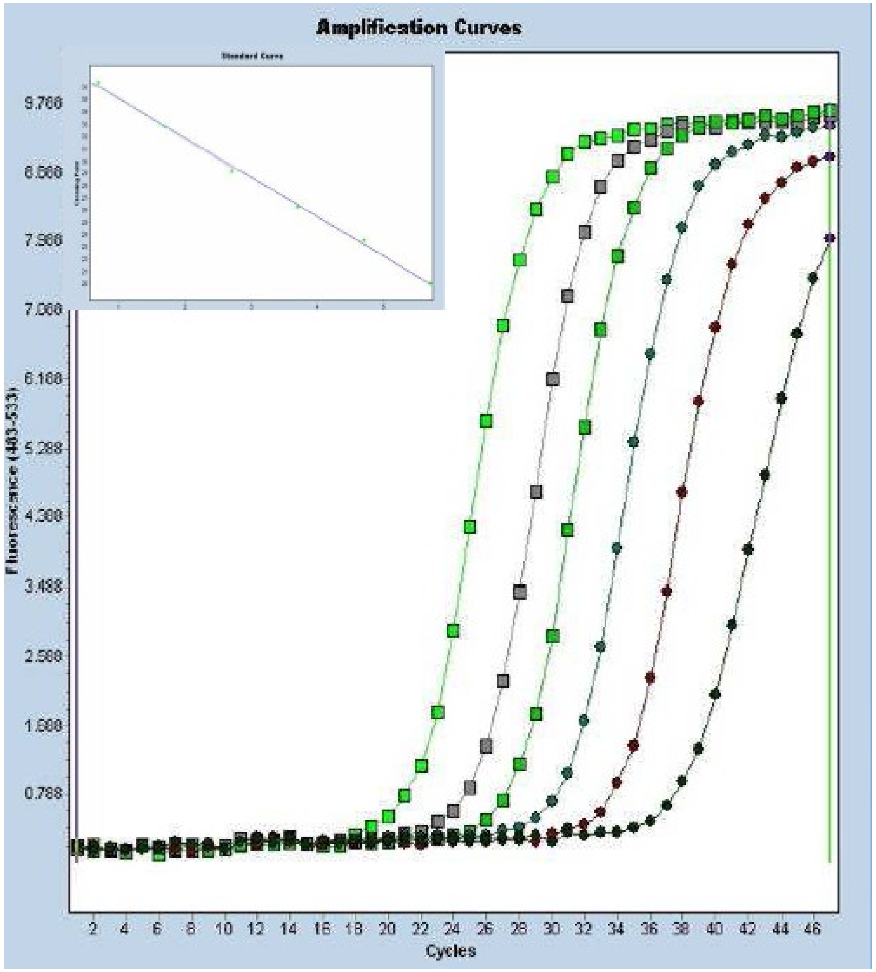
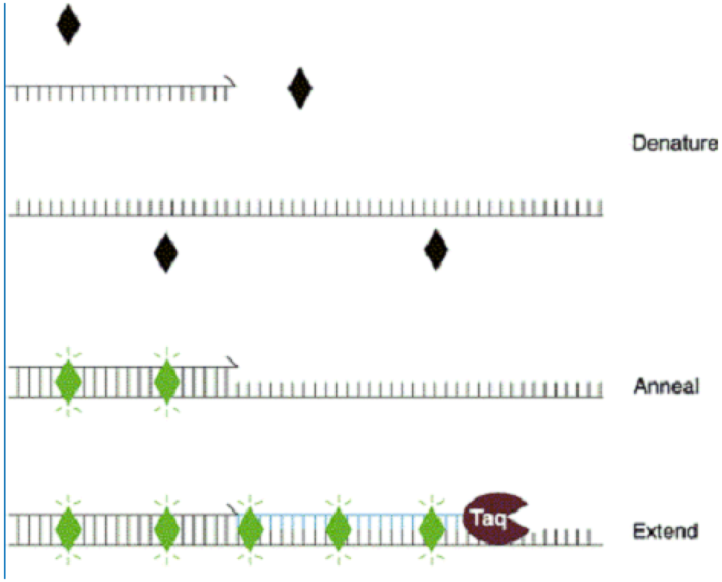


Figure 6 Lipopolysaccharide-induced expression of tgTLR4 in peritoneal macrophages (A) in comparison with 18S RNA expression (B): 1, stimulated both *in vivo* and *in vitro*; 2, stimulated only *in vitro* and 3, unstimulated.

Studium kandidátních genů - qPCR



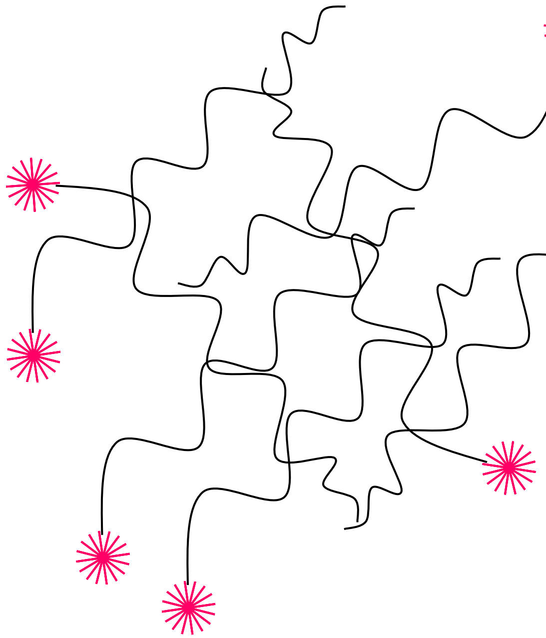


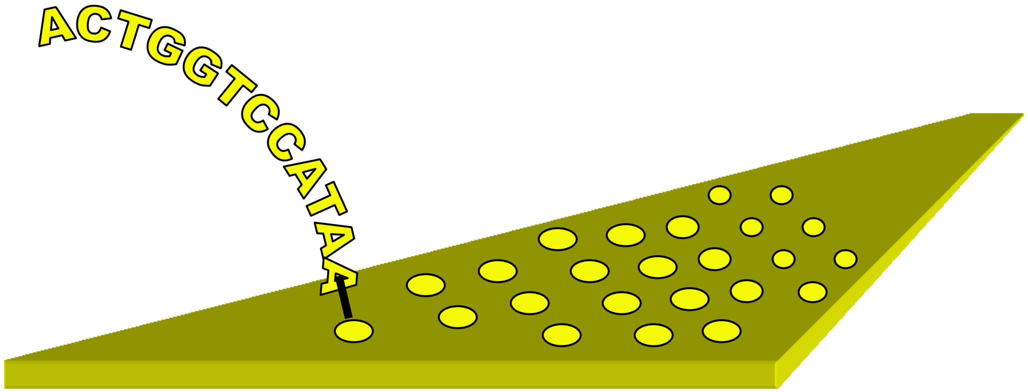
- většinou nespecifická detekce (SYBR Green, aj.)
- nutné srovnání s house-keeping geny

https://www.youtube.com/watch?v=y8tHiH0BzGY

Microarray transcriptomics

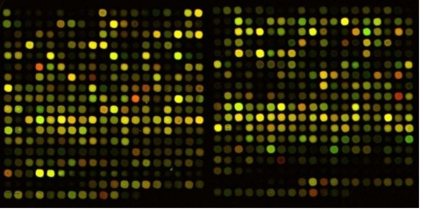
Microarray analysis of transcriptome (~ specific DNA hybridization)





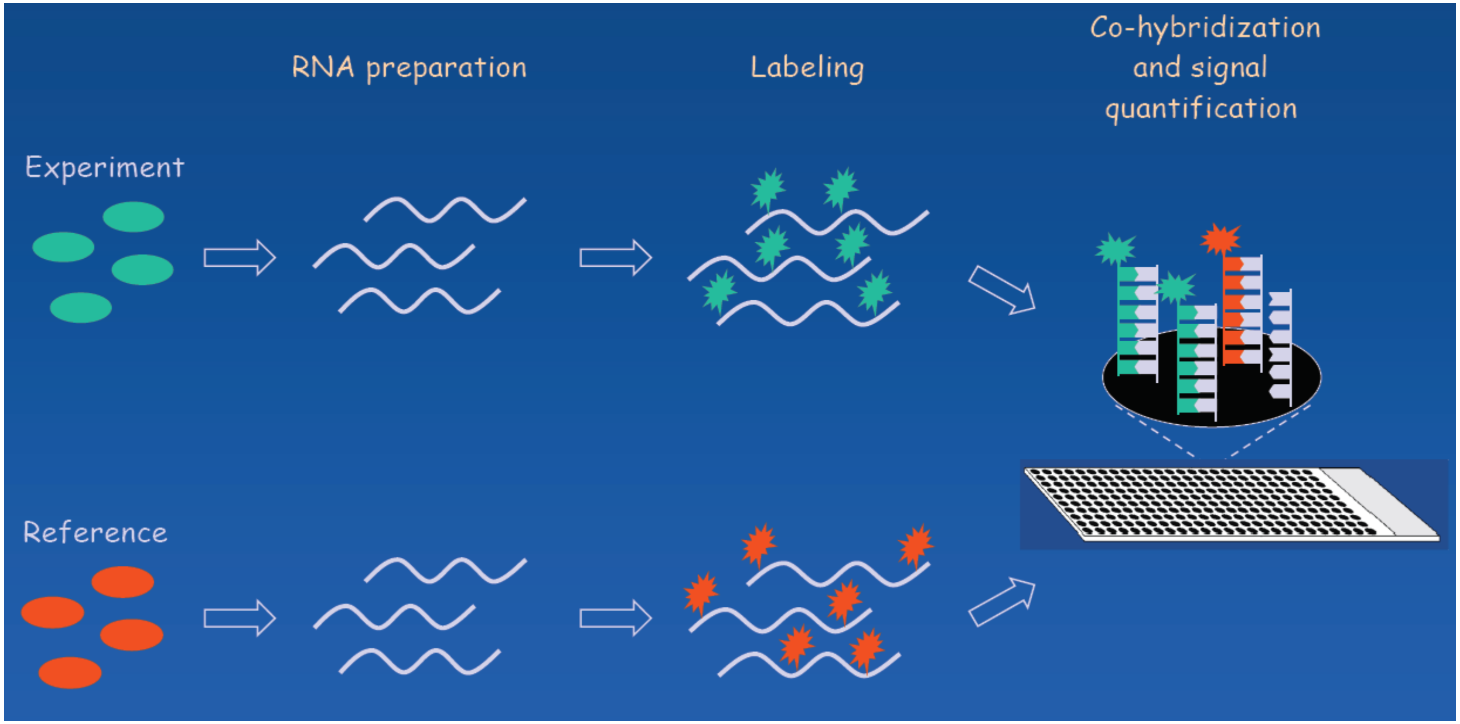
Target (i.e. mix of transcripts in a form of cDNA = mRNA přepsaná do DNA reverzní transkriptázou, tj. neobsahuje introny)

Probe
(i.e. synthesized oligonucleotides complementary to particular genes)

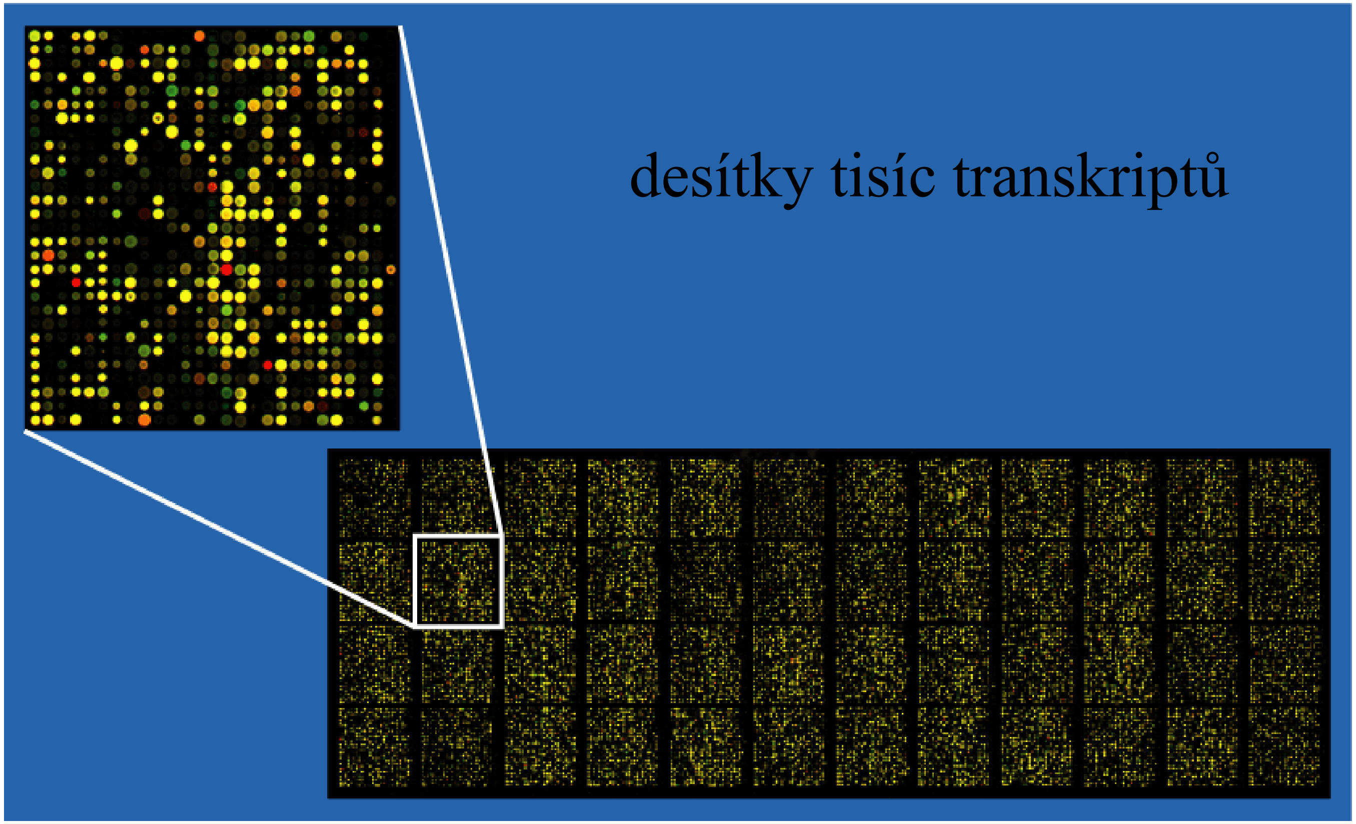


Sledování exprese genů (microarrays)

- Sledování exprese mnoha (tisíce) genů najednou
- Založeno na hybridizaci
- Sleduje se rozdíl vůči kontrole ("heterologous hybridization") = dvoukanálový experiment



Vyhodnocení chipu – analýza obrazu
(srovnání úrovně exprese mezi kontrolou a experimentem)



desítky tisíc transkriptů

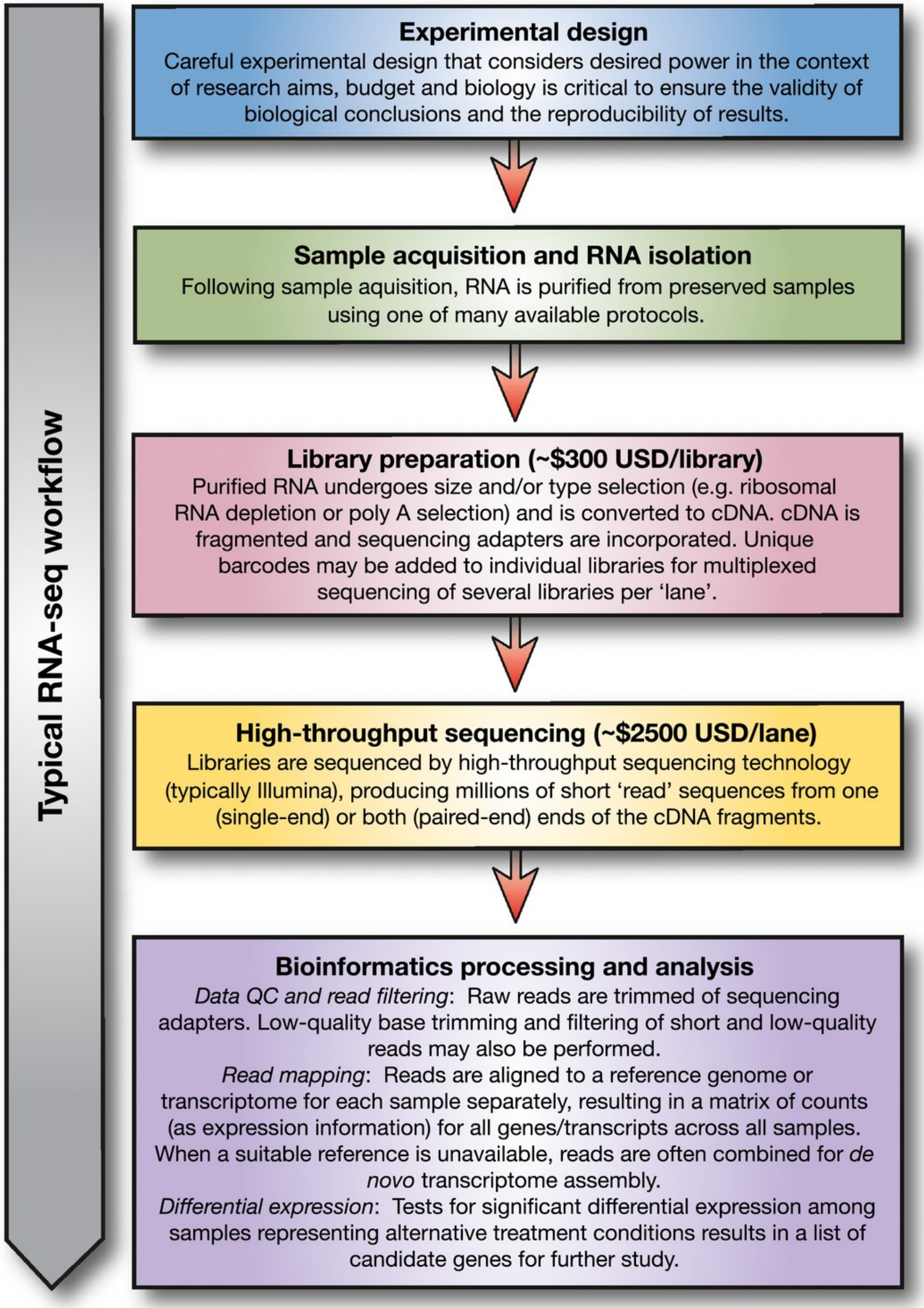
Komerčně dostupné pro kompletní transkriptom cca 25 druhů
(další jsou rychle vyvíjeny, i na zakázku)

Commercial or custom microarrays

Agilent	Affymetrix	Custom	
Arabidopsis thaliana	*A. thaliana*	*Acropora millepora*	*Locusta migratoria*
Bovinae spp. (bovine)	*Bovinae* spp. (bovine)	*Acropora palmata*	*Loxodonta africana*
Brassica spp.	*Caenorhabditis elegans*	*Ambrosia artemisiifolia*	*Megachile rotundata*
C. elegans	*Callitrichidae* spp. (Marmoset)	*Amphiura filiformis*	*Melitaea cinxia*
Canis spp. (canine)	*Canis* spp. (canine)	*Anemonia viridis*	*Meliteaea cinxia*
Culicidae spp. (mosquito)	*Danio rerio* (Zebrafish)	*Anopheles gambiae*	*Metarhizium robertsii*
D. rerio (Zebrafish)	*Drosophila melanogaster*	*Apis mellifera carnica*	*Montastraea faveolata*
D. melanogaster	*Equus ferus* (horse)	*Azospirillum brasilense*	*Mytilus californianus*
E. ferus (horse)	*Felidae* spp. (feline)	*Brugia malayi*	*Neotoma lepida*
Gallus gallus (chicken)	*G. gallus* (chicken)	*Calanus finmarchicus*	*Neurospora crassa*
Gossypium spp. (cotton)	*Glycine max* (soybean)	*Carpodacus mexicanus*	*Neurospora tetrasperma*
Homo sapiens	*H. sapiens*	*Coregonus cupleaformis*	*Onthophagus taurus*
Hordeum vulgare (barley)	*Macaca mulatta* (rhesus)	*Crassostrea gigas*	*Ostreococcus tauri*
Leporidae spp. (rabbit)	*Medicago* spp.	*Crassostrea virginica*	*Petrolisthes cinctipes*
M. mulatta (rhesus)	*Mus musculus*	*Daphnia magna*	*Picea glauca*
Magnaporthe spp.	*Oryza sativa* (rice)	*Daphnia pulex*	*Pimephales promelas*
Medicago spp.	*Ovis aries* (sheep)	*Dreissena polymorpha*	*Pinus pinaster*
M. musculus (mouse)	*Saccharomyces cerevisiae*	*Drosophila mojavensis*	*Pinus taeda*
Nicotiana tabacum (tobacco)	*Schizosaccharomyces pombe*	*Eucalyptus grandis*	*Platichthys flesus*
O. sativa (rice)	*Suidae* spp. (porcine)	*Folsomia candida*	*Ruditapes philippinarum*
O. aries (sheep)	*Taeniopygia guttata* (Zebra Finch)	*Fundulus heteroclitus*	*Ruditapes philippinarum*
Rattus norvegicus (rat)		*Gadus morhua*	*Salmo salar*
S. cerevisiae		*Gasterosteus aculeatus*	*Salvelinus fontinalis*
Salmonidae spp. (salmon)		*Helianthus annus*	*Schistocerca gregaria*
Solanum lycopersicum (tomato)		*Karenia brevis*	*S. lycopersicum*
Suidae spp. (porcine)		*Lagopus lagopus*	*Solenopsis invicta*
Triticum spp. (wheat)		*Lagopus lagopus scoticus*	*Tribolium castaneum*
Xenopus spp.		*Laternula elliptica*	*Tursiops truncatus*
		Lepeophtheirus salmonis	

RNA-seq

Differential gene expression (DE) studies in nonmodel species without the need for prior genomic resources



Experimental design
Careful experimental design that considers desired power in the context of research aims, budget and biology is critical to ensure the validity of biological conclusions and the reproducibility of results.

Sample acquisition and RNA isolation
Following sample aquisition, RNA is purified from preserved samples using one of many available protocols.

Library preparation (~$300 USD/library)
Purified RNA undergoes size and/or type selection (e.g. ribosomal RNA depletion or poly A selection) and is converted to cDNA. cDNA is fragmented and sequencing adapters are incorporated. Unique barcodes may be added to individual libraries for multiplexed sequencing of several libraries per 'lane'.

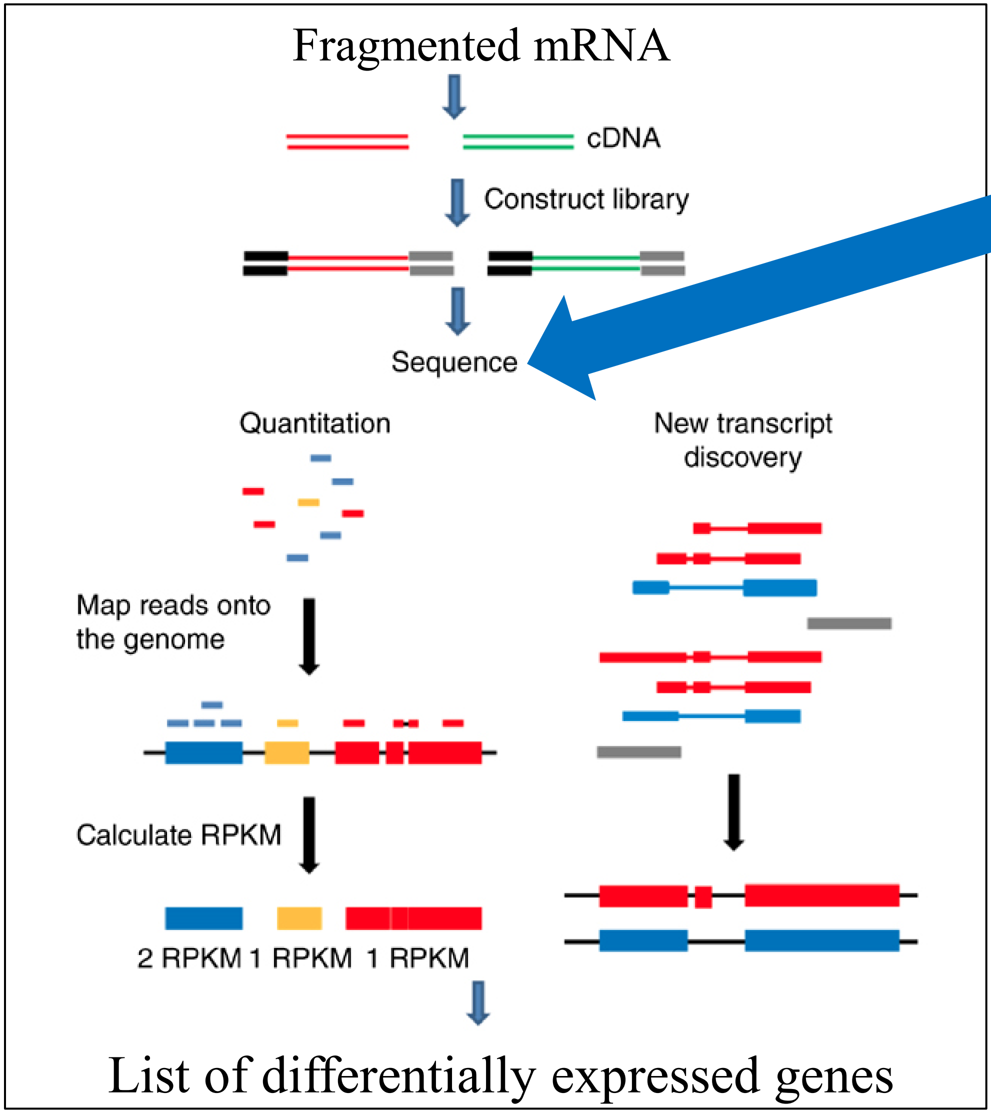
High-throughput sequencing (~$2500 USD/lane)
Libraries are sequenced by high-throughput sequencing technology (typically Illumina), producing millions of short 'read' sequences from one (single-end) or both (paired-end) ends of the cDNA fragments.

Bioinformatics processing and analysis
Data QC and read filtering: Raw reads are trimmed of sequencing adapters. Low-quality base trimming and filtering of short and low-quality reads may also be performed.
Read mapping: Reads are aligned to a reference genome or transcriptome for each sample separately, resulting in a matrix of counts (as expression information) for all genes/transcripts across all samples. When a suitable reference is unavailable, reads are often combined for *de novo* transcriptome assembly.
Differential expression: Tests for significant differential expression among samples representing alternative treatment conditions results in a list of candidate genes for further study.

Todd et al. 2016

RNA-Seq = sekvenování transkriptomu

RNA-Seq workflow for gene expression analysis


Fragmented mRNA

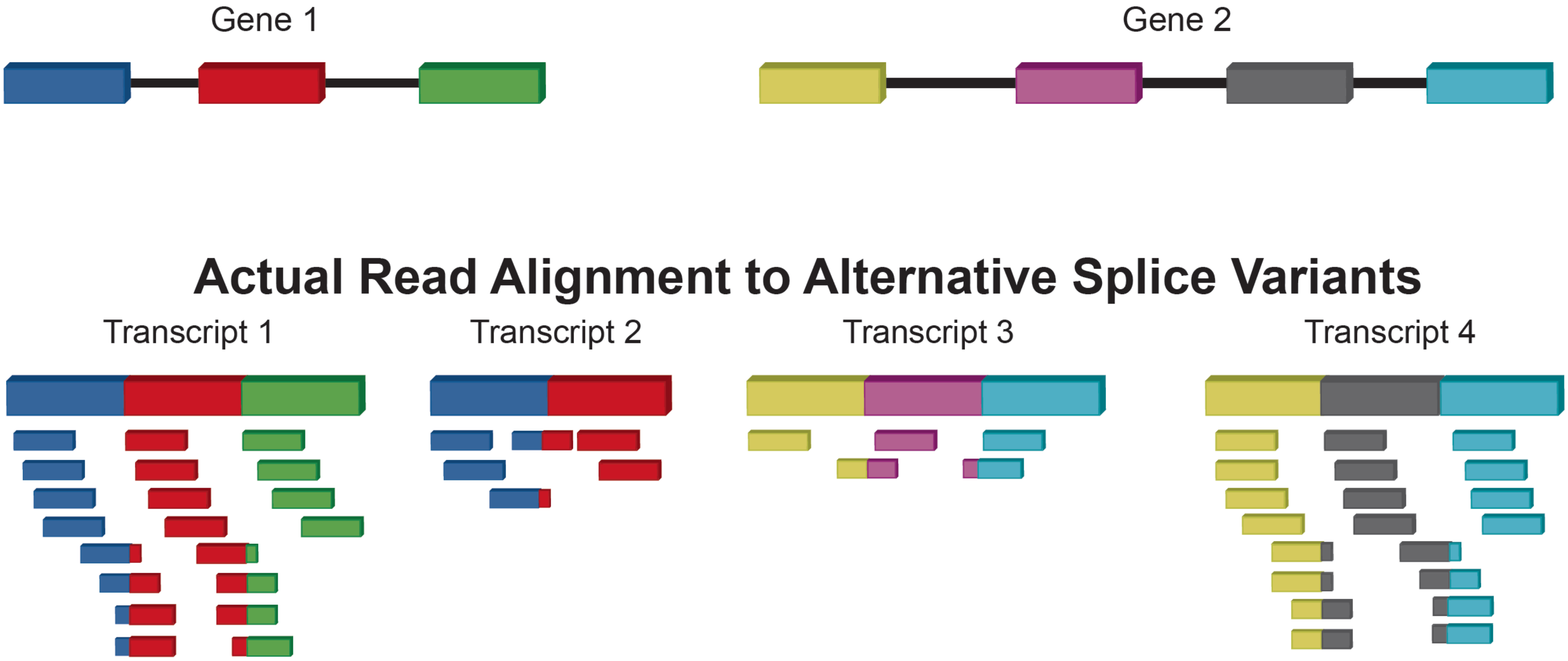
List of differentially expressed genes

Využití NGS

Adapted from Zeng and Mortazavi, *Nature Immunology* **13** (2012)

RNA-Seq quantification

(RPKM = reads per kilobase per million reads)



Gene 1

Gene 2

Actual Read Alignment to Alternative Splice Variants

Transcript 1

Transcript 2

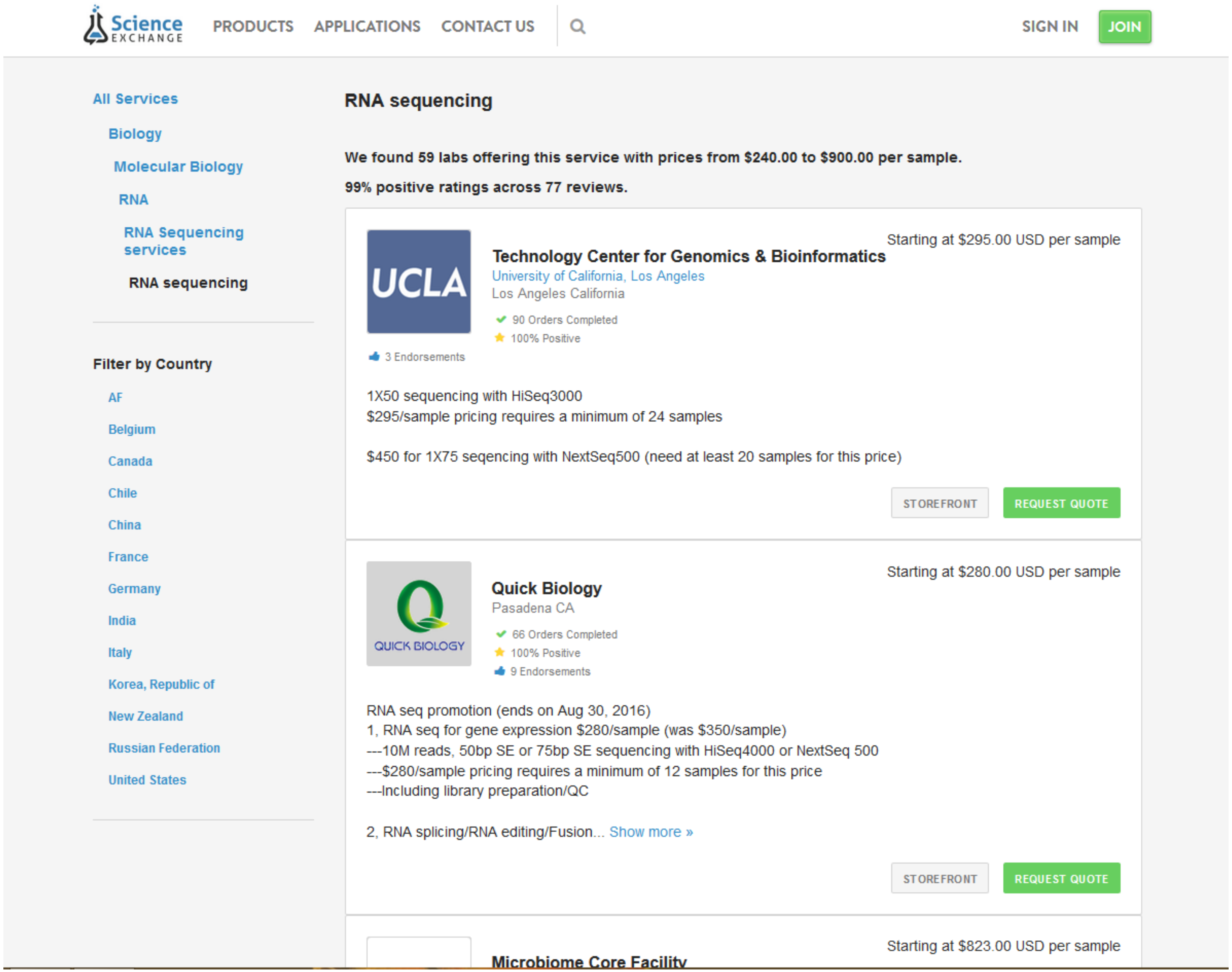
Transcript 3

Transcript 4

Musí být znám referenční transkriptom

RNA-seq commercially available

from 245 USD/sample



All Services
Biology
Molecular Biology
RNA
RNA Sequencing services
RNA sequencing

Filter by Country

- AF
- Belgium
- Canada
- Chile
- China
- France
- Germany
- India
- Italy
- Korea, Republic of
- New Zealand
- Russian Federation
- United States

RNA sequencing

We found 59 labs offering this service with prices from $240.00 to $900.00 per sample.

99% positive ratings across 77 reviews.

Starting at $295.00 USD per sample

Technology Center for Genomics & Bioinformatics
University of California, Los Angeles
Los Angeles California
✓ 90 Orders Completed
★ 100% Positive
3 Endorsements

1X50 sequencing with HiSeq3000
$295/sample pricing requires a minimum of 24 samples

$450 for 1X75 seqencing with NextSeq500 (need at least 20 samples for this price)

Starting at $280.00 USD per sample

Quick Biology
Pasadena CA
✓ 66 Orders Completed
★ 100% Positive
9 Endorsements

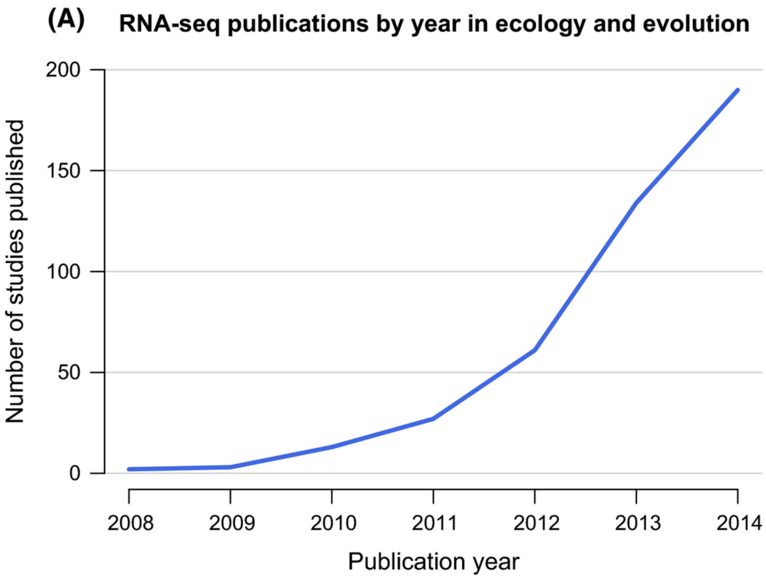
RNA seq promotion (ends on Aug 30, 2016)
1, RNA seq for gene expression $280/sample (was $350/sample)
---10M reads, 50bp SE or 75bp SE sequencing with HiSeq4000 or NextSeq 500
---$280/sample pricing requires a minimum of 12 samples for this price
---Including library preparation/QC

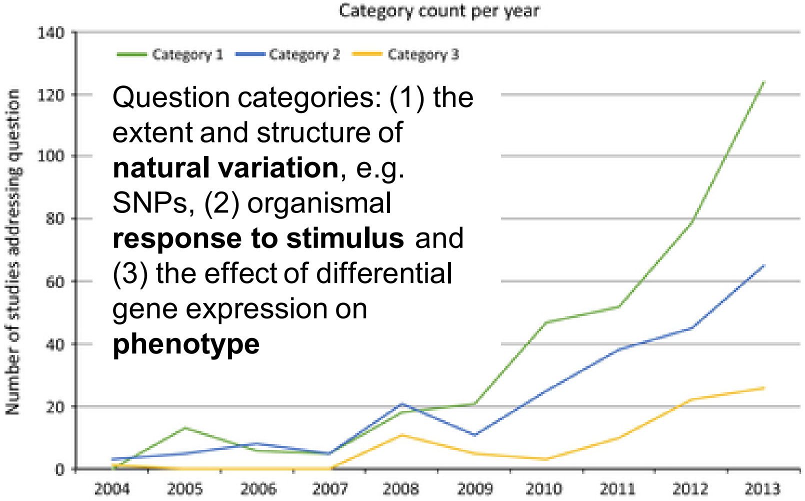
2, RNA splicing/RNA editing/Fusion... Show more »

Starting at $823.00 USD per sample

Microbiome Core Facility

RNAseq - review



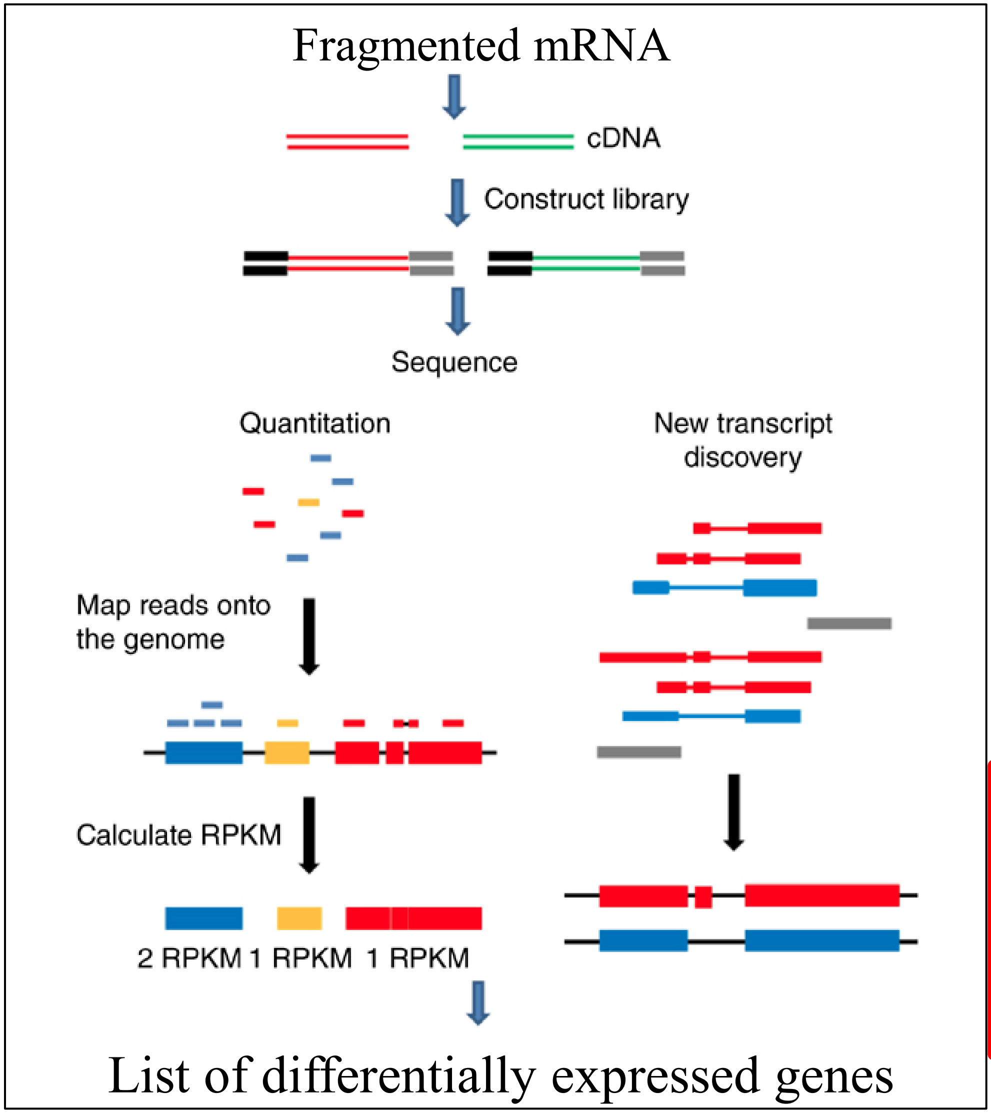


Question categories: (1) the extent and structure of **natural variation**, e.g. SNPs, (2) organismal **response to stimulus** and (3) the effect of differential gene expression on **phenotype**

Alvarez et al. 2015, Todd et al. 2016

RNA-Seq = sekvenování transkriptomu

RNA-Seq workflow for gene expression analysis



gene ontology
identifikace funkce daného genu

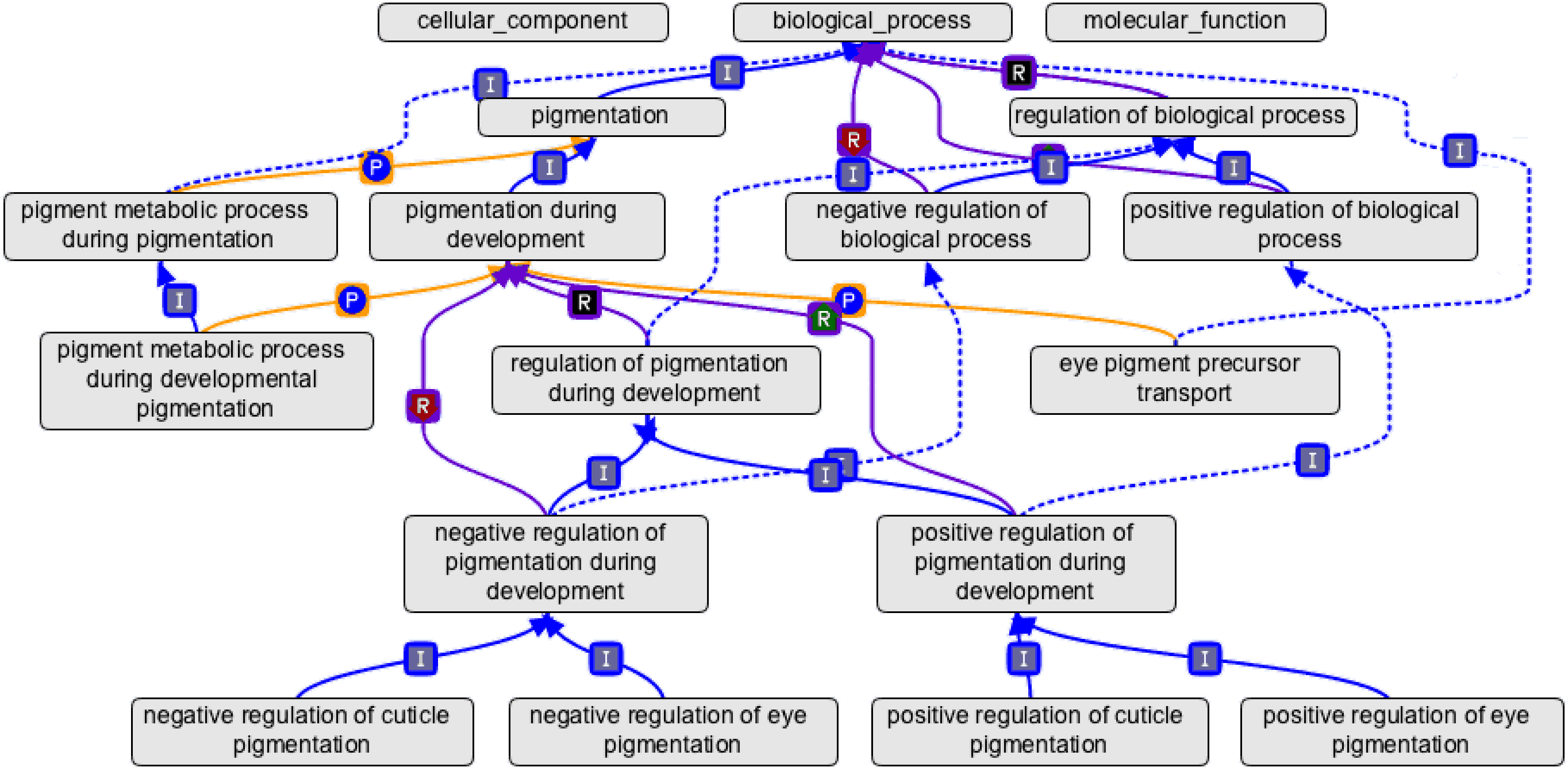
Adapted from Zeng and Mortazavi, *Nature Immunology* **13** (2012)

Gene ontology (http://geneontology.org/)

= functional annotation analysis

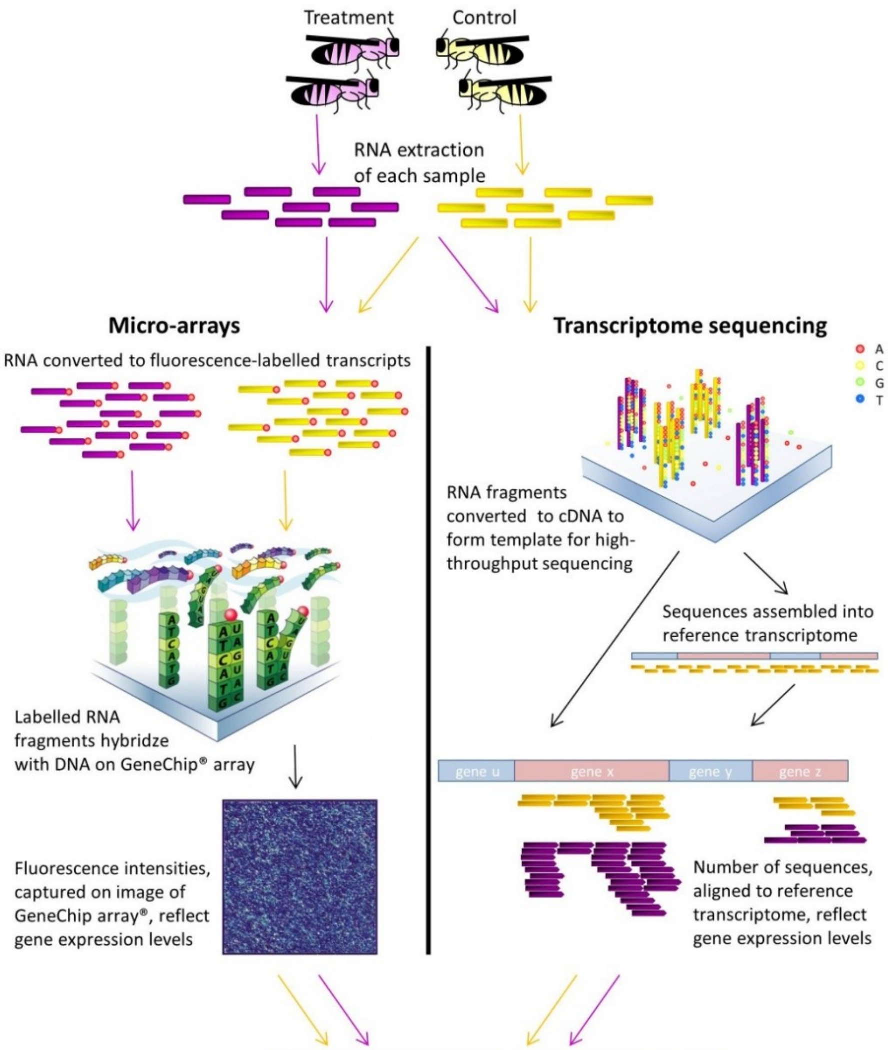
- založena na databázích dostupných anotovaných genů u modelových organismů
- **Cellular Component** - the parts of a cell or its extracellular environment
- **Molecular Function** - the elemental activities of a gene product at the molecular level, such as binding or catalysis
- **Biological Process** - operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.

Gene ontology as a graph



Example: A set of terms under the biological process node pigmentation.

Microarrays vs. RNA-seq

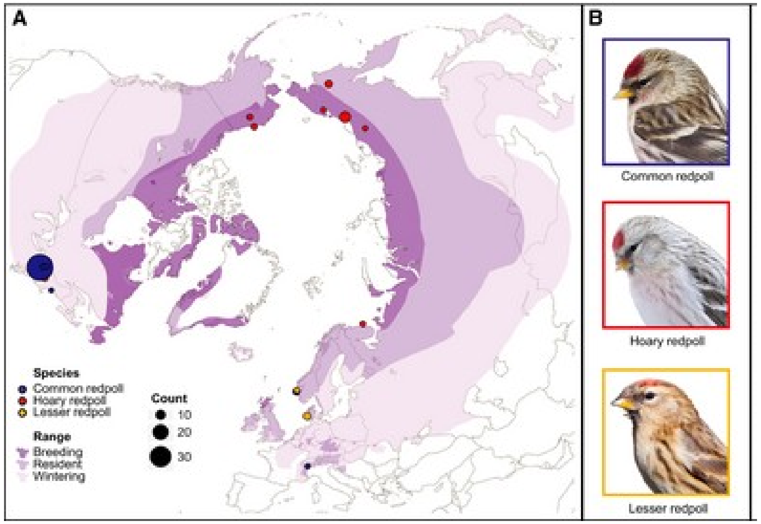


Gene ID	Transcript abundance in Treatment sample	Transcript abundance in Control sample
ABc00001	3778	3894
ABc00002	768	189
ABc00003	1087	1476

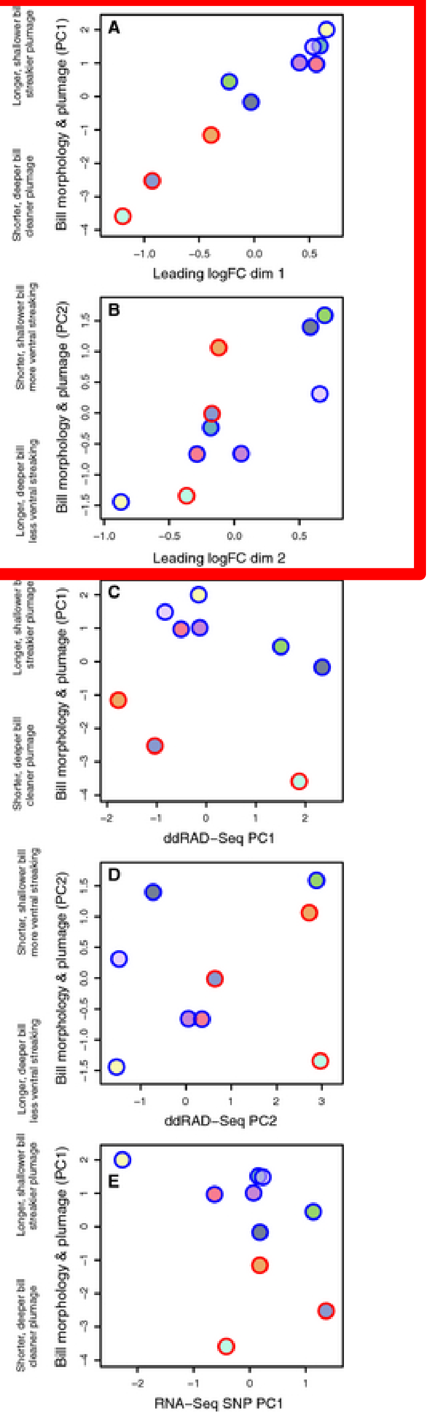
Měly by dávat stejné výsledky, ale srovnávacích studií je velmi málo

Příklady

Mason and Taylor 2015

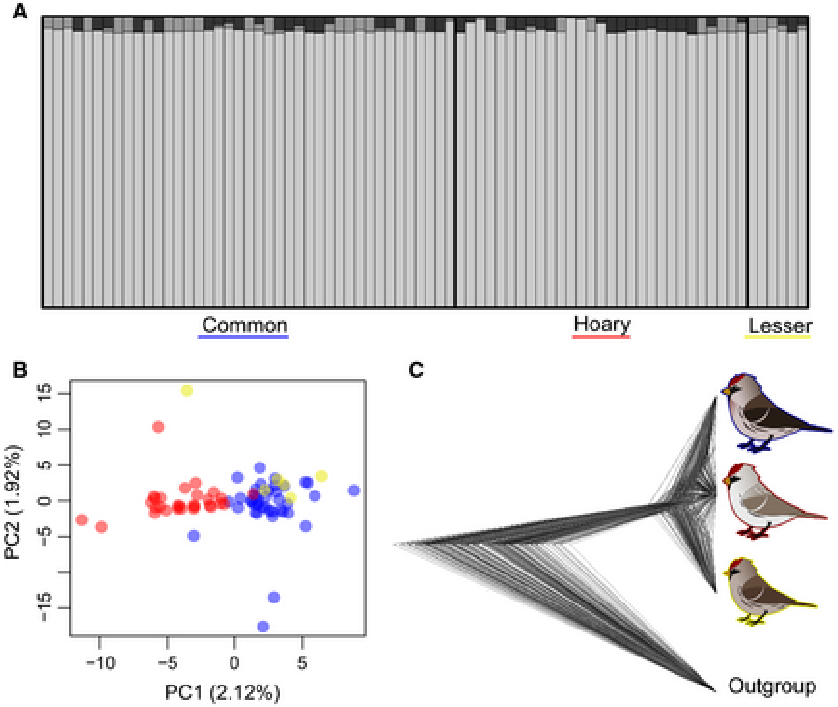


Only differentially expressed genes are responsible for morphological changes (zobák, zbarvení)



20721 SNPs (ddRAD)

215825 SNPs (RNAseq)



20721 SNPs (ddRAD) – no genetic difference at neutral loci

Ng'oma *et al. BMC Genomics* 2014, **15**:754
http://www.biomedcentral.com/1471-2164/15/754

BMC Genomics

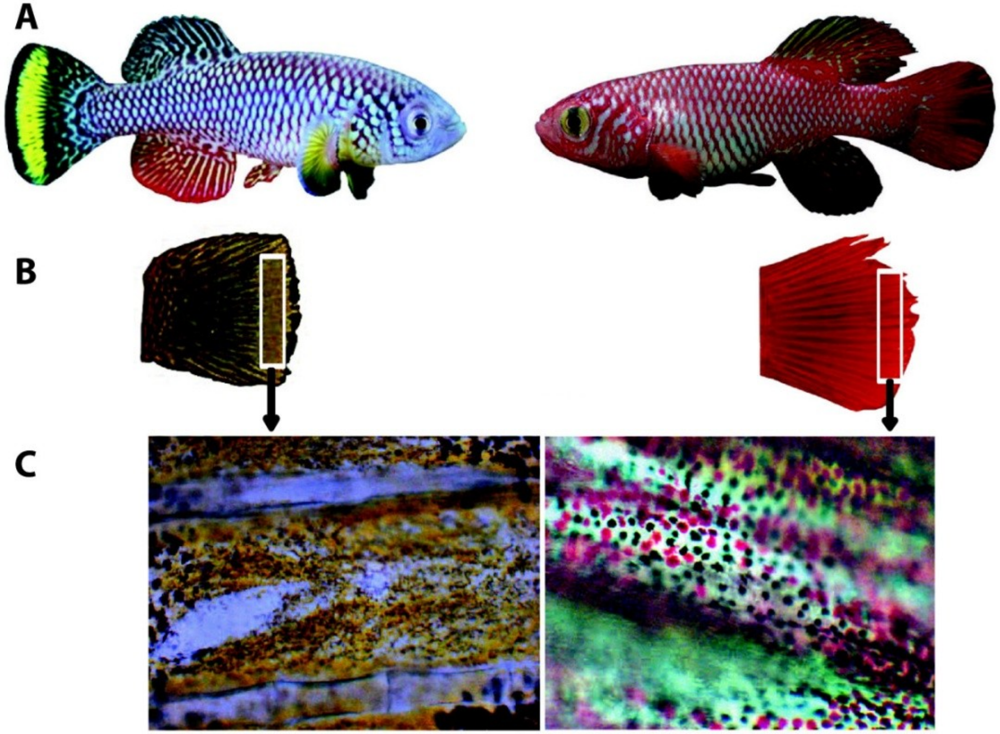
RESEARCH ARTICLE **Open Access**

Transcriptome profiling of natural dichromatism in the annual fishes *Nothobranchius furzeri* and *Nothobranchius kadleci*

Enoch Ng'oma[1], Marco Groth[1], Roberto Ripa[2], Matthias Platzer[1] and Alessandro Cellerino[1,2*]

- cca 334 miliónů sekvencí („reads“); 42 mil./sample

- 210 DEGs („differentially expressed genes“) – 119 up-regulated, 91 down-regulated u žlutých jedinců



xanthophory

melanophory
erythrophory

Konzistence výsledků

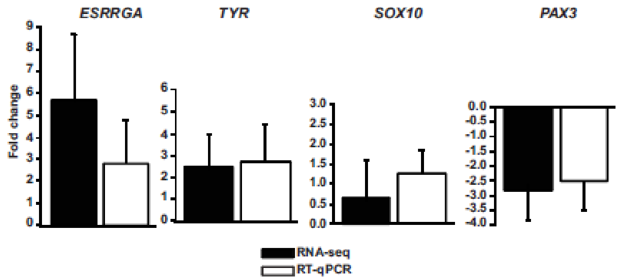


Figure 2 RT-qPCR validation of sample genes regulated in RNA-seq data. Four sample genes were selected based on fold change or known role in melanocyte biology ($\log_2$ fold change yellow vs. red morph). The same individuals whose expression was quantified with RNA-seq were used in qPCR. Expression was normalized to *TBP*, as internal control; error bars represent standard deviation.

- změny v expresi jdou stejným směrem u RNA-seq i RT-qPCR vybraných genů

Functional annotation clustering (= gene ontology)

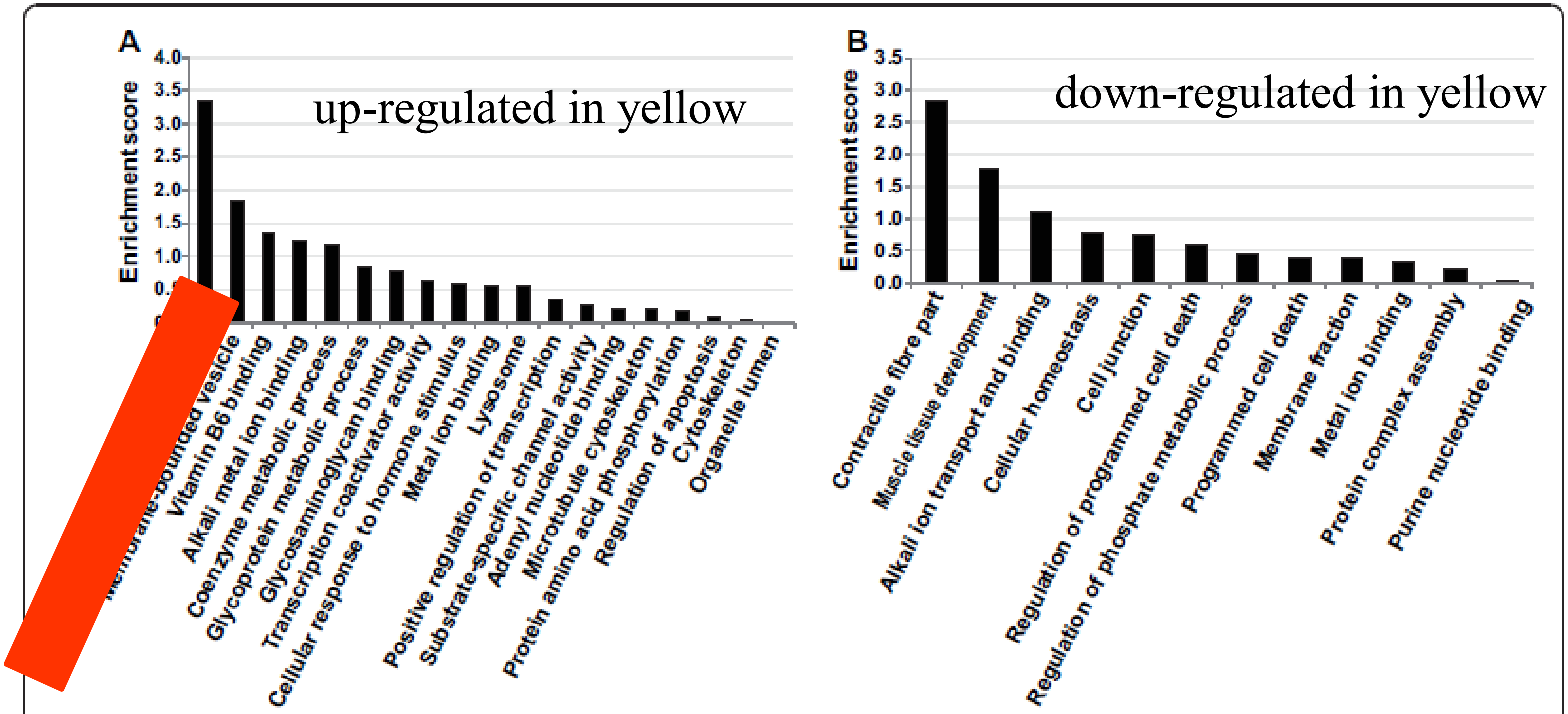


Figure 3 Gene Ontology. DAVID Functional Annotation Clustering (FAC) analysis of differentially expressed genes (DEGs) obtained by comparison of yellow vs. red tail RNA samples. **A.** Enriched functional gene clusters for the 119 up-regulated genes. **B.** Enriched functional gene clusters for the 91 down-regulated genes. Significance is determined by corresponding enrichment scores.

- xanthophory u žlutých jedinců jsou asociovány s melanogenezí
- v dalším kroku je možné studovat roli jednotlivých kandidátních genů

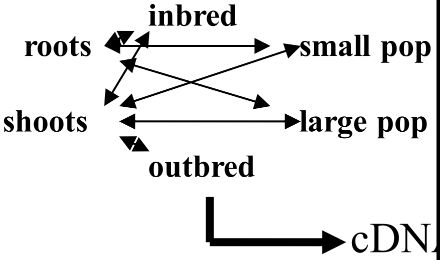
Case study: Joop Ouborg et al.

Transcriptional profiling of inbreeding depression and genetic erosion in *Scabiosa columbaria*: the balance between genetic drift and selection in the genetic erosion process.

2009

Example:

Scabiosa columbaria



inbred

roots

small pop

shoots

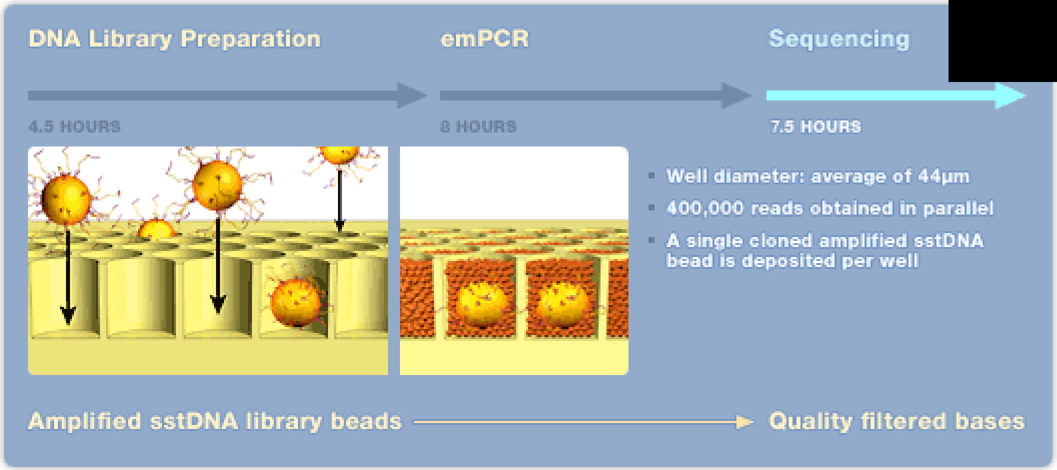
large pop

outbred

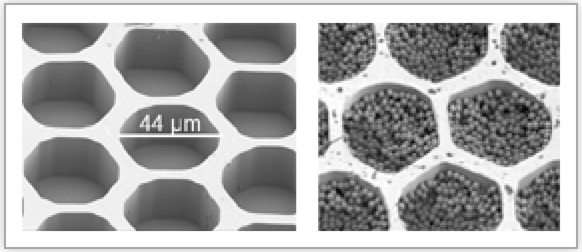
cDN

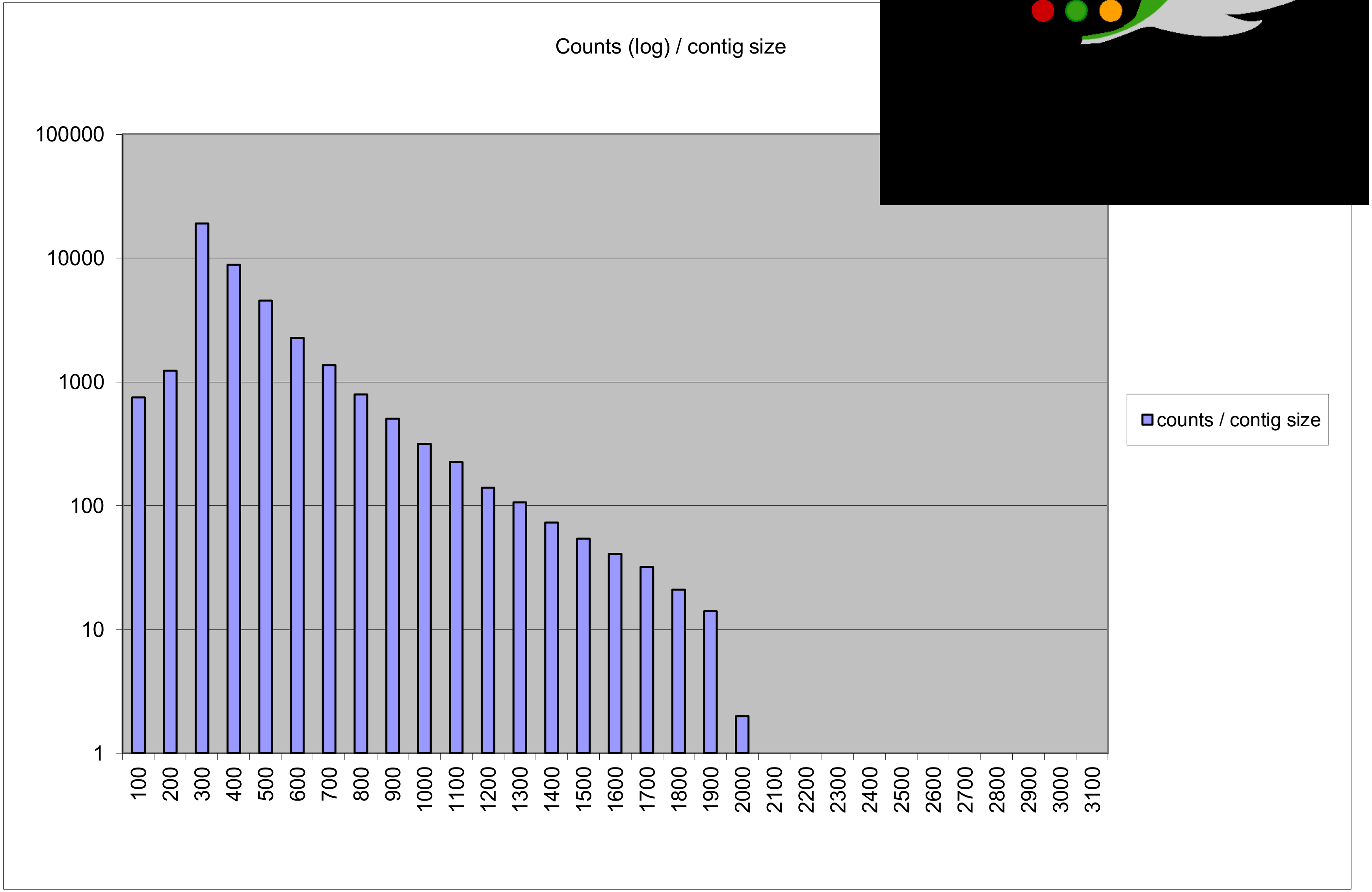
cDNA library preparation – 454 sequencing

FIGURE 9



- Well diameter: average of 44μm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well





Total number of reads: **528557** Number of contigs: **40302**

In the next phase:

Annotation of these 40.000^{+} ESTs („e
sequence tags“)

Automated programs available, like **BLAST2GO (**http://www.blast2go.de/**)**:

just feed a file with the ESTs into the program, and turn it on……

1 week later you will have the results, being:

- Homology with known sequences
- Known function

The sequences may also be searched for:

EST-associated SSR markers: MISA (http://pgrc.ipk-gatersleben.de/misa/)

SNP markers: SNP-mining software like PolyBayes

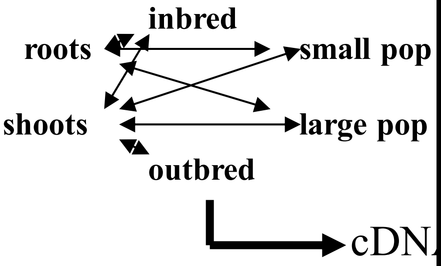
(http://genome.wustl.edu/tools/software/polybayes.cgi)

Again by using search software, freeware

ALMOST HALF OF GENES (ESTs) ARE UNKNOWN !!!

Example:

Scabiosa columbaria



inbred

roots

small pop

shoots

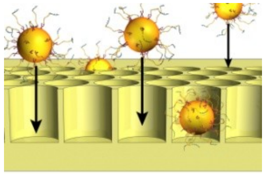
large pop

outbred

cDN

enomic

454 LIFE SCIENCES



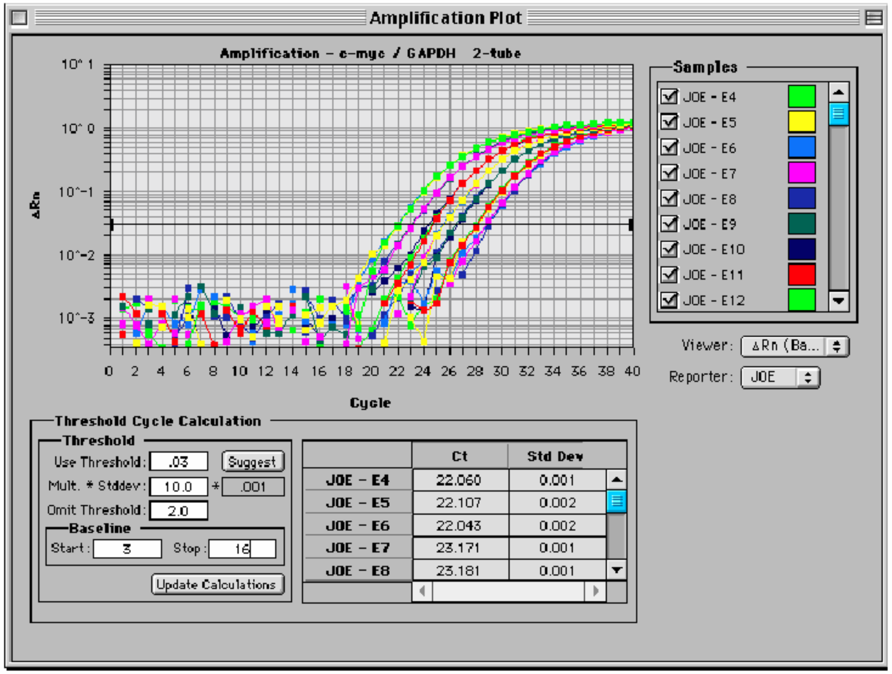
NCBI

530.000 sequences in one run, leading to ~ 40.000 ESTs

Two methods of detecting transcri

1. Design of quantitative RealTime-PCR met
EST sequences

2. Design of a *Scabiosa* specific microarray



☑ JOE - E4 🟩
☑ JOE - E5 🟨
☑ JOE - E6 🟦
☑ JOE - E7 🟪
☑ JOE - E8 🟦
☑ JOE - E9 🟩
☑ JOE - E10 🟦
☑ JOE - E11 🟥
☑ JOE - E12 🟩

Viewer: ΔRn (Ba...

Reporter: JOE

Threshold Cycle Calculation

Threshold

Use Threshold: .03

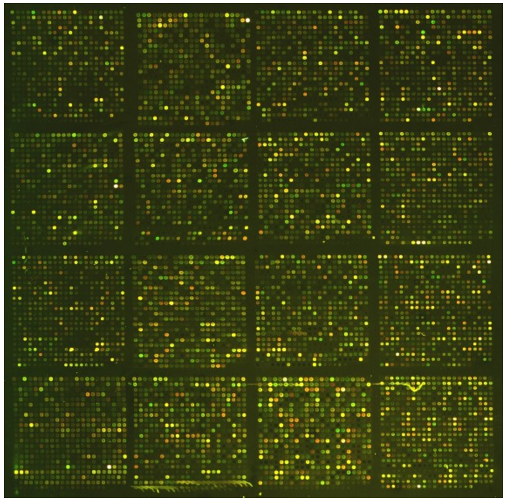
Mult. * Stddev: 10.0 * .001

Omit Threshold: 2.0

Baseline

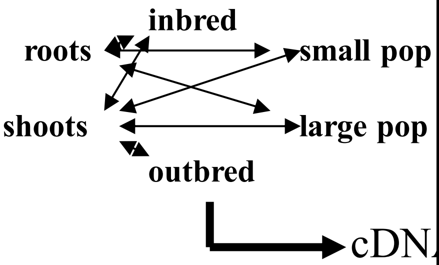
Start: 3 Stop: 16

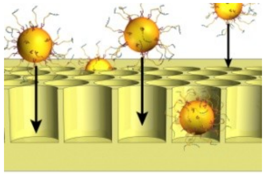
	Ct	Std Dev
JOE – E4	22.060	0.001
JOE – E5	22.107	0.002
JOE – E6	22.043	0.002
JOE – E7	23.171	0.001
JOE – E8	23.181	0.001



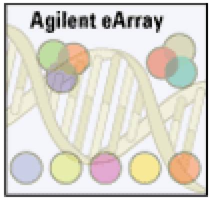
Example:

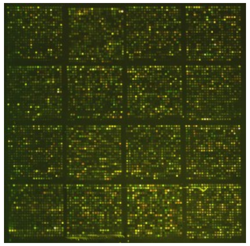
Scabiosa columbaria


inbred
roots
small pop
shoots
large pop
outbred
cDN



530.000 sequences in one run, leading to ~ 40.000 ESTs





15k – 30k
60-mer
microarrays

Experiment: transcriptional profiling of inbreeding depression