

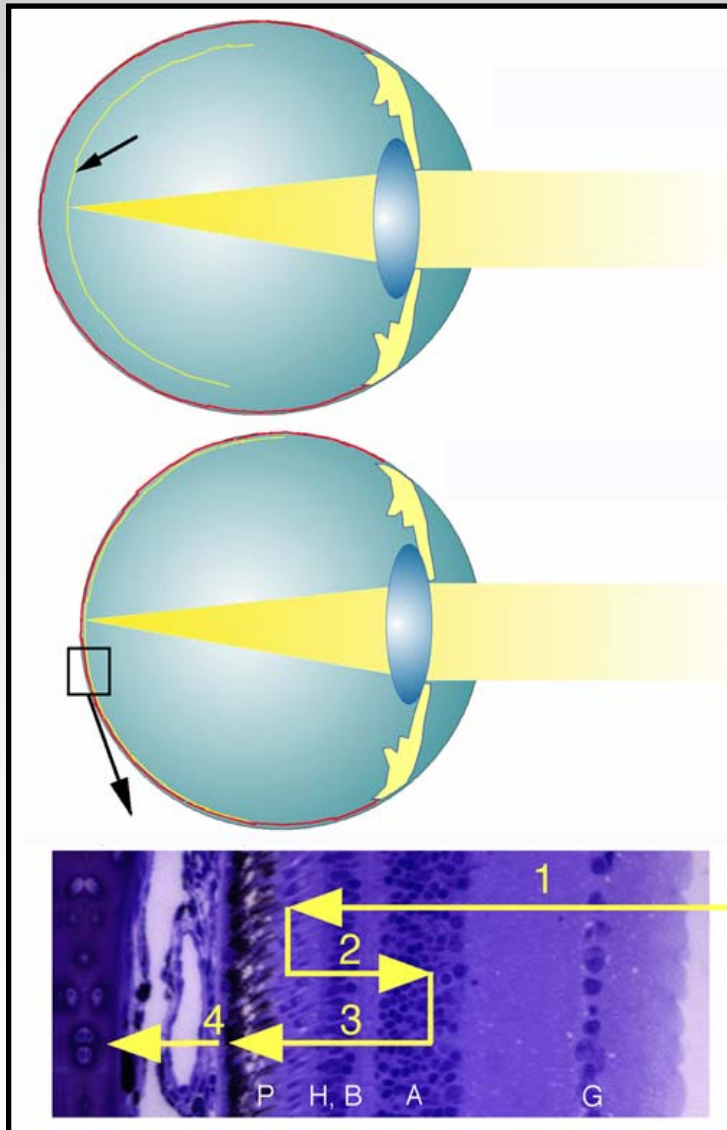
# Gene expression analyses in the eye and ear: recent findings and future possibilities

Marita Feldkaemper

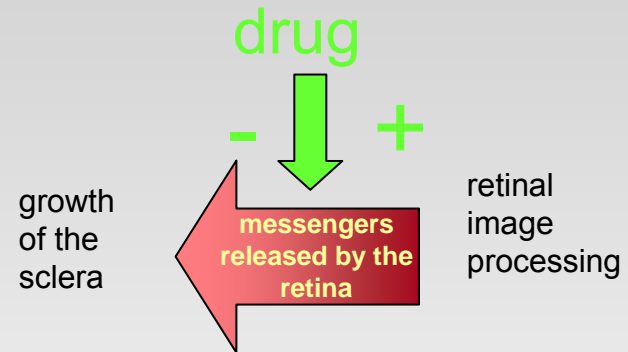
Centre for Ophthalmology, Institute for Ophthalmic Research Tuebingen, Germany



# Signaling pathways during eye growth regulation



- modifying visually induced signals

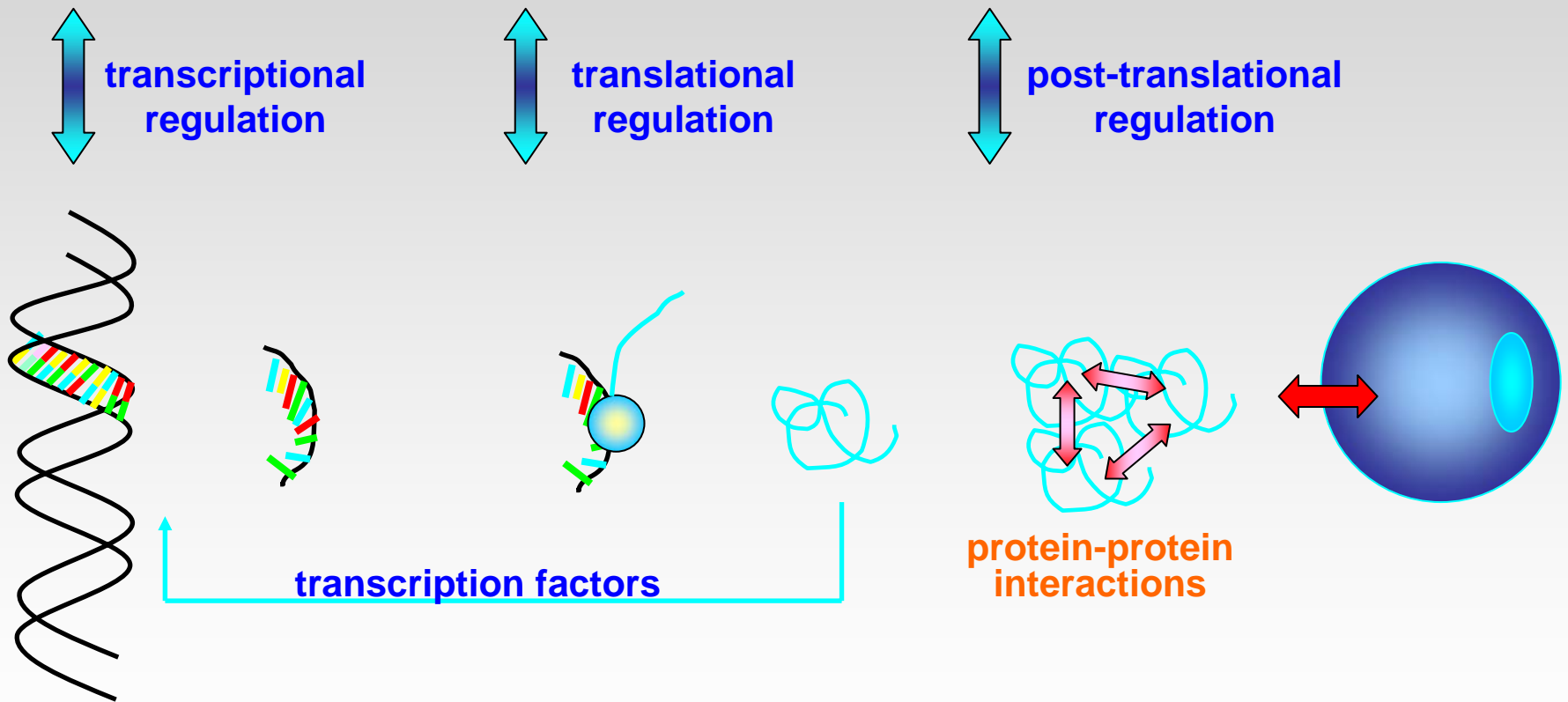


How can the best targets be identified ?

- find new targets by screening for genes that respond to experimentally induced myopia
- optimize compounds that act on already known targets

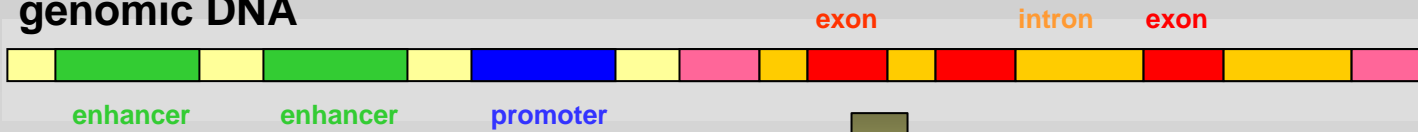
# Different levels of visual and genetic control

- Eye growth is controlled at several levels



# Transcription of eukaryotic genes

## genomic DNA



transcription

## hnRNA (hetero-nuclear, very short lived)



splicing

## mRNA (messenger)



translation

## protein



RNA Polymerase I transcribes ribosomal RNA (rRNA).

RNA Polymerase II transcribes messenger RNA (mRNA) and most small nuclear RNAs (snRNAs).

RNA Polymerase III transcribes transfer RNA (tRNA) and other small RNAs (including the small 5S rRNA).

# Types of RNA

About 97% of the transcriptional output is non-protein-coding in eukaryotes.

## Non-coding RNAs involved in translation

**transfer RNA** (tRNA); each kind of tRNA carries (at its 3' end) one of the 20 **amino acids**  
**ribosomal RNA** (rRNA): 18S rRNA, 28S, 5.8S, and 5S rRNA

## Non-coding RNAs involved in gene regulation

**micro RNAs** (miRNA, single stranded, 21-22 nucleotide): can down-regulate gene expression  
**small interfering RNA** (siRNA, double-stranded, 20-25 nucleotides), involved in RNA interference pathway.  
**long noncoding RNAs** that regulate genes.

## Protein coding RNAs

**Messenger RNA** (mRNA) carries information about a protein sequence to the ribosomes. mRNA may contain regulatory elements itself, in the 5' untranslated region or 3' untranslated region. These cis-regulatory elements regulate the activity of that mRNA. The untranslated regions can also contain elements that regulate other genes.

*(not complete)*

# Transcriptome

## Transcriptome definition:

The **transcriptome** is the set of all RNA molecules, including mRNA, rRNA, tRNA and other non-coding RNA.

The term can be applied to the total set of transcripts in a given organism, or to the specific subset of transcripts present in a particular cell type.

The transcriptome can vary with external environmental conditions.

Because it includes all *mRNA* transcripts in the cell, the transcriptome reflects the genes that are being actively expressed at any given time, with the exception of mRNA degradation phenomena such as transcriptional attenuation.

**Transcriptomics** is the branch of molecular biology that deals with the study of messenger RNA molecules produced in an individual or population of a particular cell type.

# Control of the expression of the mRNA genes

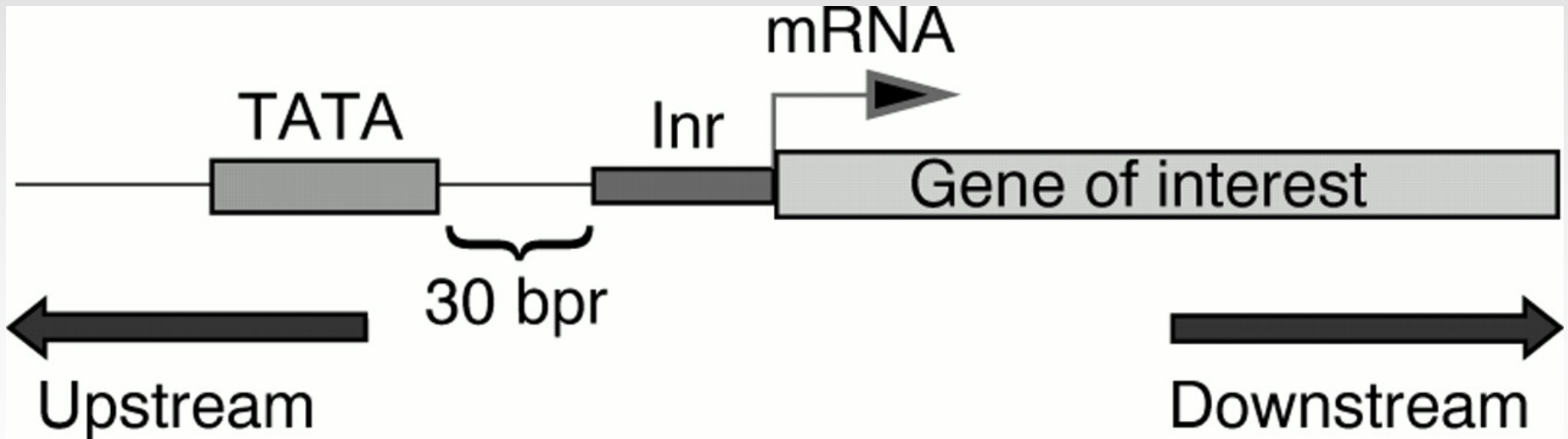
## Transcriptional initiation

This is the **most important mode for control of eukaryotic gene expression**.

## Transcription start site

RNA polymerase II binds. Pol II is a complex of different proteins.

## The basal promoter

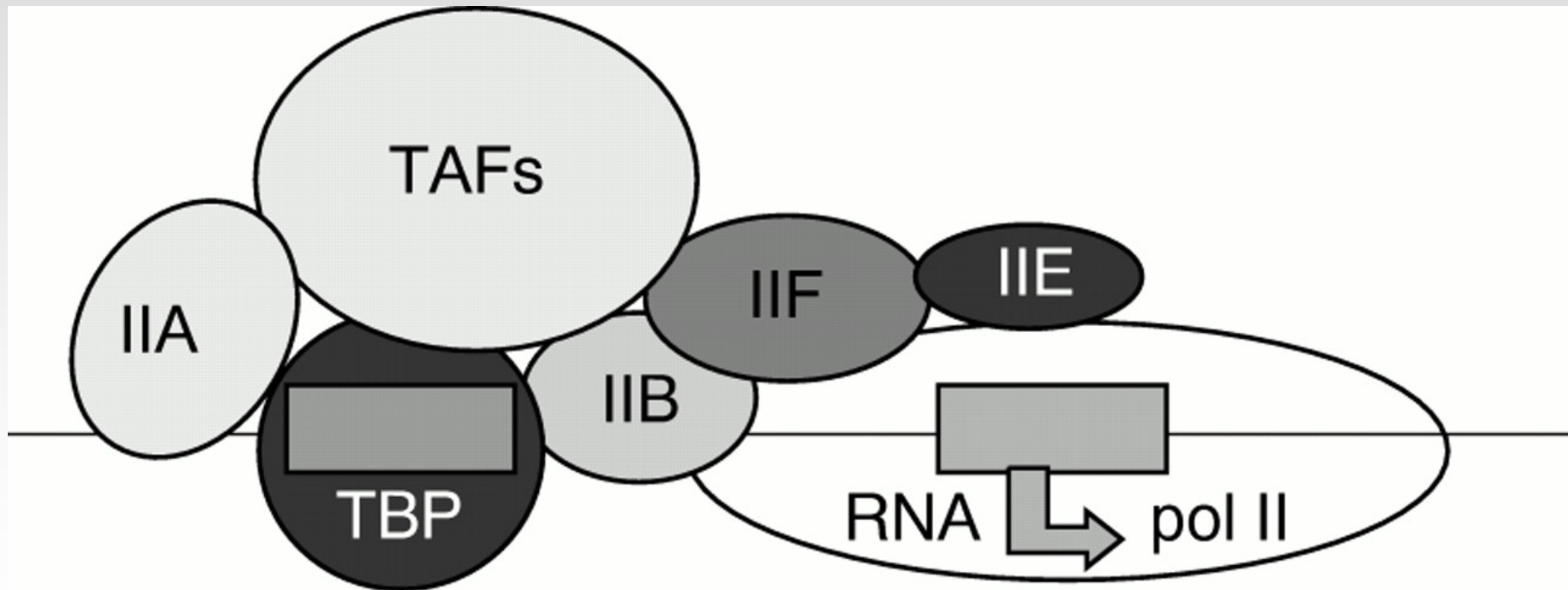


The basal promoter contains a sequence of 7 bases (TATAAAA) called the **TATA box** and an initiator element (Inr).

# The basal promoter

The basal promoter is bound by a large complex of some 50 different proteins, including **Transcription Factor IID (TFIID)**, which is a complex of **TATA-binding protein (TBP)**, which recognizes and binds to the TATA box ; 14 other protein factors which bind to TBP — and each other — but not to the DNA; **Transcription Factor IIB (TFIIB)** which binds both the DNA and pol II.

The basal or core promoter **is found in all protein-coding genes.**





# Transcription of eukaryotic genes

## The upstream enhancer/promoter

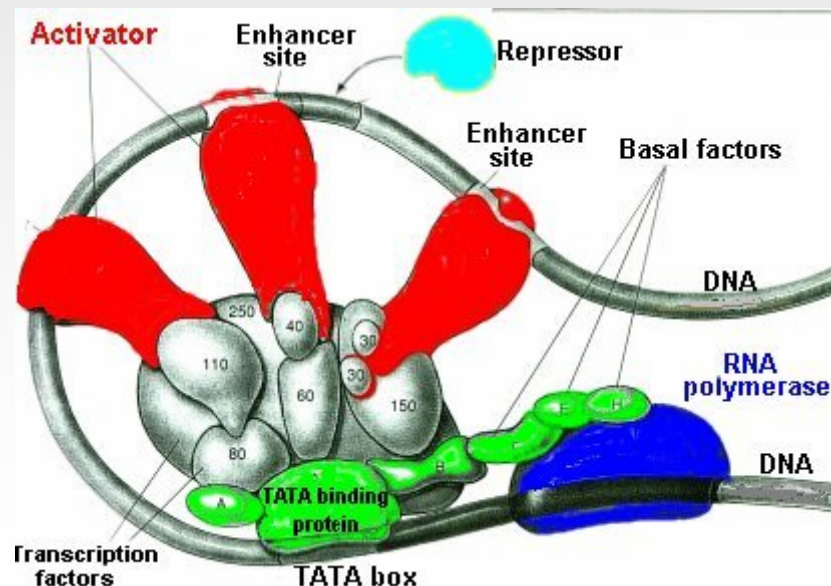
Structure and associated binding factors **differ from gene to gene**.

Regulatory sequences are predominantly located upstream (5') of the transcription initiation site, although some elements occur downstream (3') or even within the genes themselves.

The **number** and **type** of regulatory elements to be found varies with each mRNA gene.

**Different combinations** of transcription factors also can exert differential regulatory effects upon transcriptional initiation.

Various cell types each express **characteristic combinations** of transcription factors, this is the **major mechanism for cell-type specific regulation** of mRNA expression.



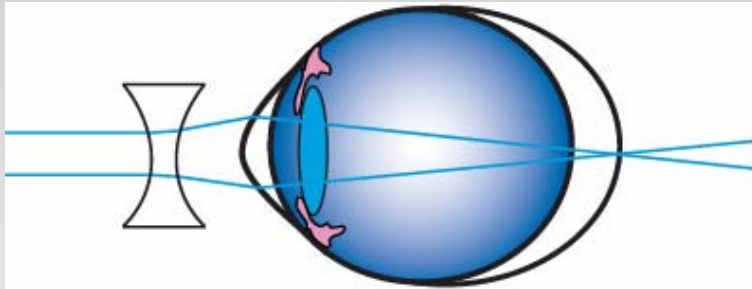
# Methods used in transcriptomics studies

## Analysis of induced gene expression changes

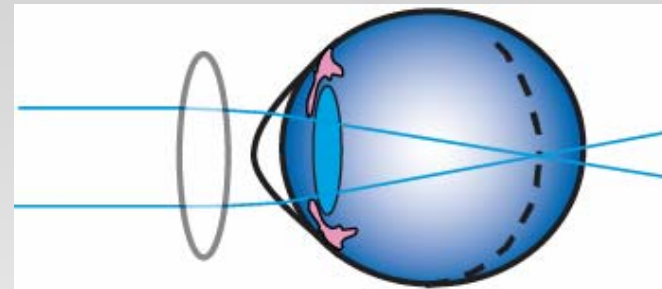
- PCR and real-time PCR
- Northern blot
- In-situ hybridisation
- Oligonucleotide microarrays
- Next generation sequencing

# Measuring induced changes in transcription rate of a (known) gene in the visual system and the ear

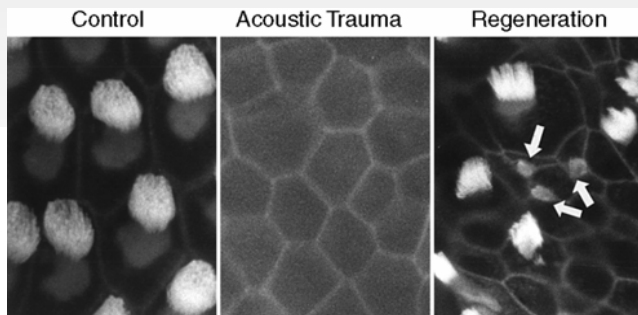
Examples for induced changes in the eye and ear:



Minus lens-wear – increased ocular growth



Plus lens-wear – decreased ocular growth



Margaret Lomax et al. (2001). *Noise & Health* 3 :19-35.

When noise exposure kills hair cells in birds, these cells can regenerate and hearing will recover. In mammals, however, the hair cell loss, and resulting hearing loss, is permanent.

# Reverse Transcriptase Polymerase Chain Reaction (PCR)

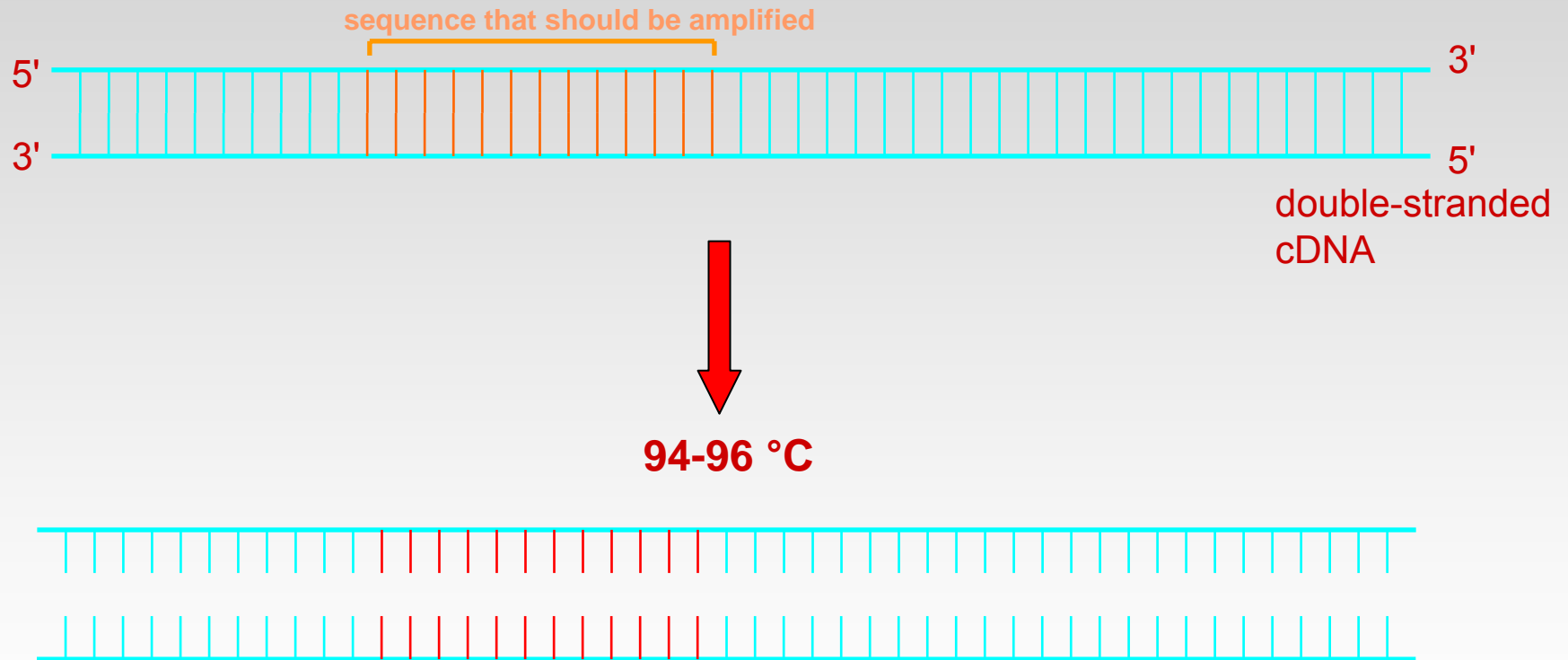
- invented by Dr. Kary Mullis 1983 (awarded with the Nobel prize 1993)
- makes it possible to produce very high numbers of copies of a sequence
- clever tricks, using repetitive "melting" of cDNA



temperature raised  
to 94-96° C

# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- RNA extraction, conversion into cDNA with reverse transcriptase  
**STEP 1: heating and denaturation**

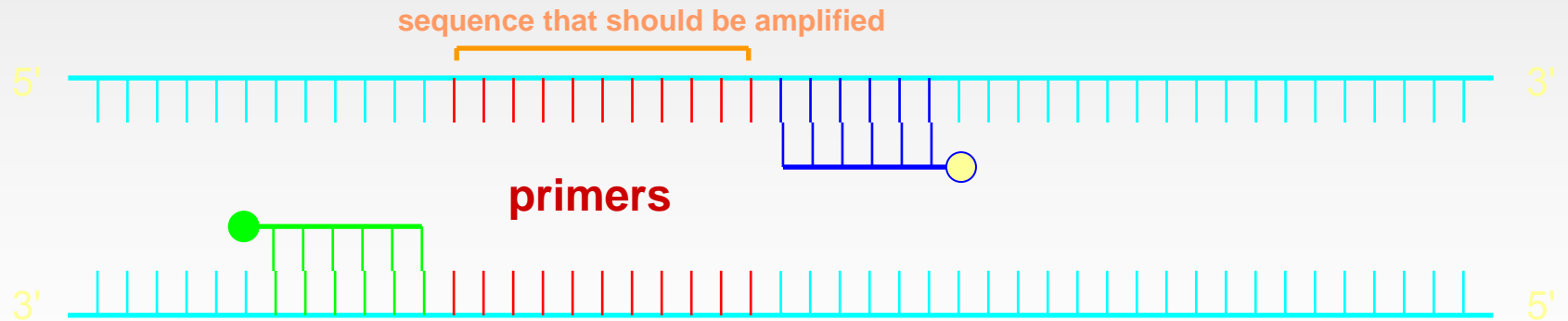


# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- STEP 2: annealing of primers



50-65 °C

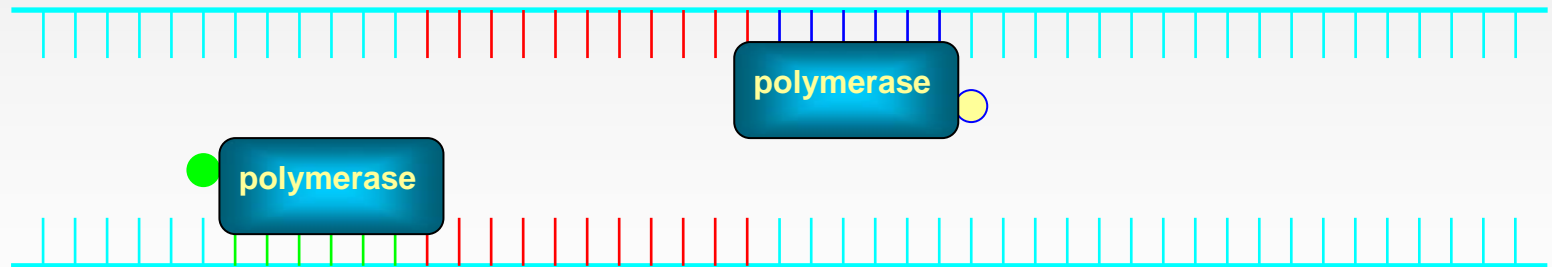


# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- STEP 3: action of the Taq polymerase



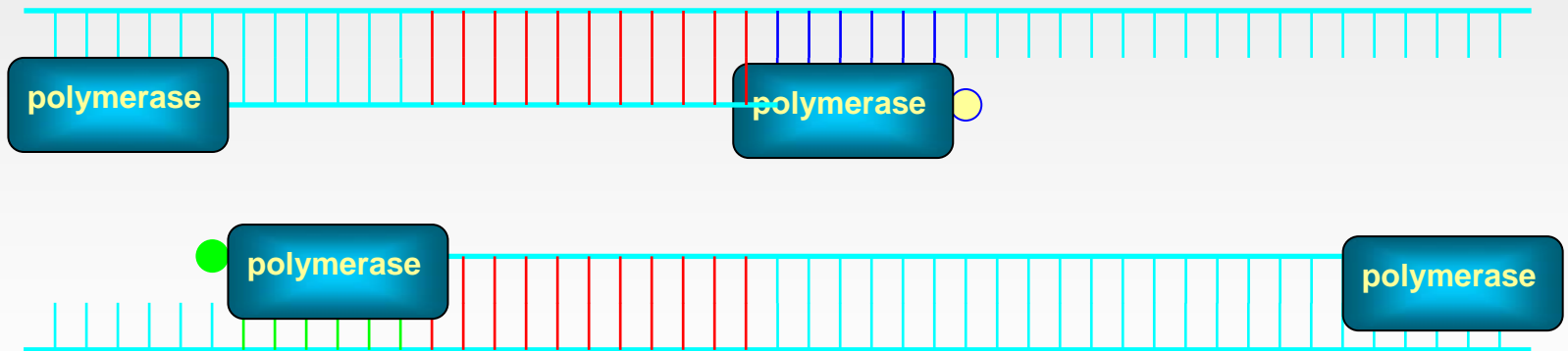
72 °C



# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- STEP 4: synthesis (elongation)

72 °C





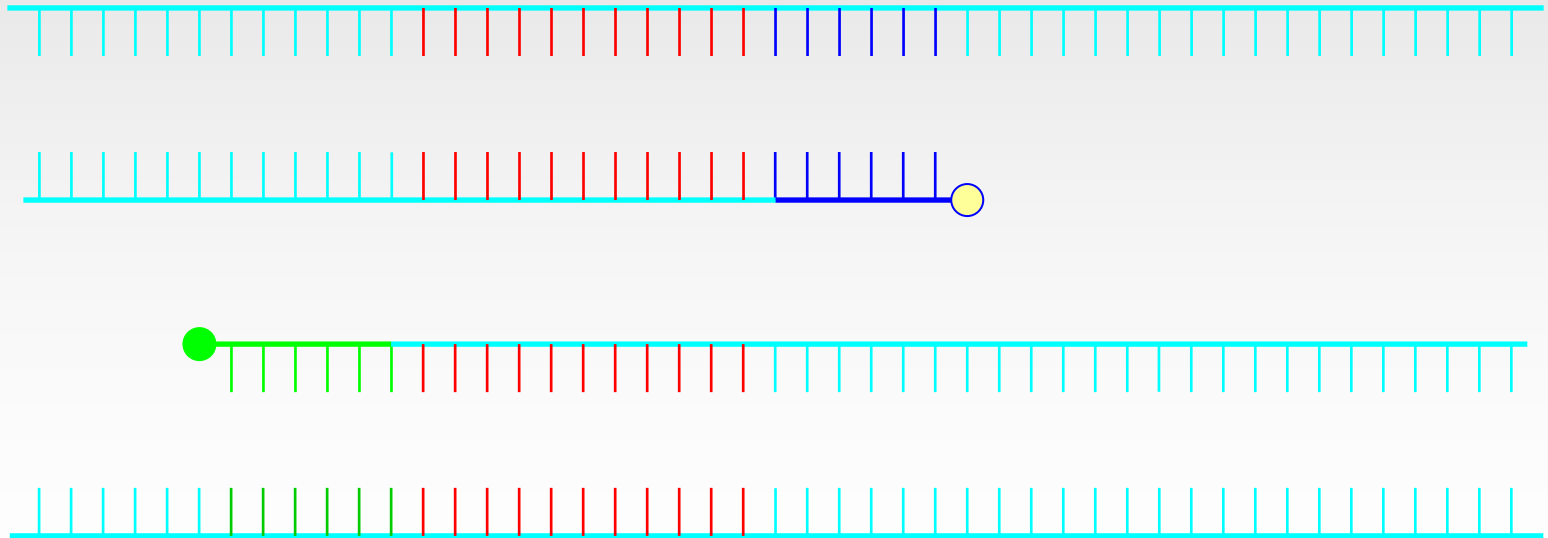
# Reverse Transcriptase Polymerase Chain Reaction (PCR)



STEP 1  
Cycle 2: heating and denaturation



94-96 °C



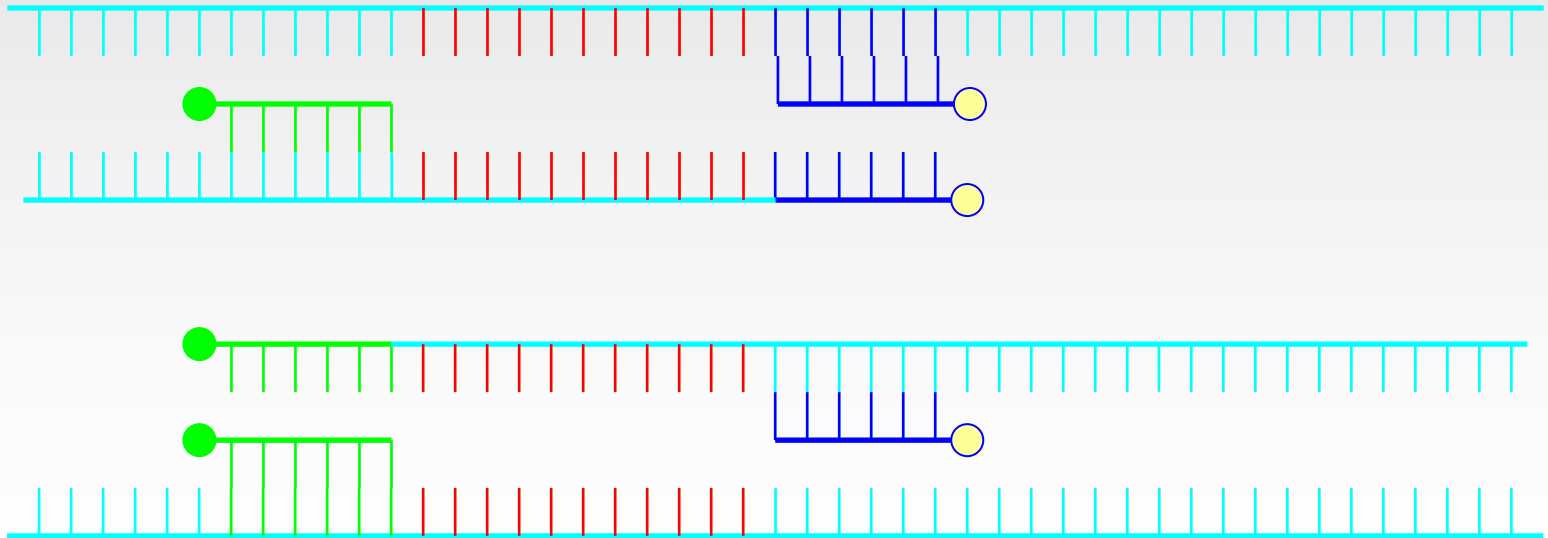
# Reverse Transcriptase Polymerase Chain Reaction (PCR)



STEP 2:  
Cycle 2: annealing of primers



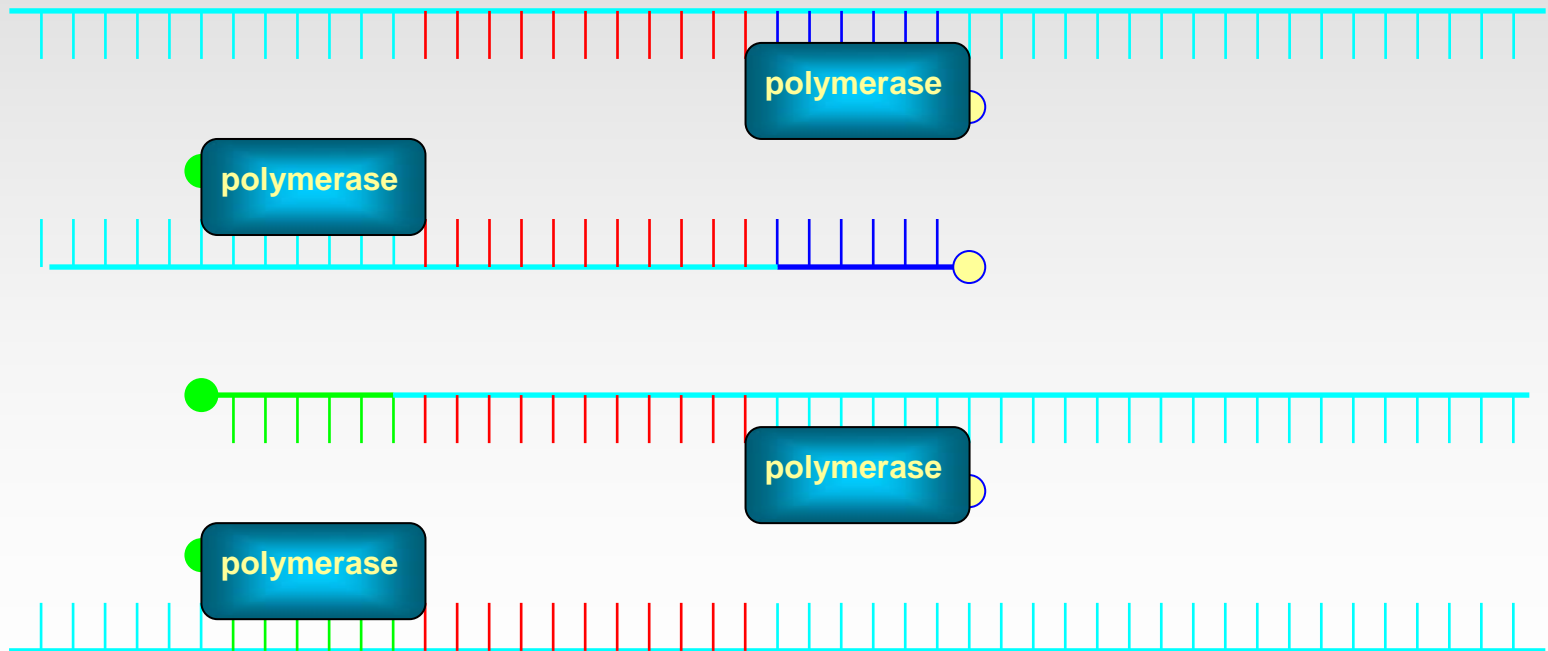
50-65 °C



# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- STEP 3  
Cycle2: attaching the Taq polymerase

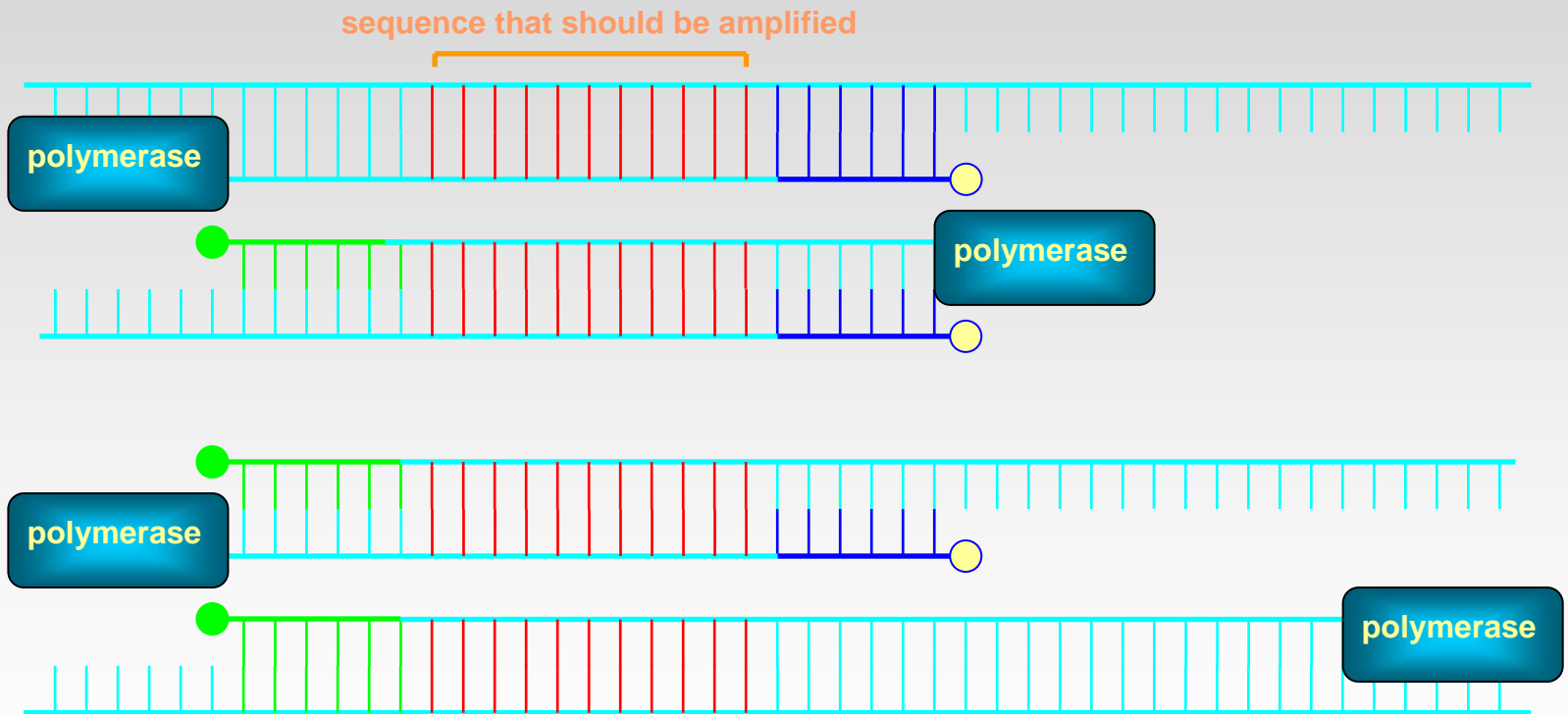
72 °C



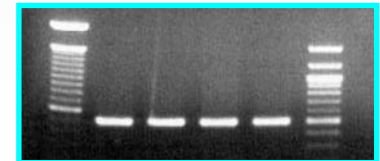
# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- STEP 4  
Cycle 2: synthesis

72 °C



after 35-45 cycles:  
Ethidiumbromide-stained agarose gel  
 $2^{40} = 10^{12}$  amplification factor



# Reverse Transcriptase Polymerase Chain Reaction (PCR)

- Problems of standard PCR: not exact quantification possible
- Amplification functions may have different shapes and they saturate

## Quantitative methods:

- a) Northern blot
- b) Real-time PCR

# Quantitative measurement of RNA: Northern Blot

area ~ amount RNA



tissue

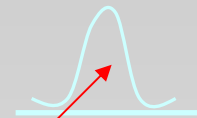
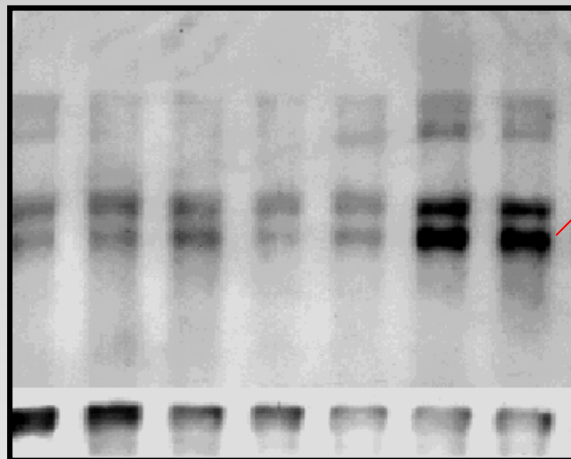
total RNA 1-10  $\mu\text{g}$ !  
(about 6 pg/cell)



Large fragments

Small fragments

18S-rRNA



enzyme-linked antibody against DIG: amplification!

Digoxigenin

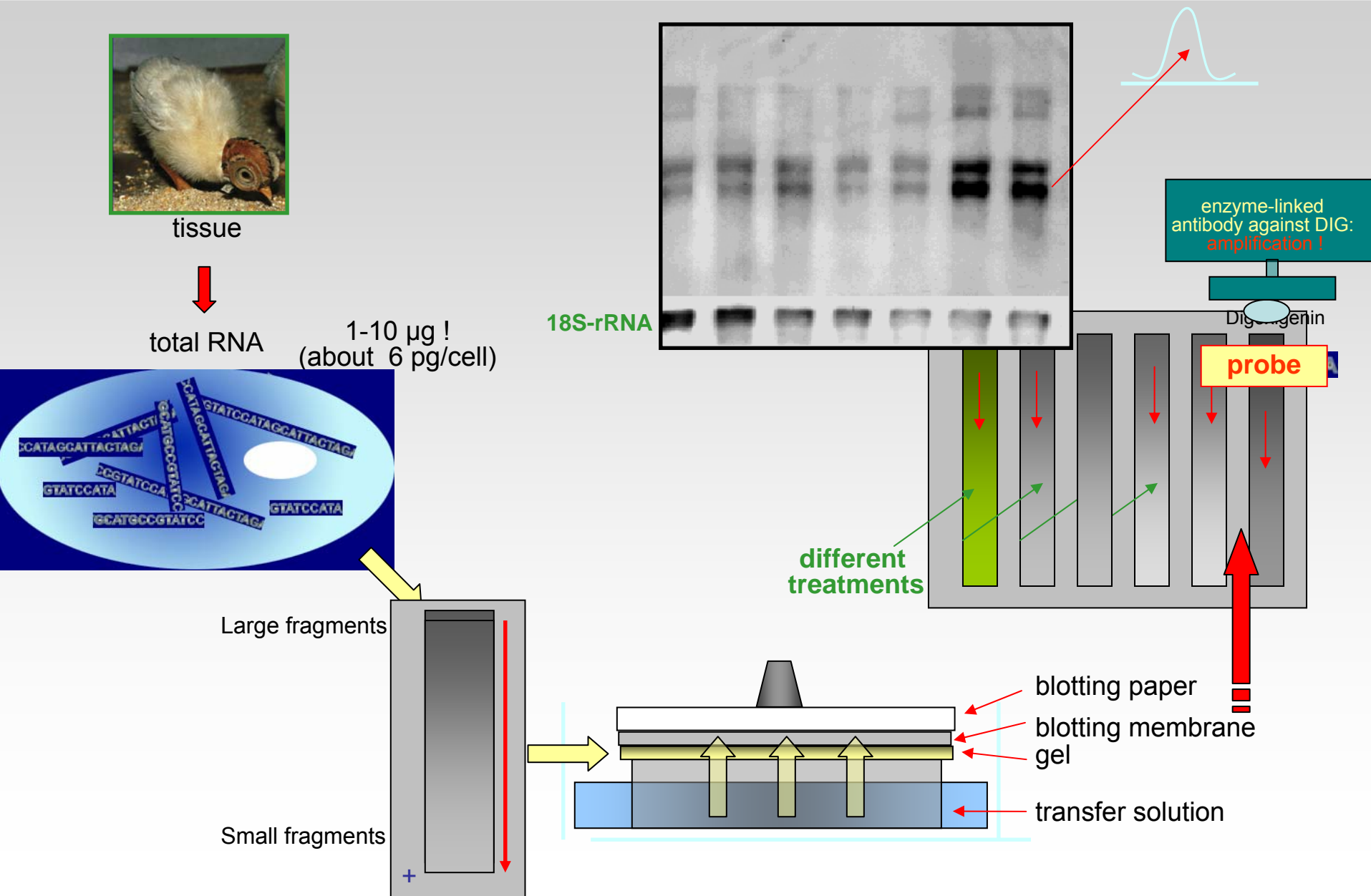
probe

different treatments

blotting paper

blotting membrane

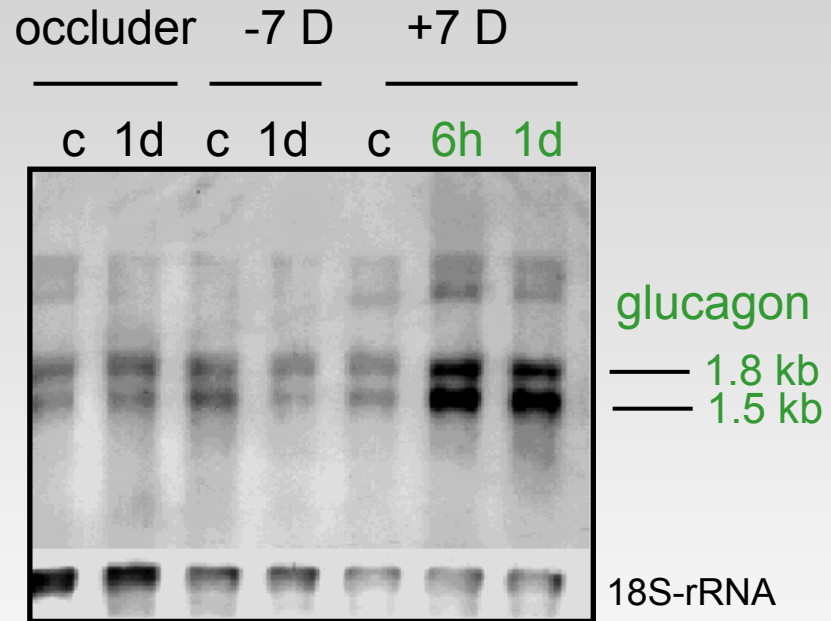
transfer solution



# Northern Blot: examples of results



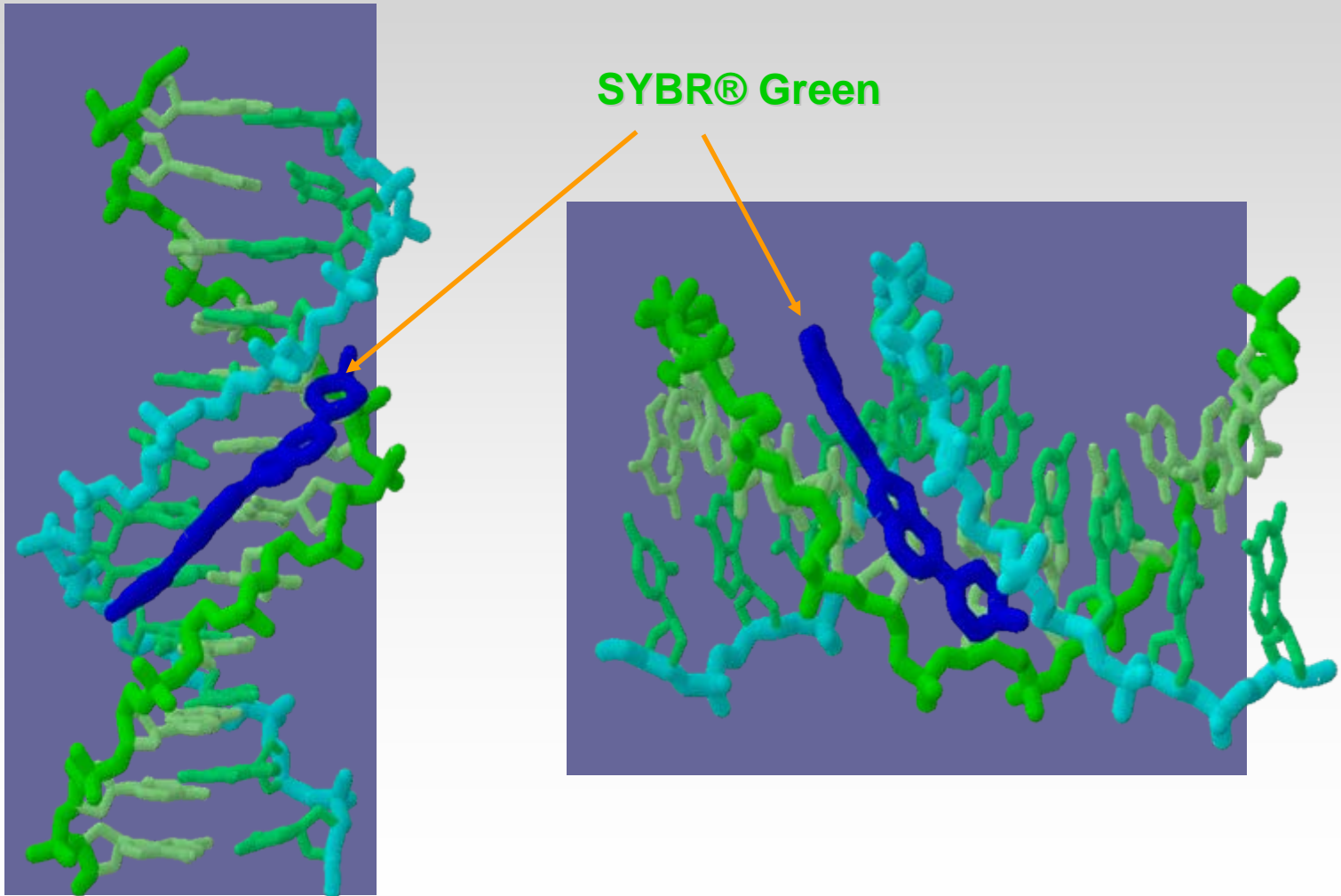
After 6 hours with positive lenses, glucagon mRNA is increased (Feldkaemper et al, IOVS 41, 1623-1628, 2002).



- Advantages:  
additional information about the size of the mRNA transcripts
- Disadvantages:  
need of high amount of RNA

# Real Time Polymerase Chain Reaction (real-time PCR)

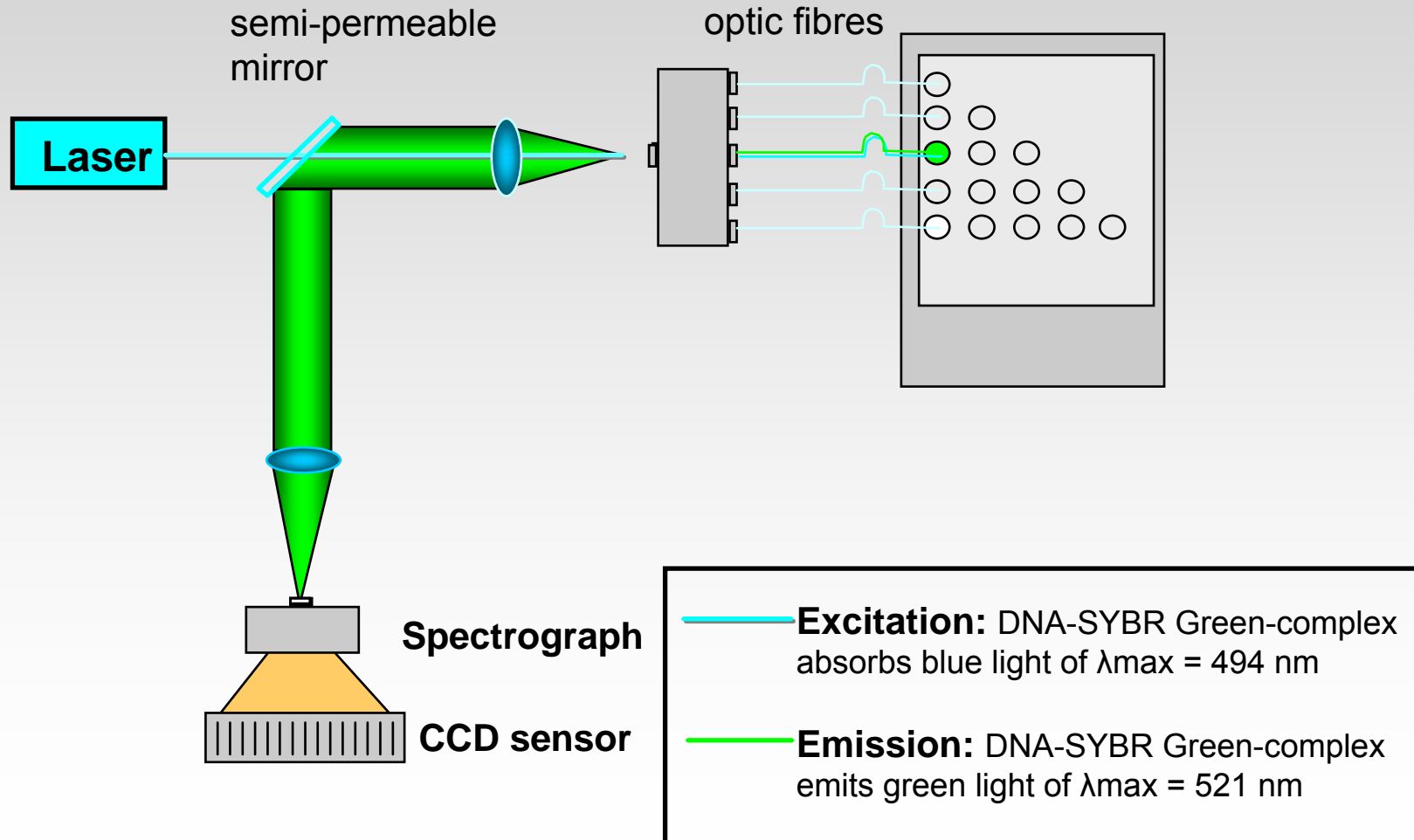
- simplest method: SYBR® Green fluorescence ~ double-stranded DNA





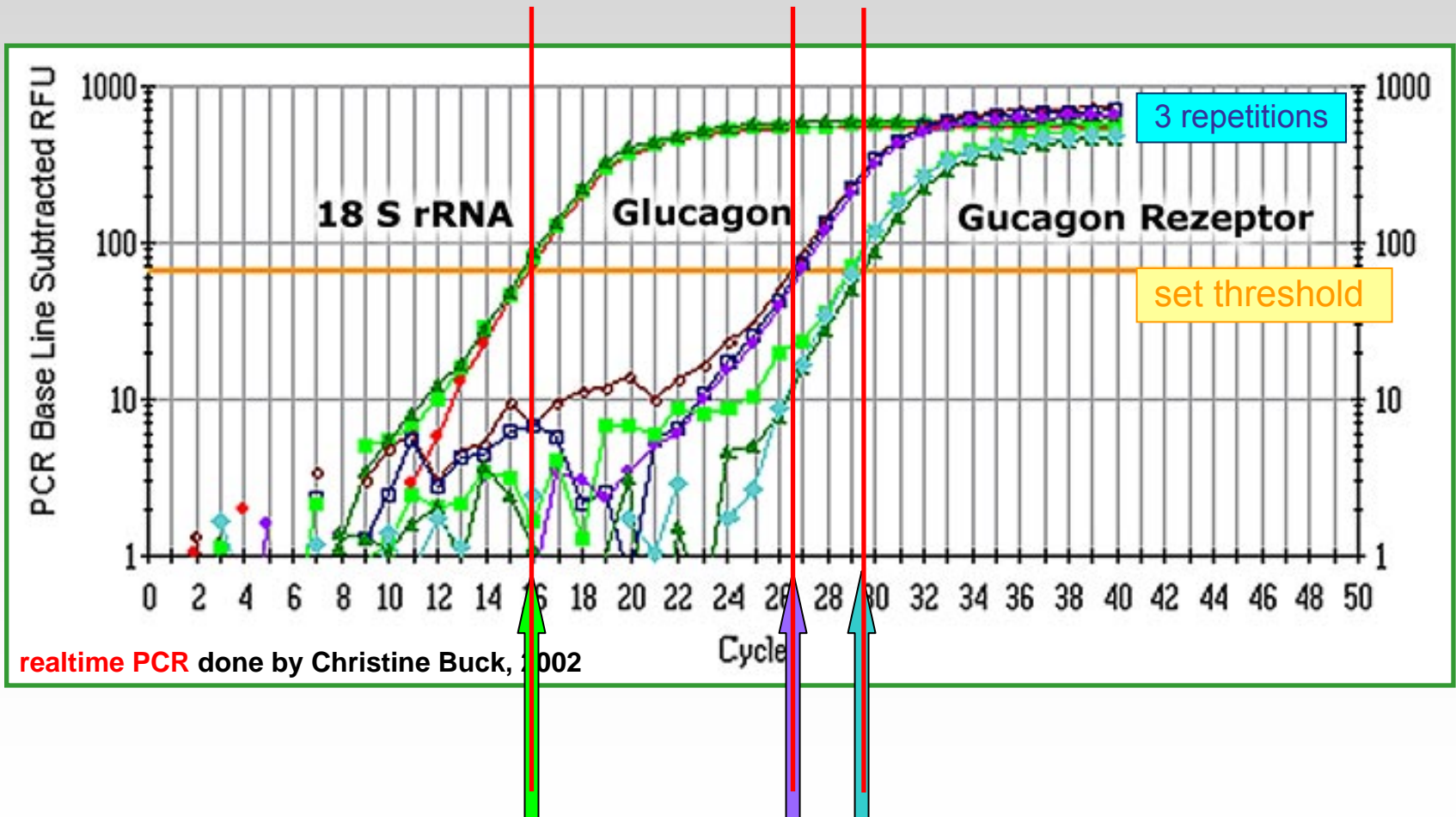
# Real Time Polymerase Chain Reaction (real-time PCR)

- quantitative Real Time PCR: thermocycler with integrated photometer



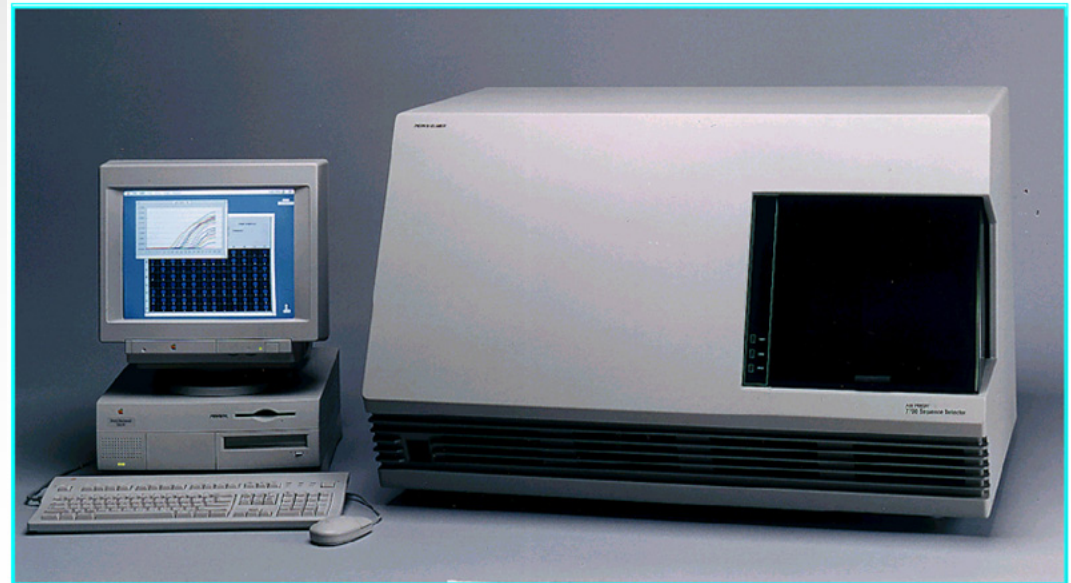
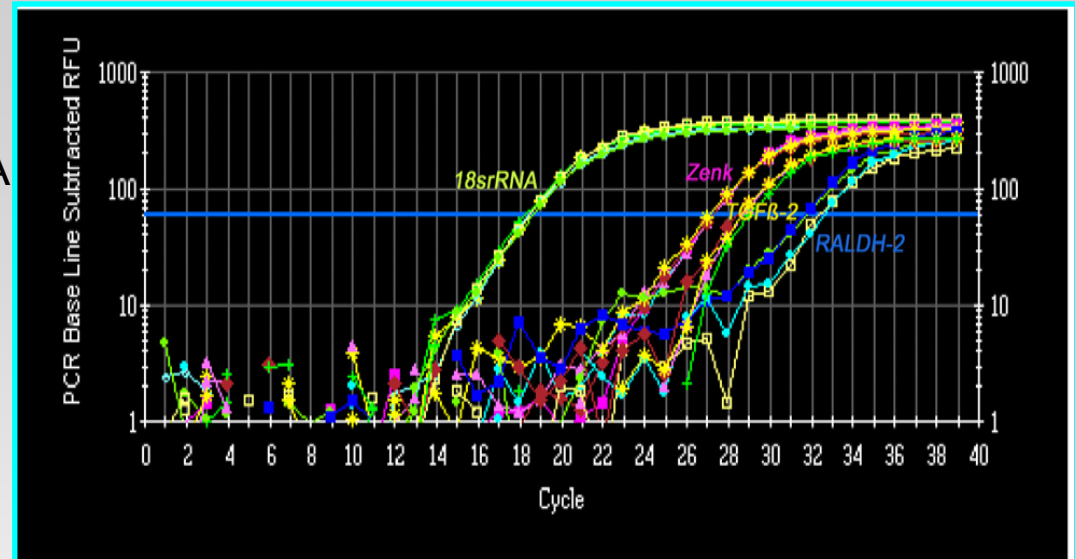
# Real time Polymerase Chain Reaction (real-time PCR)

- "on-line" measurement of the increase in copies of the target sequence

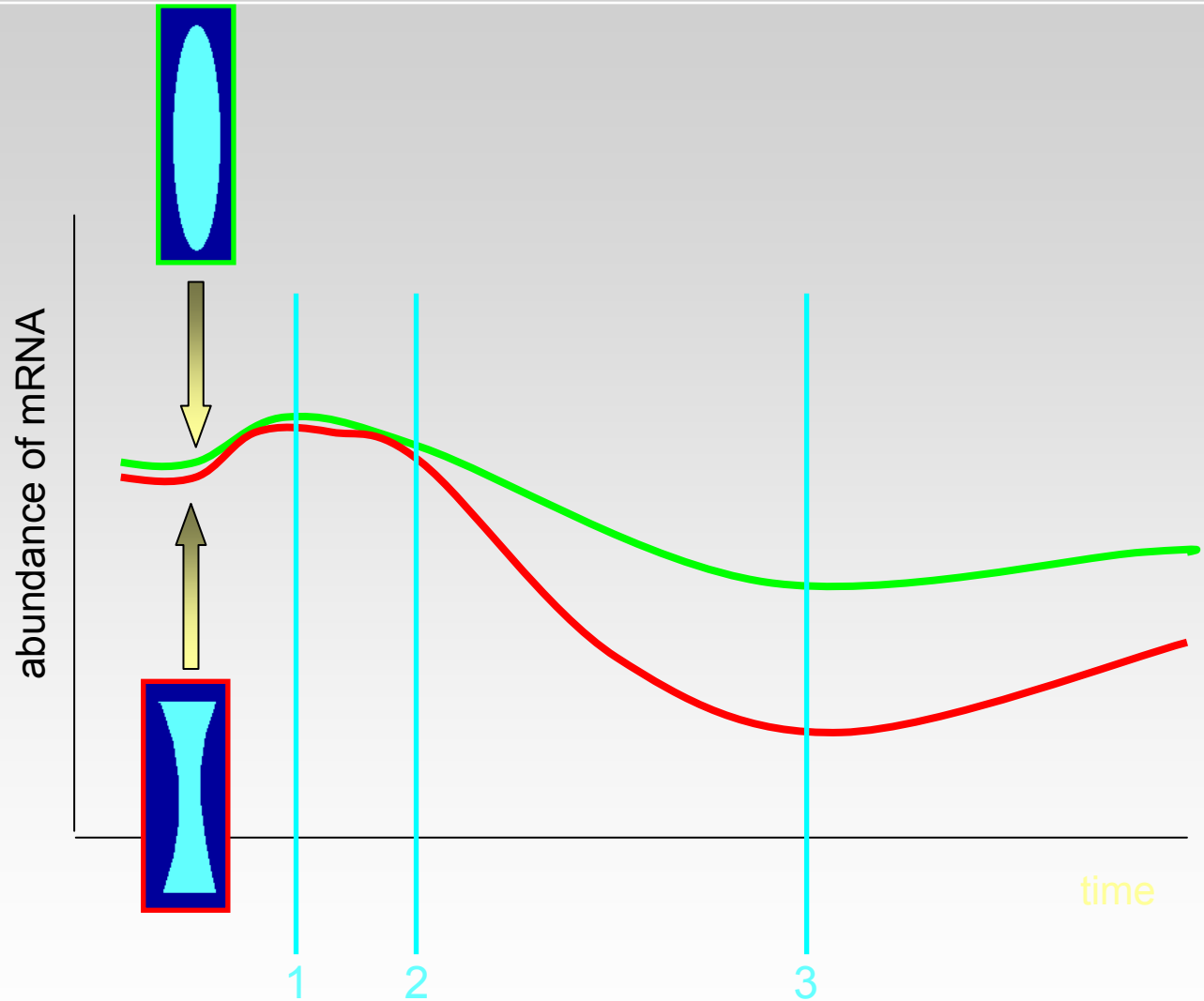


# Real Time Polymerase Chain Reaction (real-time PCR)

- Advantages
  - only very low amount of RNA/cDNA is needed (1 ng)
  - quick
- Disadvantages:
  - relatively expensive



# Time course problem

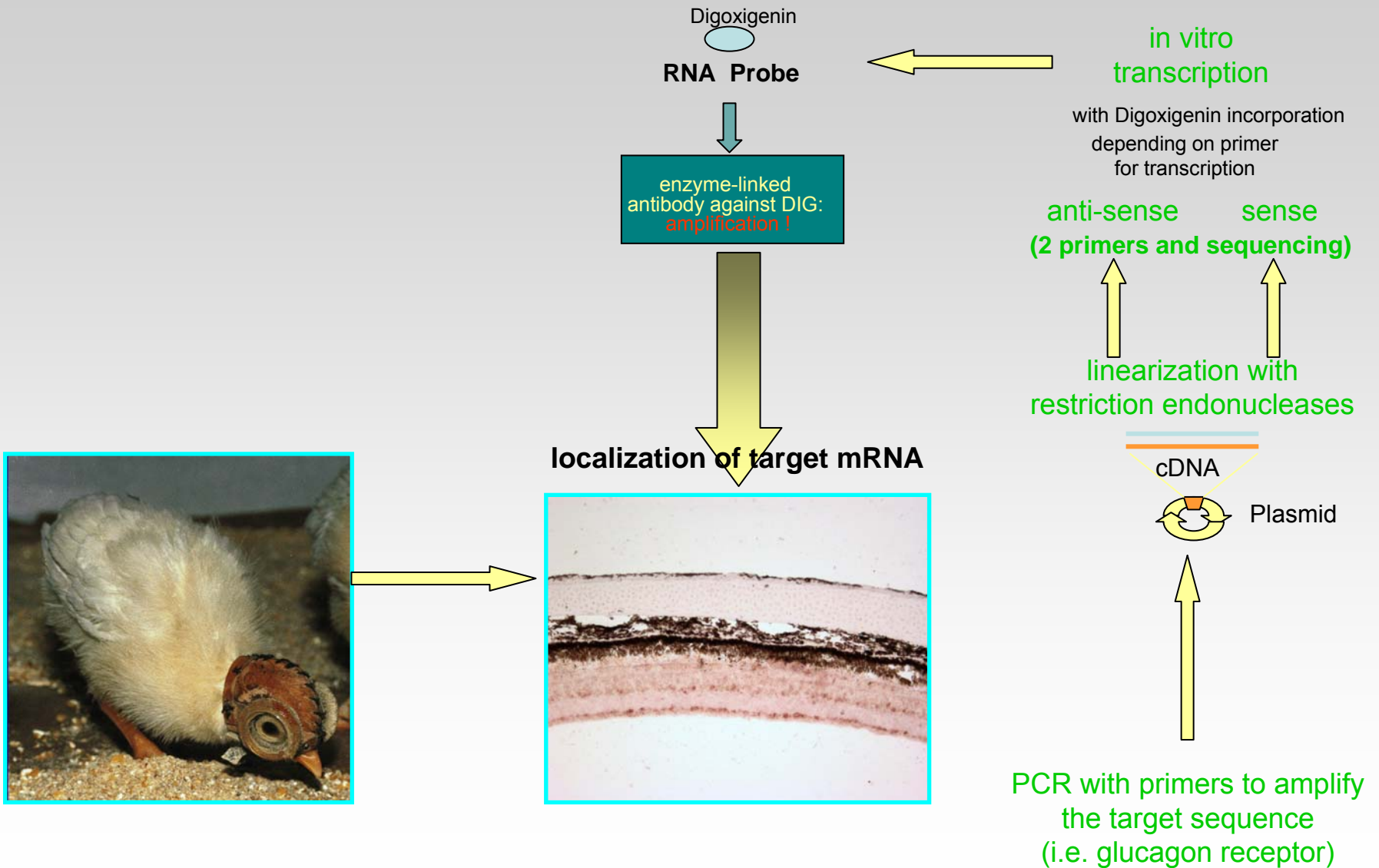


- compare effects of lenses at 1,2, 3

---

## Localizing the transcript in the tissue

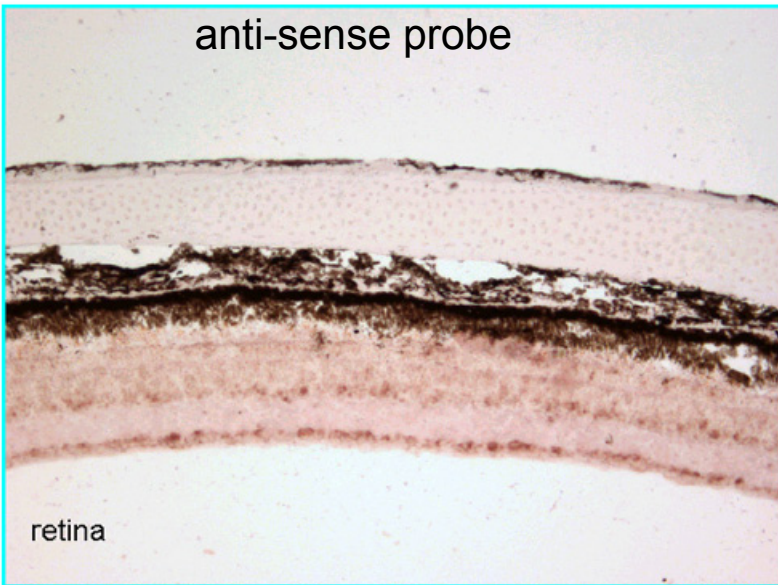
# Localisation of mRNA transcripts: In-situ hybridisation



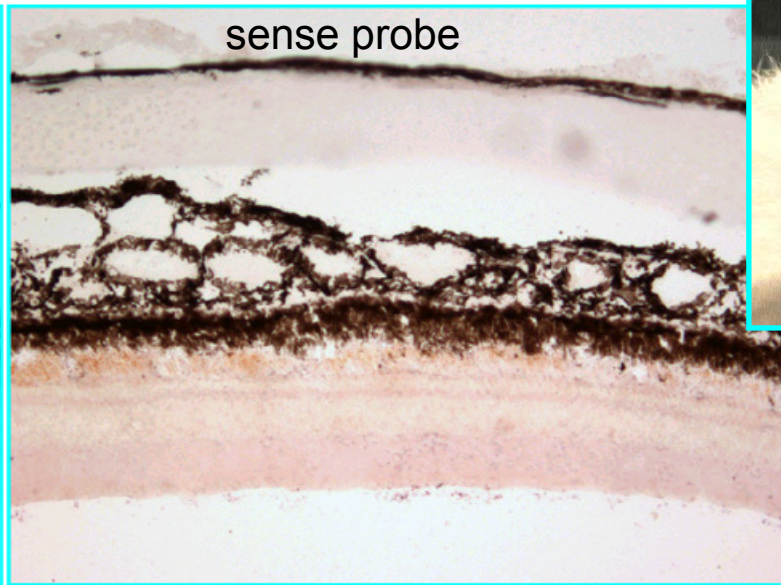


# In-situ hybridisation: the chicken glucagon receptor

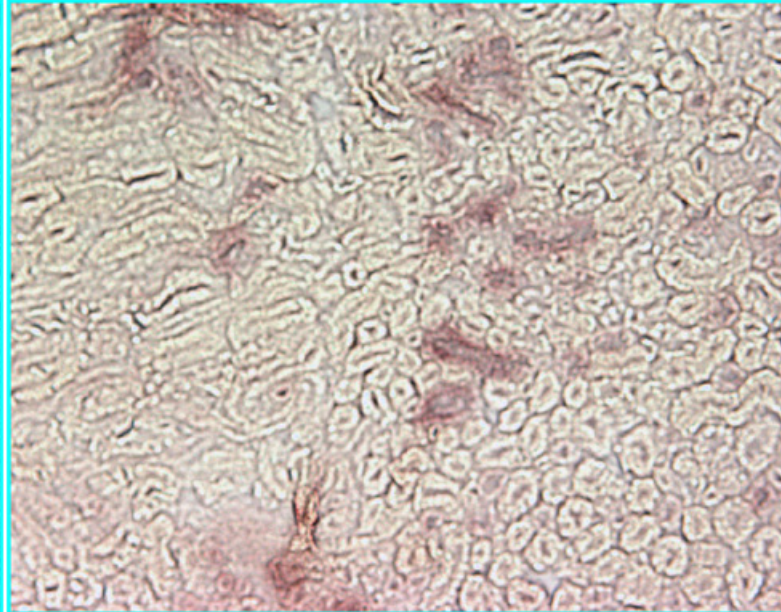
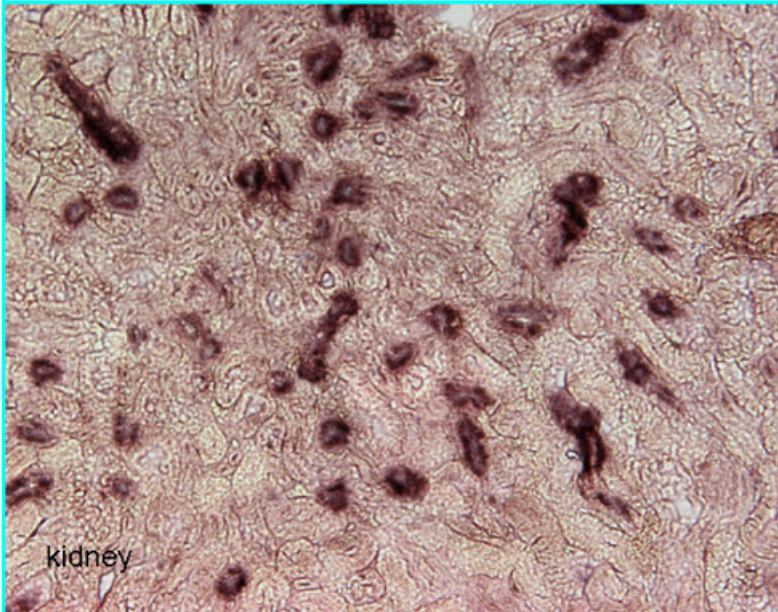
anti-sense probe



sense probe

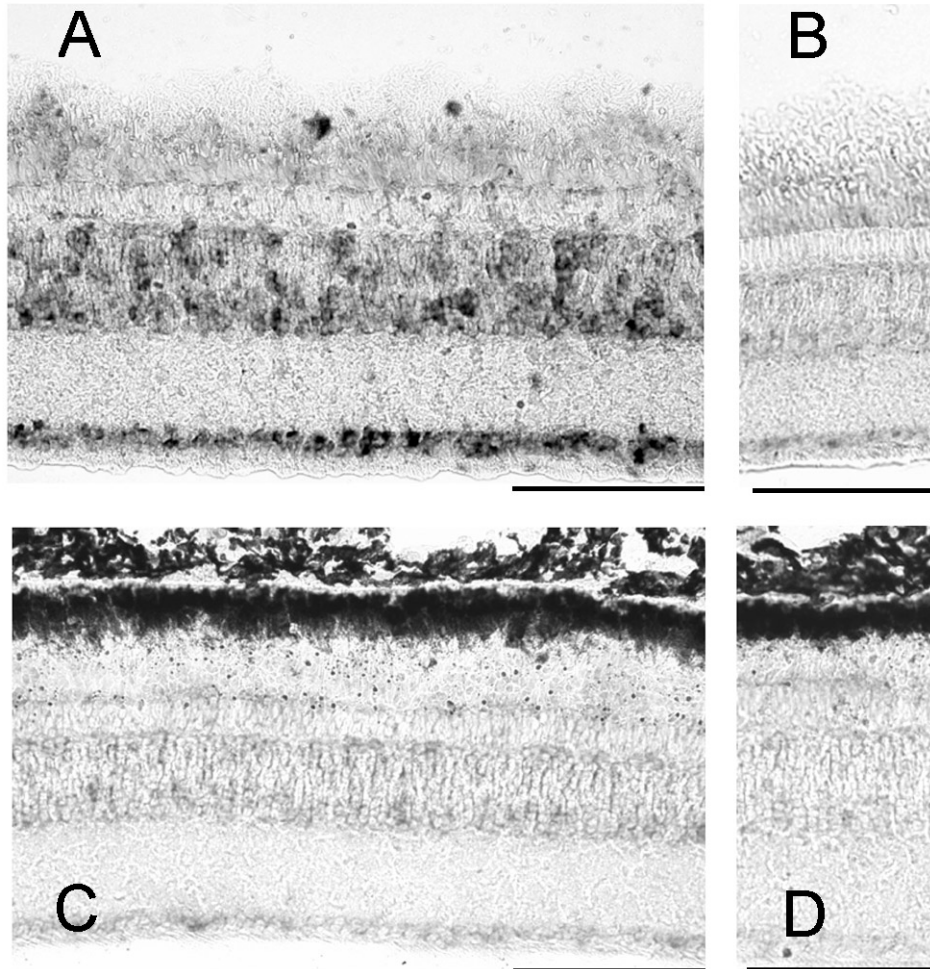


kidney



Burkhardt,  
Feldkaemper,  
Schaeffel, 2001

# In-situ hybridisation: the chicken glucagon receptor



(A) Glucagon receptor expression after injection of water, antisense probe. (B) Same section as in (A) labelled with the sense probe.

(C) Glucagon receptor expression after injection of a glucagon antagonist. (D) Same section as in (C) labelled with the sense probe.

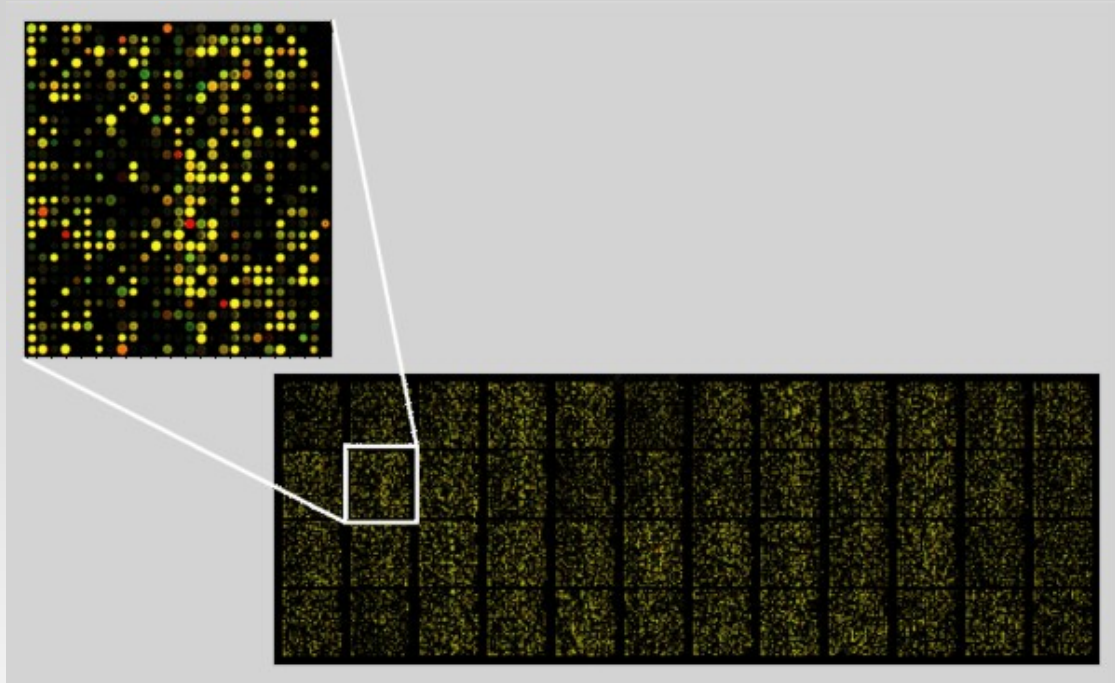
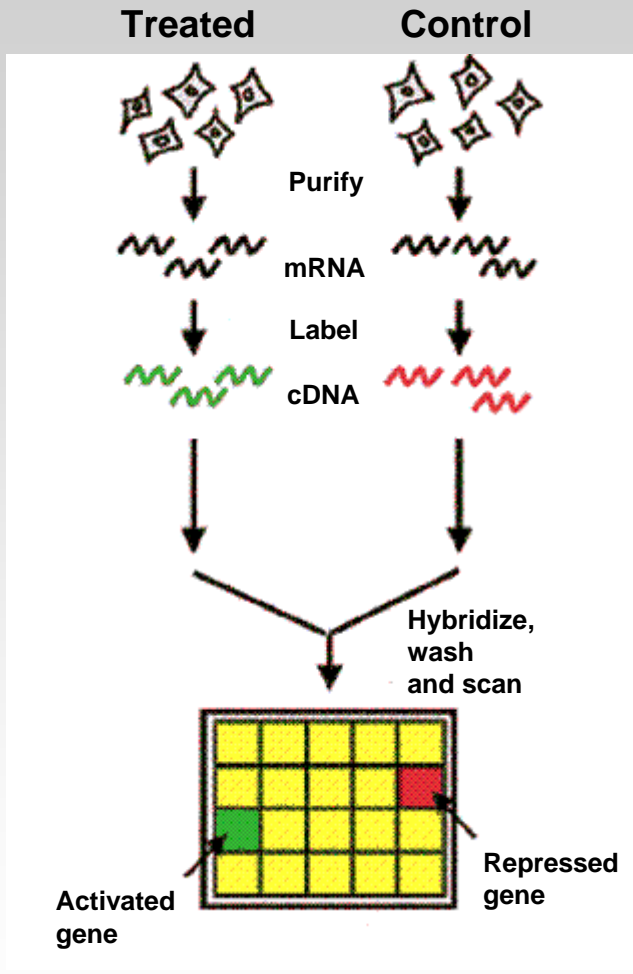
The scale bar represents 100  $\mu$ m.



---

Searching for new genes  
that are controlled by visual experience  
(e.g. defocus) or hearing loss

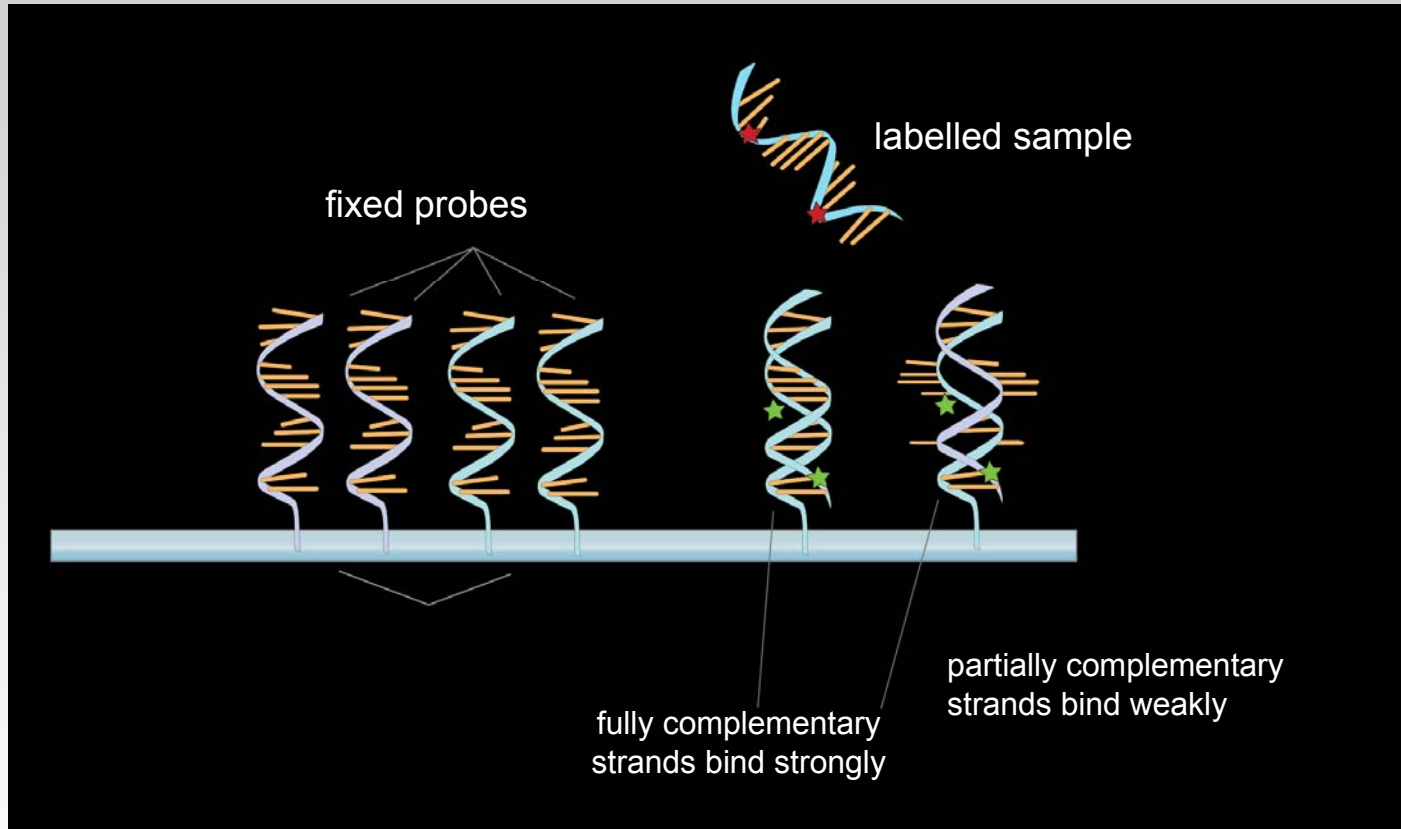
# Microarray Workflow



**Core principle** behind microarrays is hybridization between two DNA strands

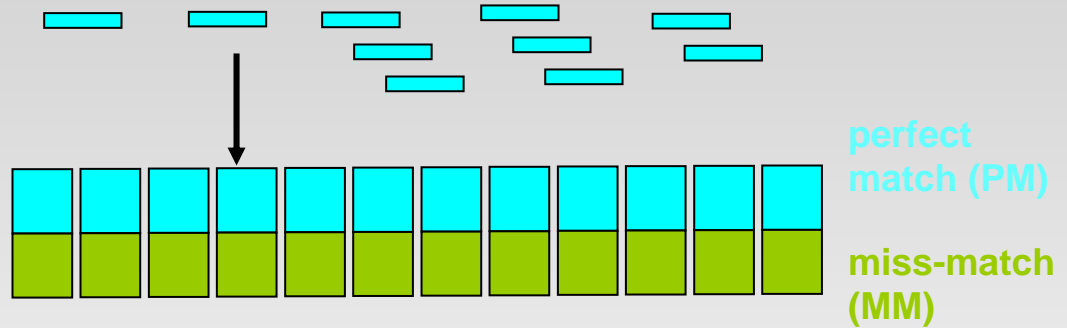
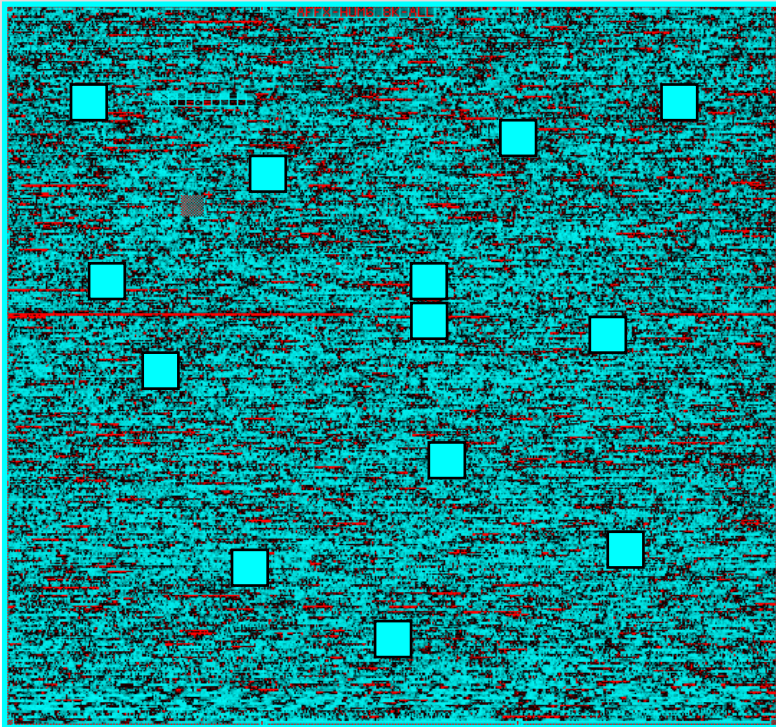
**Outcome:** Relative signal intensities for each transcript → Correlated to mRNA levels in the samples

# Hybridization of the target to the probe



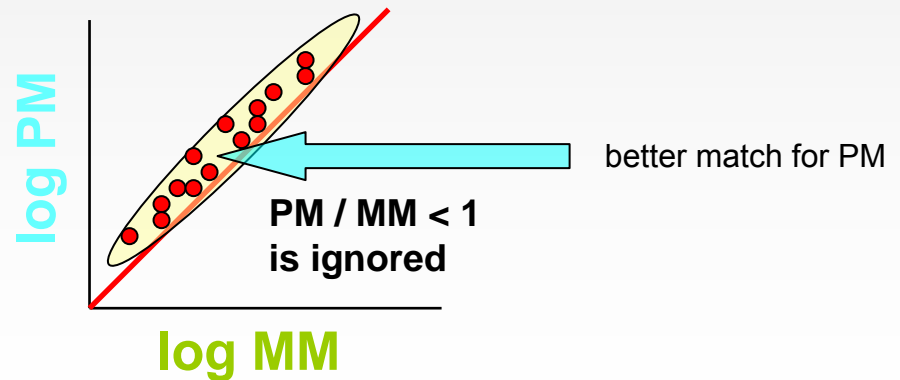
# Microarray Technique (Affymetrix chips)

5' 11 to 20 different oligonucleotide sequences complementary to different parts of the coding region of the gene (unique parts pref) AAAAAA

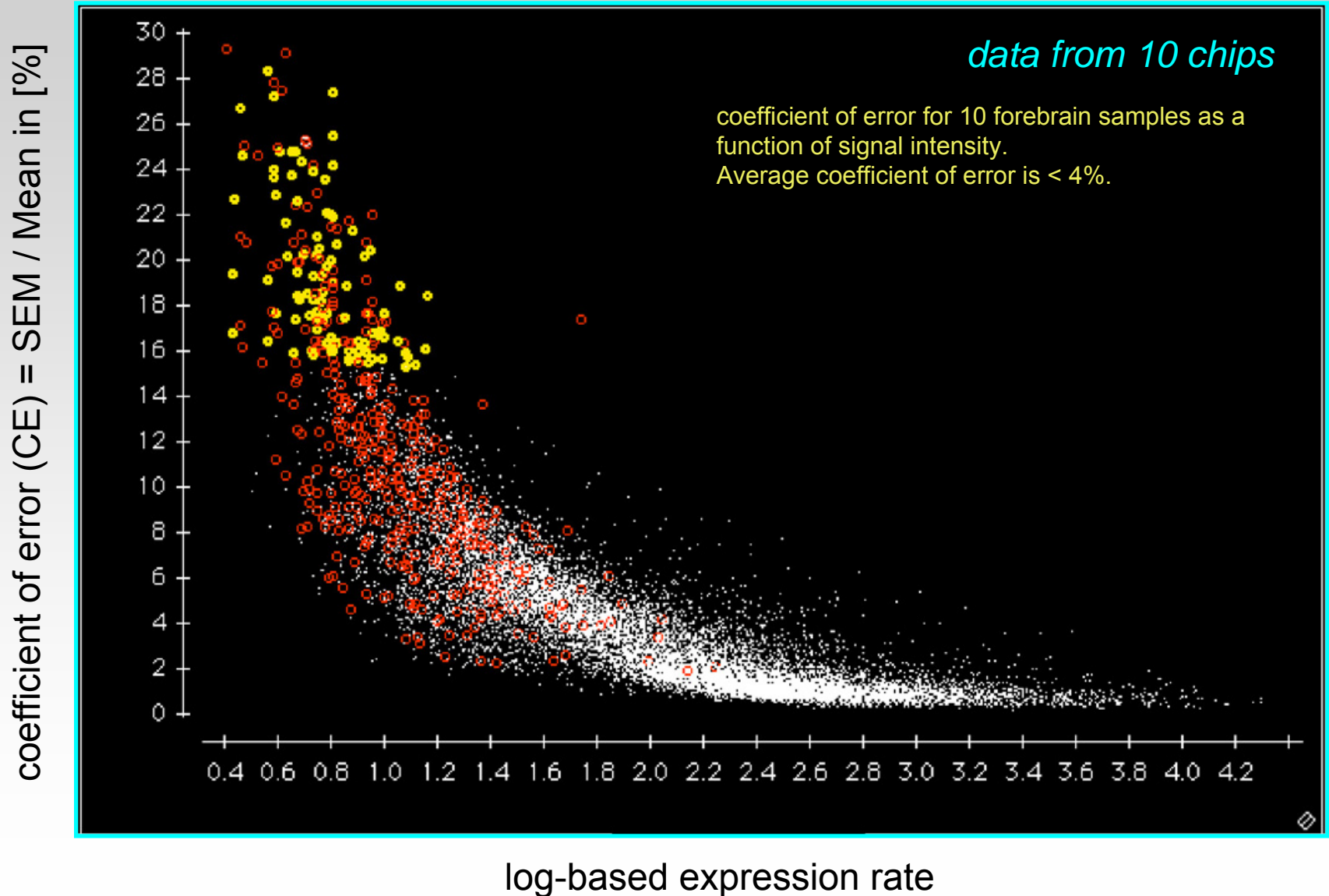


PM - 25 bases complementary to mRNA

MM - 13th base is AT or GC swap

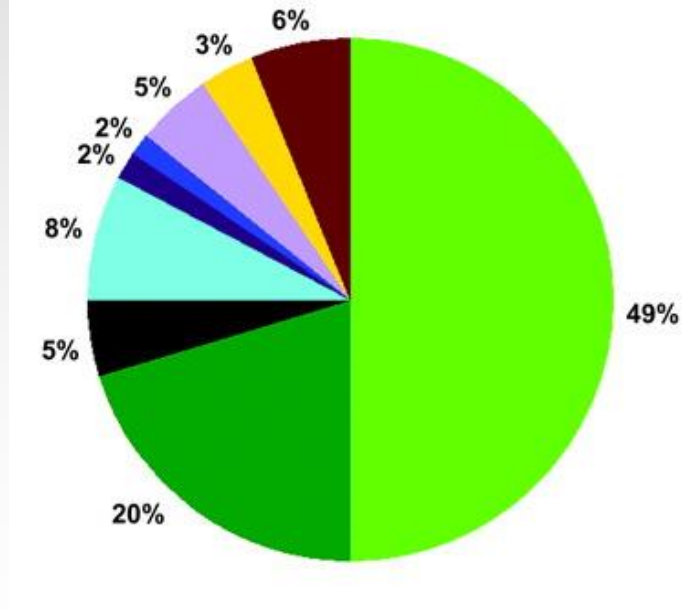


# Microarray technique - increase in detection reliability with expression rate

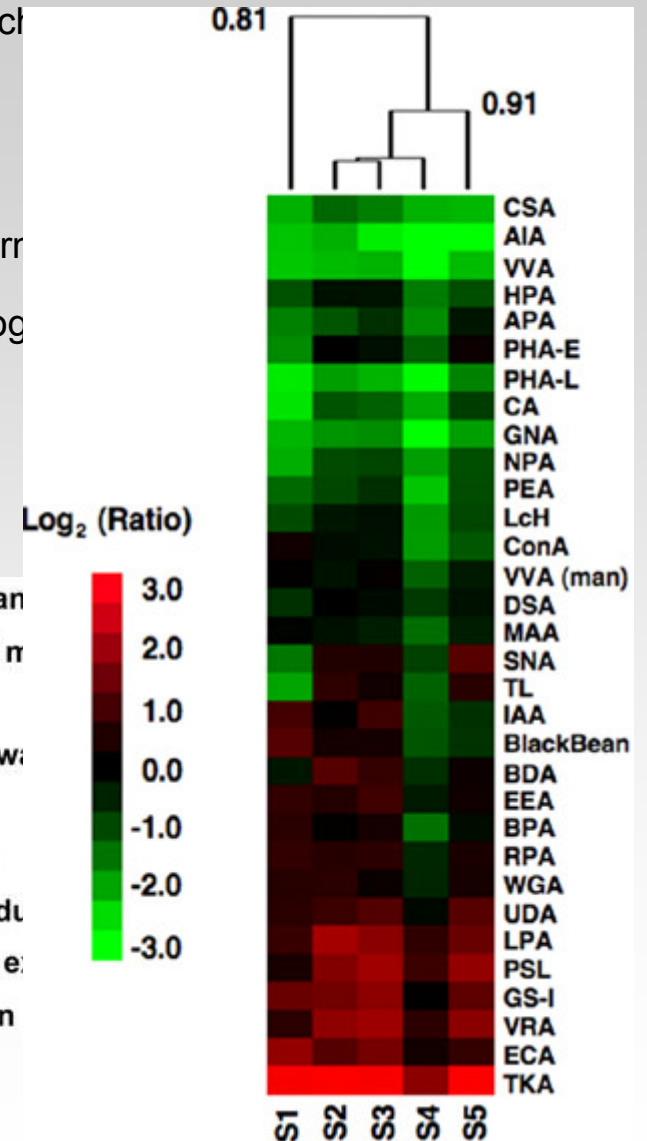


# Microarray Workflow

- Statistics: correction for multiple testing. Threshold-setting (fold-change) of transcripts
- Scaling/normalisation of the intensity
- Validation of selected transcripts (usually real-time PCR, Northern blot)
- Cluster the transcripts with respect to function (e.g. Gene Ontology: biological process, function, cellular component)
- Cluster the transcripts with respect to expression level



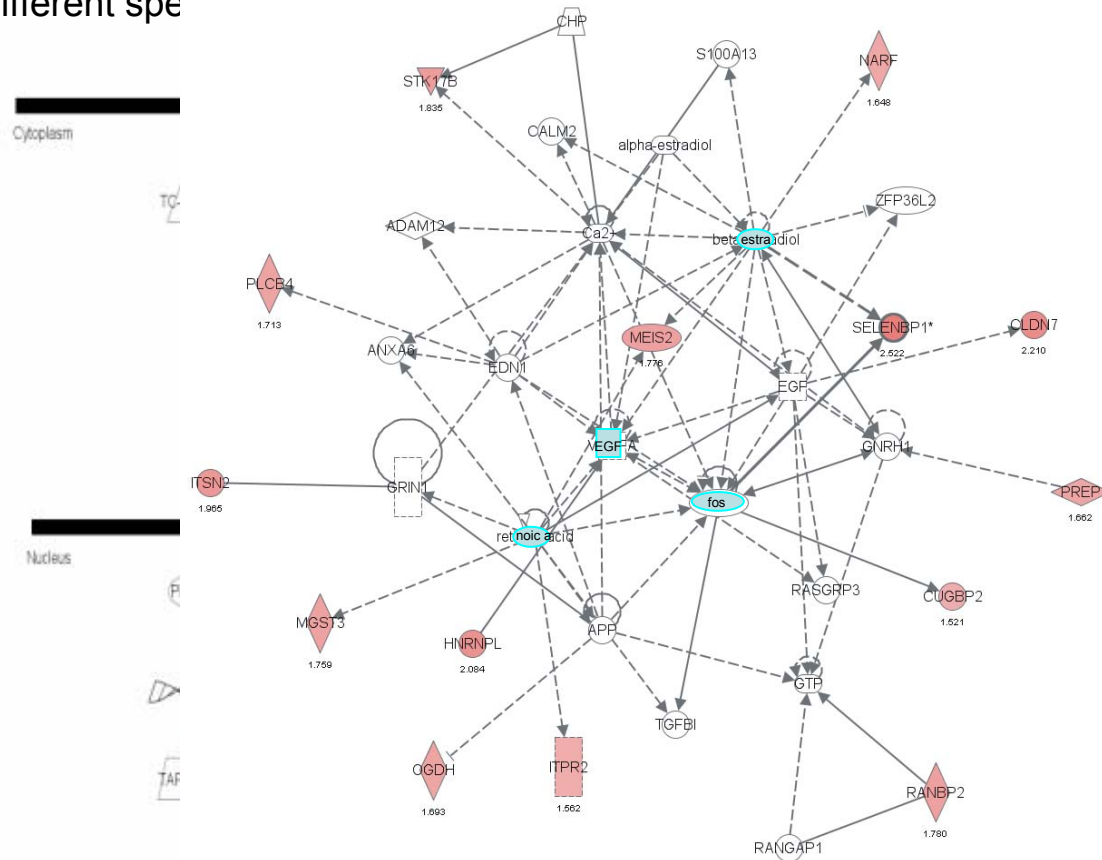
- cell growth and proliferation
- nucleic acid metabolism
- cell death
- energy pathways
- catabolism
- metabolism
- signal transduction
- response to environment
- cell adhesion





# Microarray Workflow

- **Canonical pathway** analysis: e.g. interferon signalling, glutamate receptor signalling
- **Network analysis**: creating large and highly complex graphic representations of cellular and molecular processes, and the integration of a variety of biological data, including RNA expression. Information is based on manually created databases of interactions, and metabolic and signalling pathways for different spe



# Databases

The screenshot shows the Gene Expression Omnibus (GEO) main page in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://www.ncbi.nlm.nih.gov/geo/>. The page features the NCBI logo on the left and the GEO logo on the right. Below the logos, there are navigation links for HOME, SEARCH, and SITE MAP, along with links for GEO Publications, FAQ, MIAME, and Email GEO. The page is titled "NCBI » GEO" and indicates the user is "Not logged in | Login".

The main content area includes a description of the Gene Expression Omnibus: "Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »".

The page is divided into several sections:

- GEO navigation:** This section contains two main categories: "QUERY" and "BROWSE". Under "QUERY", there are links for "DataSets" (with a search box containing "cochlea" and a "GO" button), "Gene profiles" (with a search box and a "GO" button), "GEO accession" (with a search box and a "GO" button), and "GEO BLAST". Under "BROWSE", there are links for "DataSets" and "GEO accessions". "GEO accessions" further branches into "Platforms", "Samples", and "Series".
- Site contents:** This section provides a overview of the site's data and documentation. It includes a "Public data" table with the following statistics:

Category	Count
Platforms	7,995
Samples	491,532
Series	19,536

Below this, there is a "Documentation" section with links for Overview, FAQ, Find, Submission guide, Linking & citing, Journal citations, Construct a Query, Programmatic access, DataSet clusters, GEO announce list, Data disclaimer, and GEO staff. A "Query & Browse" section includes links for Repository browser, Submitters, SAGEmap, FTP site, GEO Profiles, and GEO DataSets. Finally, a "Submit" section includes a link for New account.
- Submitter login:** This section contains a form for user login with fields for "User id:" and "Password:", a "LOGIN" button, and links for "» New account" and "» Recover password".

At the bottom of the page, there is a footer with links for "NLN | NIH | Email GEO | Disclaimer | Section 508". The browser's status bar at the bottom shows the system tray with the time 15:43 and a notification that "You saved: 125.4 m². All users saved: 209,043,806 m²."

<http://www.ncbi.nlm.nih.gov/geo/>



# Gene expression omnibus

[GSM109464](#): Cochlea\_WT\_9mo\_1  
[GSM109465](#): Cochlea\_WT\_9mo\_5  
[GSM109466](#): Cochlea\_WT\_9mo\_2  
[GSM109467](#): Cochlea\_D257A\_9mo\_1  
[GSM109468](#): Cochlea\_D257A\_9mo\_2

**2: GDS3028 record: Age-related hearing loss model: cochlea [ *Mus musculus* ]** [GEO Profiles](#), [Links](#)

**Summary:** Analysis of cochleas of 7 and 36 week old DBA/2J animals. The DBA/2J animal suffers from age-related hearing loss (AHL), a progressive disease characterized by an age-associated loss of hair cells and spiral ganglion cells in the cochlea. Results provide insight into the molecular basis of AHL.  
Parent Platform: [GPL339](#)  
Reference Series: [GSE6045](#)

**Type:** Expression profiling by array, count

**Subsets:** 2 age, 2 disease state sets.

**Supplementary Files:** [CEL download...](#)

**Samples:** 6

[GSM140222](#): Cochlea DBA2J 7 weeks 1  
[GSM140223](#): Cochlea DBA2J 7 weeks 2  
[GSM140224](#): Cochlea DBA2J 7 weeks 3  
[GSM140225](#): Cochlea DBA2J 36 weeks 1  
[GSM140226](#): Cochlea DBA2J 36 weeks 2  
[GSM140227](#): Cochlea DBA2J 36 weeks 3

**3: GDS2681 record: Caloric restriction effect on aged cochlea [ *Mus musculus* ]** [GEO Profiles](#), [Links](#)

**Summary:** Analysis of cochleas of 15 month old C57BL/6 (B6) animals on a calorie restricted diet. The B6 strain is a model of age-related hearing loss or presbycusis. Caloric restriction (CR) prevents late-onset presbycusis in B6. Results provide insight into the molecular basis of this effect of CR.  
Parent Platform: [GPL1261](#)  
Reference Series: [GSE4786](#)

**Type:** Expression profiling by array, count

**Subsets:** 2 age, 2 protocol sets.

**Supplementary Files:** [CEL download...](#)

**Samples:** 9

[GSM108106](#): Cochlea\_YC\_4mo\_1

# Gene expression omnibus

**GEO Accession viewer - Mozilla Firefox**

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM140222

Meist news KEGG sektion 123 Ecosia myopia Mail GO webmail UCSC GO Ontol. YFG LEO UniTu PubM Entrez Gene Homol Map Viewer analysis iHOP

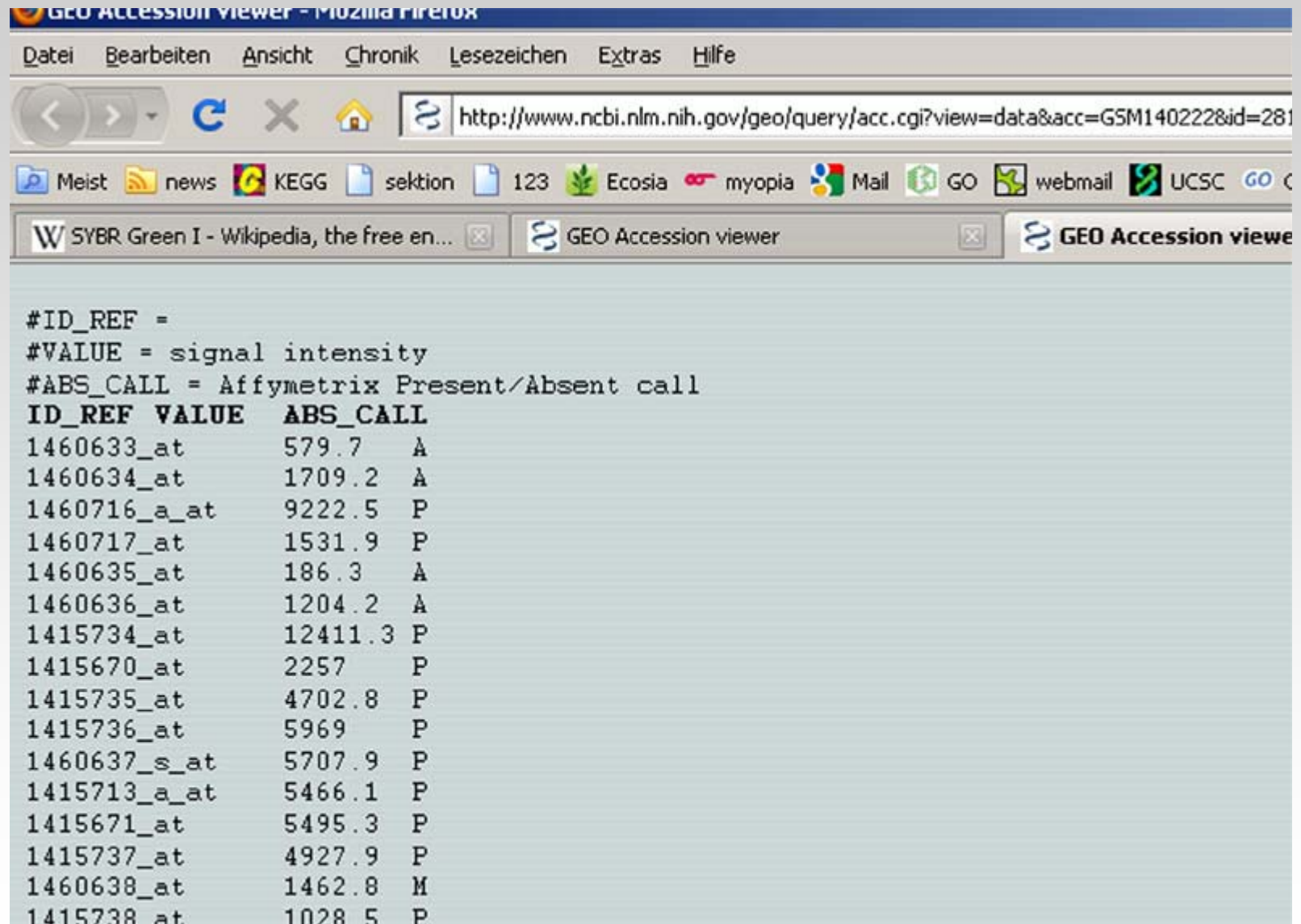
W SYBR Green I - Wikipedia, the free enc... GEO Accession viewer GDS Cluster Analysis Screenshot - Wikipedia

Source Name	RNA
Organism	<a href="#">Mus musculus</a>
Characteristics	Strain: DBA/2J mice Gender: male Age: 7-week-old Tissue: cochlea
Biomaterial provider	JAX
Treatment protocol	To identify genes associated with AHL, each 7-week-old sample (n = 3) was compared to each 36-week-old sample (n = 3), generating a total of nine pairwise comparisons.
Growth protocol	All mice were raised under 12:12 dark/light conditions, and housed in the the University of Tokyo-approved Animal Care Facility, and were provided a nonpurified diet and acidified water ad libitum.
Extracted molecule	total RNA
Extraction protocol	Total RNA was isolated from cochlea tissues using the TriZol reagent (Invitrogen) according to the manufacture's instructions.
Label	Biotin
Label protocol	cDNA synthesis from total RNA (Invitrogen), cRNA amplification and labeling (Enzo BioArray IVT kit) were accomplished according to the manufacture's instructions (Affymetrix).
Hybridization protocol	Hybridization (Affymetrix Hybridization oven 640), washing, and staining (Affymetrix Fluidics station 400) were accomplished according to the manufacture's instructions.
Scan protocol	Arrays were scanned on the Affymetrix GeneChip Scanner 3000. Image analysis was performed using the Affymetrix software.
Description	NA
Data processing	Scaled to 1500 (Affymetrix software)
Submission date	Oct 16, 2006
Last update date	Oct 17, 2006
Contact name	Shinichi Someya
E-mail(s)	<a href="mailto:someya@wisc.edu">someya@wisc.edu</a>
Phone	608-265-5205
Fax	608-262-2976
URL	<a href="http://www.genetics.wisc.edu/faculty/profile.php?id=140">http://www.genetics.wisc.edu/faculty/profile.php?id=140</a>
Organization name	University of Wisconsin-Madison
Department	Medical Genetics
Lab	Prolla Lab
Street address	425 Henry Mall Dr
City	Madison
State/province	WI
ZIP/Postal code	53706
Country	USA
Platform ID	<a href="#">GPL339</a>
Series (1)	<a href="#">GSE6045</a> Gene expression profile of the age-related hearing loss cochlea in DBA/2J mice

Fertig You saved: 125.4 m². All users saved: 209,043,806 m².

Start Posteingang - Microsoft ... Microsoft Excel - 170510i... Reference Manager - [R... GEO Accession viewer... Microsoft PowerPoint - [... Adobe Photoshop 16:46

# Gene expression omnibus



#ID\_REF =  
#VALUE = signal intensity  
#ABS\_CALL = Affymetrix Present/Absent call

ID_REF	VALUE	ABS_CALL
1460633_at	579.7	A
1460634_at	1709.2	A
1460716_a_at	9222.5	P
1460717_at	1531.9	P
1460635_at	186.3	A
1460636_at	1204.2	A
1415734_at	12411.3	P
1415670_at	2257	P
1415735_at	4702.8	P
1415736_at	5969	P
1460637_s_at	5707.9	P
1415713_a_at	5466.1	P
1415671_at	5495.3	P
1415737_at	4927.9	P
1460638_at	1462.8	M
1415738_at	1028.5	P

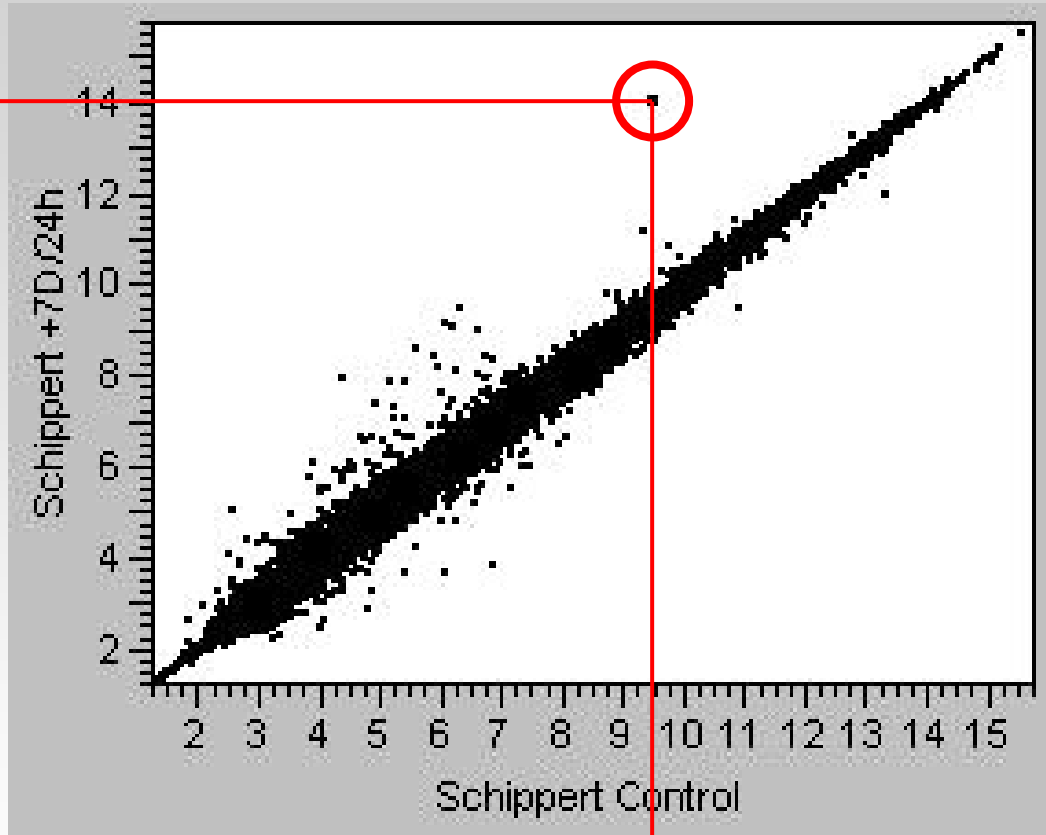
# Recent Experiments: Myopia Research

Tissue	Animal	Treatment	Duration	Author
* Retina	Monkey	Lid-fusion	2-4 months	<u>Tkatchenko et al.</u> <i>PNAS</i> 2006; 103:12
* Retina	Mouse	FDM	30min / 4h / 1d	<u>Brand et al.</u> <i>Mol. Vis.</i> 2007; 13
* Retina	KO-mouse	Age-difference WT vs. HM	30d / 42d -	<u>Schippert et al.</u> <i>Mol. Vis.</i> 2009; 15
* Retina	Chick	LIH	1d	<u>Schippert et al.</u> <i>Mol. Vis.</i> 2008; 14
Retinal AC	Chick	LIM and LIH	1d	<u>Ashby et al.</u> <i>IOVS</i> 2010; 51:7
* R/RPE	Chick	FDM	6h / 3d	<u>McGlenn et al.</u> <i>IOVS</i> 2007; 48:8
R/RPE	Chick	LIM and LIH	6h / 3d	<u>R. Stone</u> yesterday
* R/RPE/Ch	Chick	Recovery	1d / 4d	<u>Rada et al.</u> <i>Mol. Vis.</i> 2009; 15
R/RPE/Ch	Chick	Recovery	0h / 6h / 24h	<u>Giummarra et al.</u> <i>ARVO</i> 2010
* RPE/Ch	Marmoset	LIM vs. LIH	92d	<u>Shelton et al.</u> <i>Mol. Vis.</i> 2008; 14
RPE	Albino Chick	WT vs. HM	-	<u>Rymer et al.</u> <i>Exp. Eye Res.</i> 2007; 85
RPE	Chick	LIM	38d	<u>Zhang et al.</u> <i>ARVO</i> 2010
Sclera	Mouse	Age-difference	21d / 56d	<u>Zhou et al.</u> <i>IOVS</i> 2006; 47:5
Sclera	Human	no Treatment	-	<u>Young et al.</u> <i>Mol. Vis.</i> 2004; 10
Scleral FB	Human	Mechanical Stretching	30min / 24h	<u>Cui et al.</u> <i>Exp. Eye Res.</i> 2004; 78

\* Reviewed in "Gene profiling in experimental models of eye growth: Clues to myopia pathogenesis". Stone et al., *Vis. Res.* 2010

# Expression pattern comparison

LOG<sub>2</sub> of the mean relative signal intensities



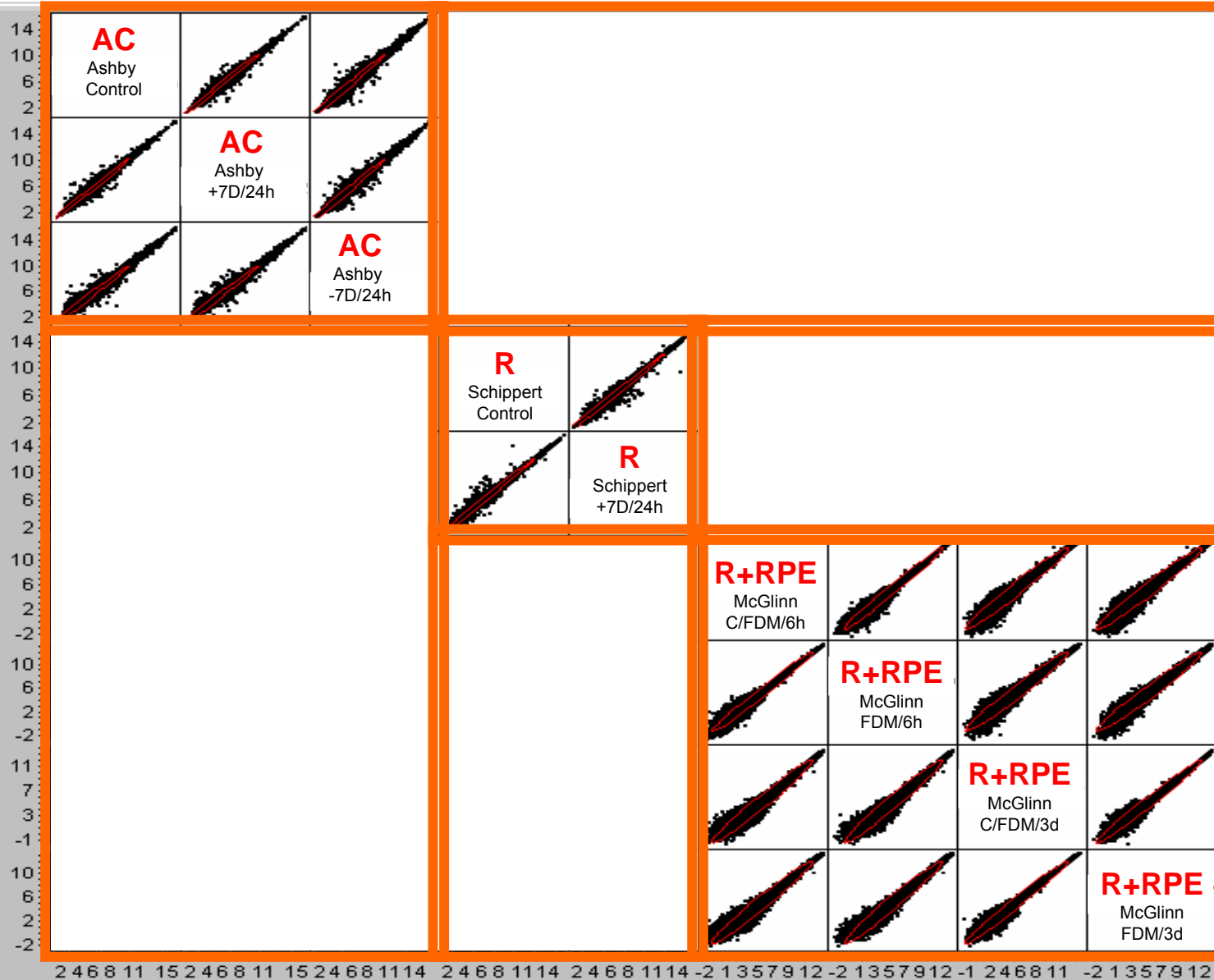
13.96  
15892

9.42  
686

Would mean that the signal intensity of this mRNA D-dopachrome tautomerase (enzyme involved in melanin synthesis) is 23-times higher in the retina of treated animals compared to control animals.

# Expression pattern comparison

LOG of the mean relative signal intensities



# Overlap between studies

Little overlap between studies so far can be explained by:

- Tissue type

Amacrine cells / Retina / Retina+RPE / Retina+RPE+Choroid / RPE / RPE+Choroid / Sclera

- Methods to induce eye growth / arrest eye growth

Lid suture / Deprivation myopia / Lens induced myopia / Lens induced hyperopia / Recovery / genetically modified animals

- Species

- Data analysis (normalisation, threshold-settings...)

# Recent Experiments, Outcomes

## Genes identified by lens or deprivation experiments:

### “Usual suspects”

- VIP (Tkatchenko et al. *PNAS* 2006, McGlinn et al. *IOVS* 2007, Ashby et al. *IOVS* 2010)
- ZENK or Egr-1 (Brand et al. *Mol.Vis.* 2007, Schippert et al. *Mol.Vis.* 2008, Ashby et al. *IOVS* 2010 )
- Ovotransferrin (Rada et al. *Mol.Vis.* 2009)
- Pre-pro glucagon (Schippert et al. *Mol.Vis.* 2008, Ashby et al. *IOVS* 2010)
- FGF2 (Shelton et al. *Mol.Vis.* 2008)

### New candidate genes

- Protein kinase Akt-2 (Brand et al. *Mol.Vis.* 2007)
- Oxoglutarate dehydrogenase (Schippert et al. *Mol.Vis.* 2009)
- Angiopoietin 2 (Ashby et al. *IOVS* 2010)
- Avian thymic hormone (Rada et al. *Mol.Vis.* 2009)
- Noggin (Zhang et al. *ARVO* 2010)
- Bone morphogenetic protein-2 and -7 (Cui et al. *Exp.Eye Res.* 2004, McGlinn et al. *IOVS* 2007, Zhang et al. *ARVO* 2010)
- Prepro-urotensin II-related peptide (McGlinn et al. *IOVS* 2007, Schippert et al. *Mol.Vis.* 2008)



# Recent Experiments, outcomes

---

## **Unexpected relationships:**

- Hint that RPE/choroid in LIH shares similarities to hypoxia-induced pathways  
(Shelton et al. *Mol. Vis.* 2008)
- Different pathways involved in LIM/LIH and in initiation / sustained eye growth in R/RPE  
(R.Stone, talk IMC 2010)

# Recent Experiments: cochlea after noise trauma

**Control versus noise-damaged cochleae:** (Margaret Lomax et al. (2001). Noise & Health 3 :19-35.)

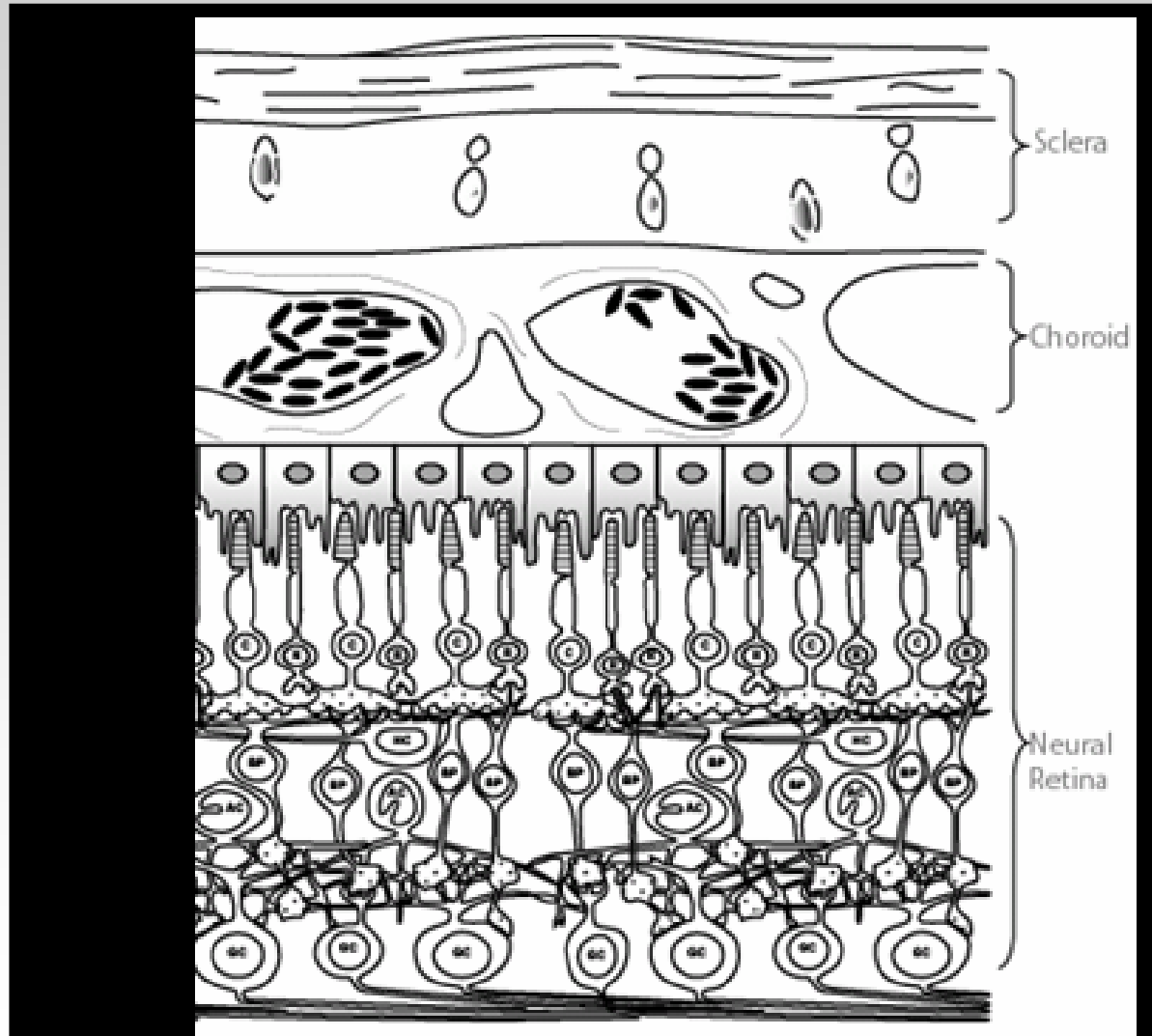
Early growth response genes:

egr-1

NGF-inducible anti-proliferative protein

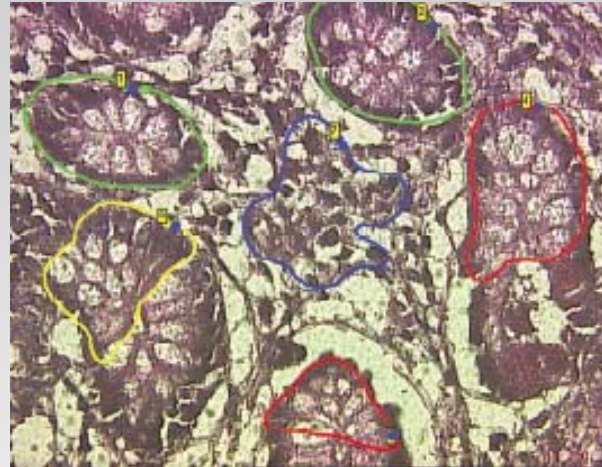
c-fos

# Problem: Retinal Cell Population is Heterogeneous

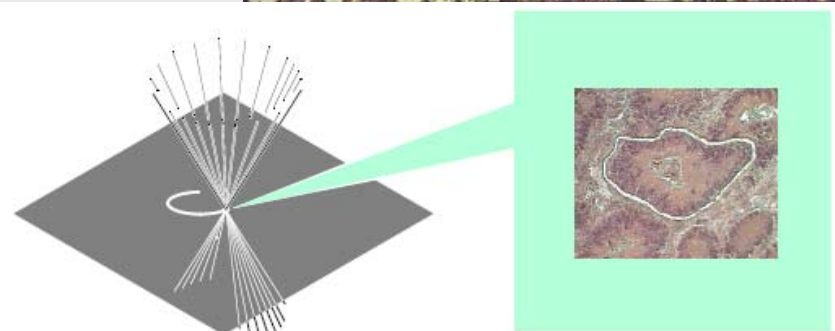


# Laser Capture Microdissection

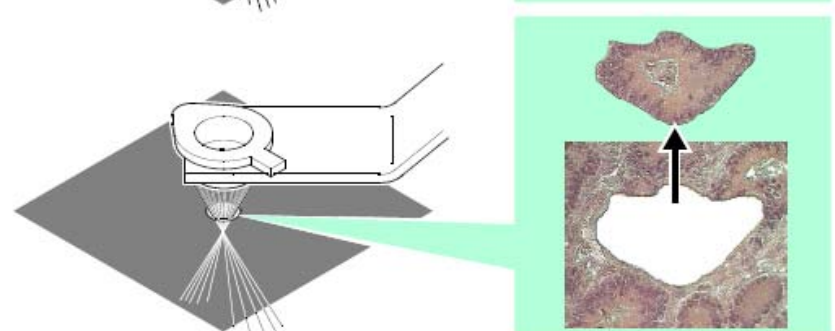
1. Draw new element that is to be extracted



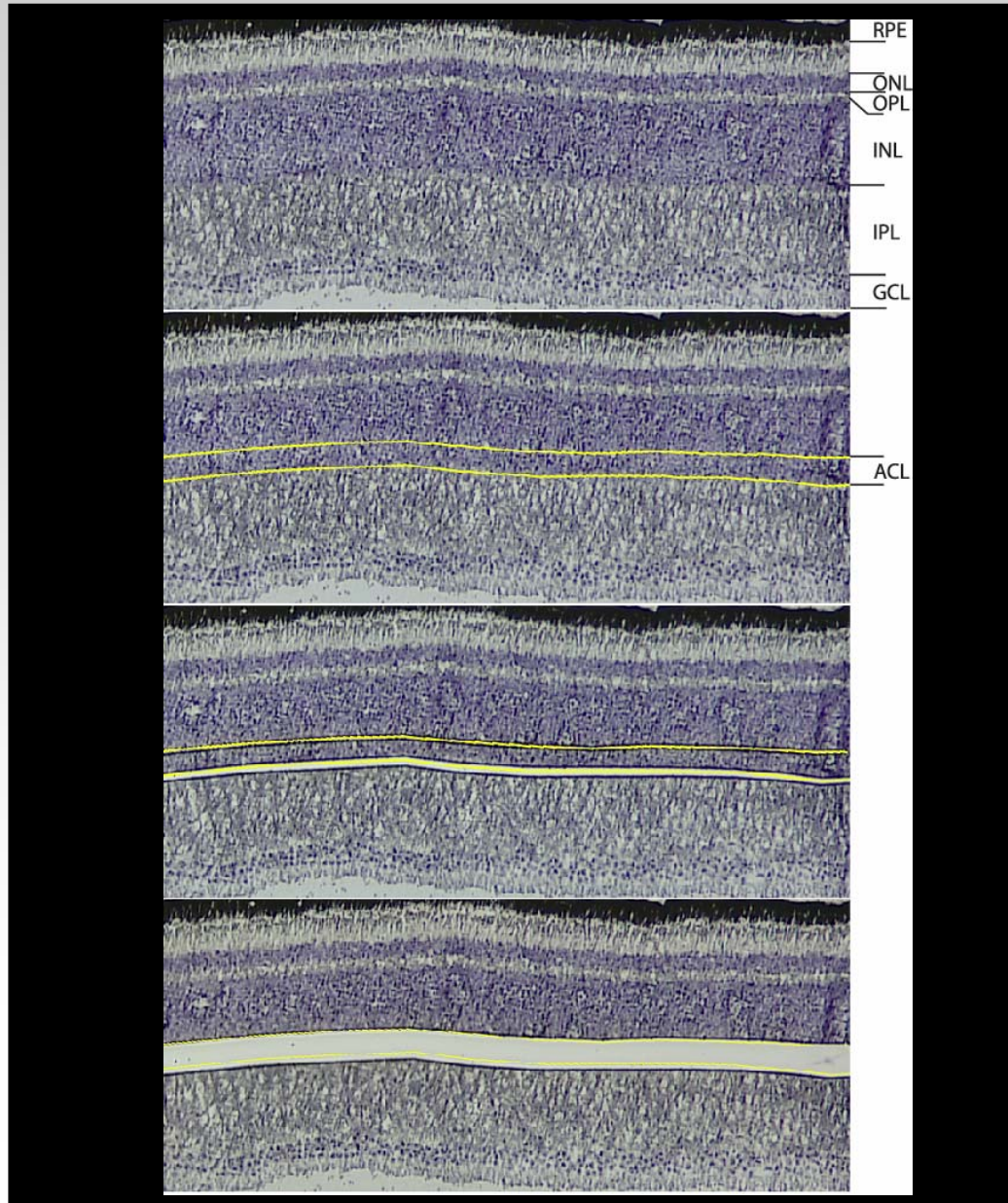
2. Initiate laser cutting



3. Catapult cut section into collection tube



# Laser Capture Microdissection



# Microarray and possibilities beyond Microarray

- Hundreds of potential signals have been identified with Microarrays, we “just” need to make sense of them
- Couple of potential signalling pathways have been proposed
- Concordance between studies is still poor, repetition using the same treatment paradigms is necessary

## Next generation sequencing (NGS)

Determines the sequence of thousands of DNA sequences in a single run (High throughput sequencing (35-400bp/transcript based on platform))

PRO:

- Very sensitive
- Quantitative
- No knowledge of genome required

CON:

- Expensive
- Complex sample preparation
- Data analysis very demanding

# Next generation sequencing

## “Seq-based” Functional Genomics Applications

ABI SOLiD

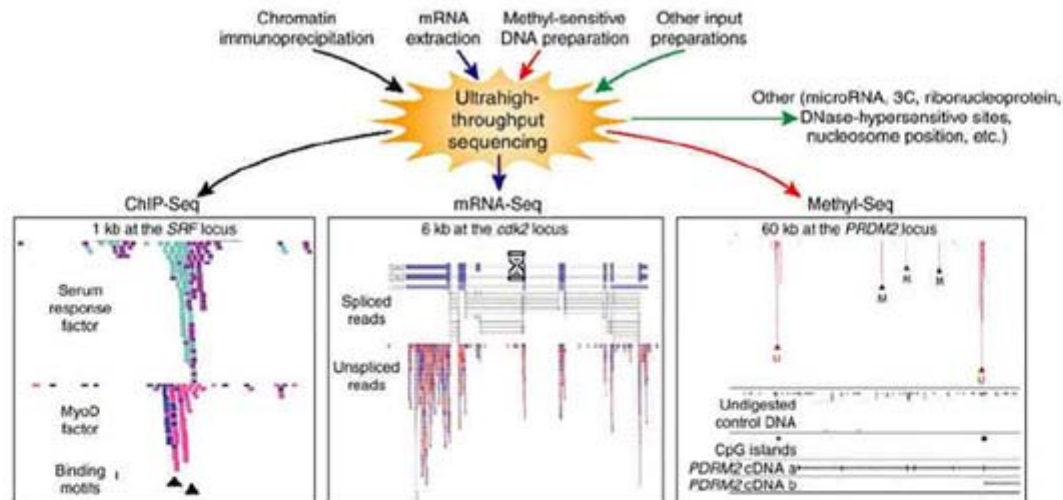


>200M 35-50bp reads/run, 1-256 samples/run

Illumina  
Genome Analyzer



>80M 35-70bp reads/run, 1-8 samples/run

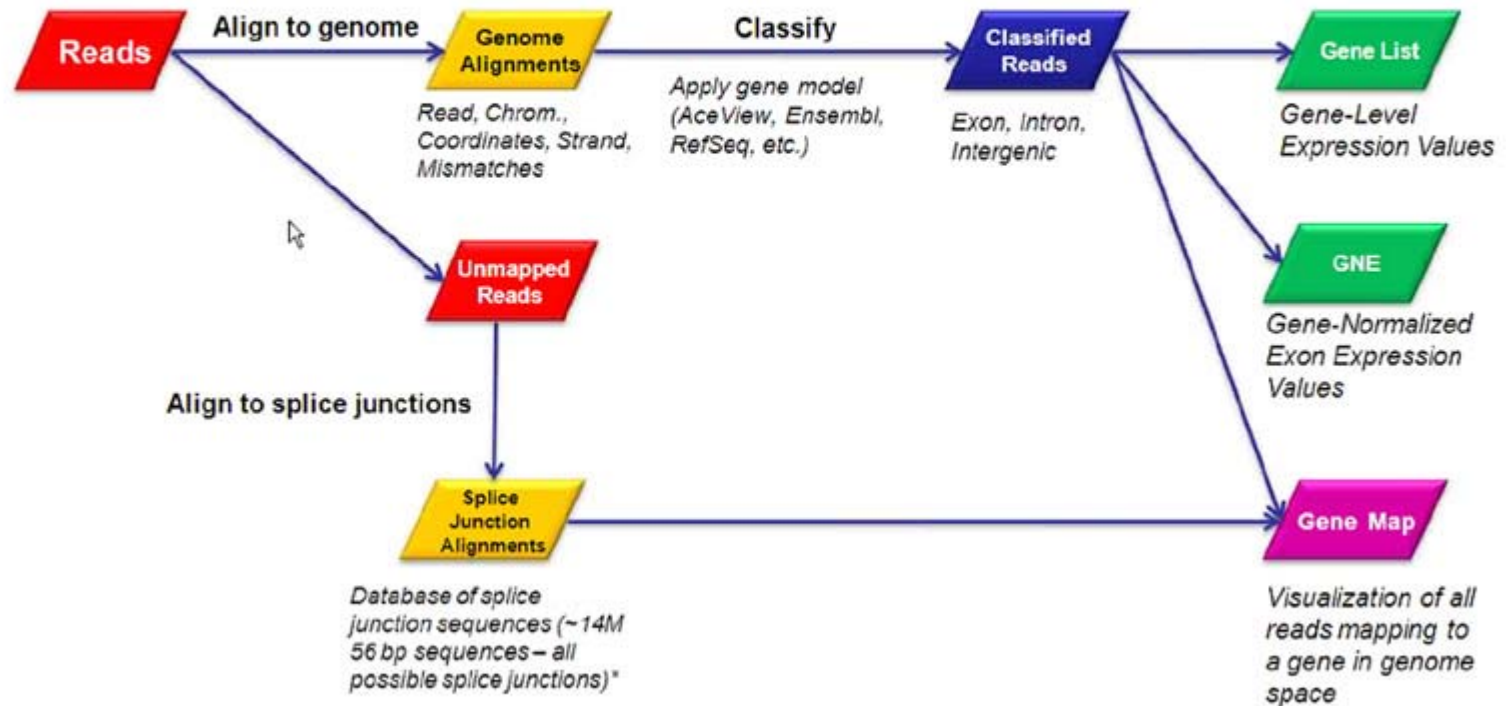


Sequence census methods for functional genomics. Wold B, Myers RM. Nat Methods. 2008 Jan;5(1):19-21.



# Next generation sequencing

## Read Processing Strategy

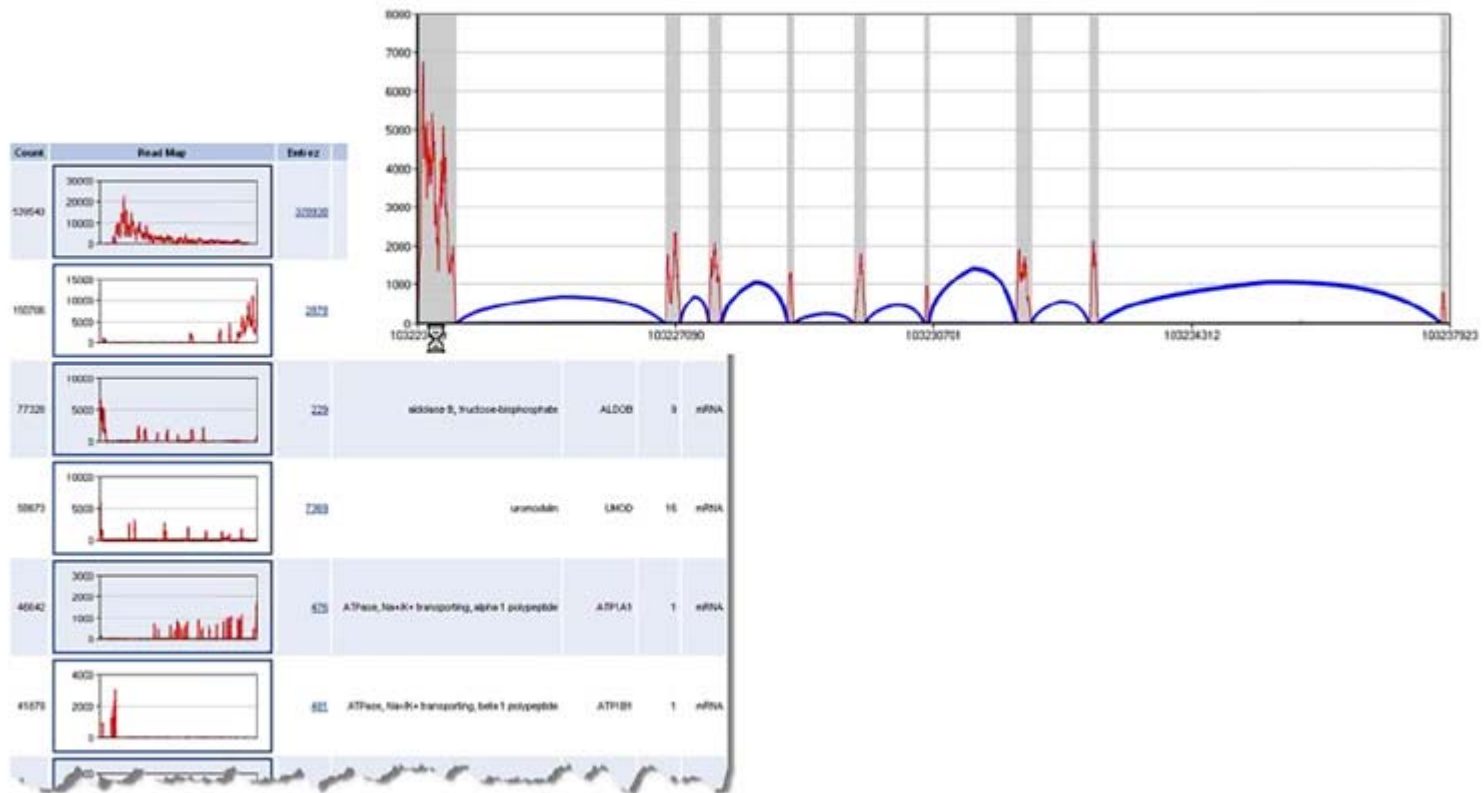


\*<http://genes.mit.edu/burgelab/mRNA-Seq/>



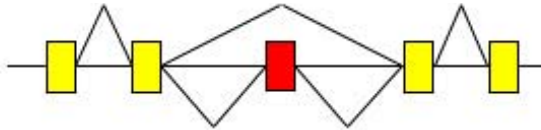
# Next generation sequencing

## Read Mapping and Splice Junction Details

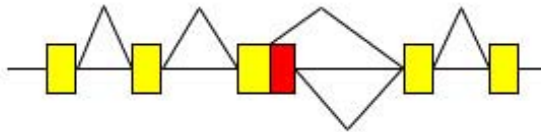


# Alternative splicing

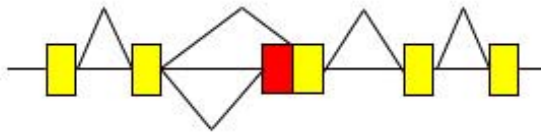
## Exon skipping



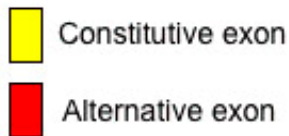
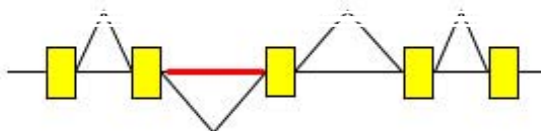
## Alternative 5' donor sites



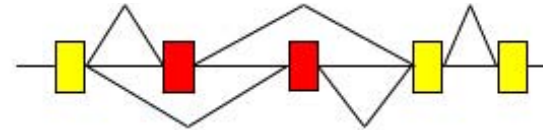
## Alternative 3' donor sites



## Intron retention



## Mutually exclusive exons



**Exon skipping:** an exon may be spliced out of the primary transcript or retained. This is the most common mode in mammalian pre-mRNAs.

**Mutually exclusive exons:** One of two exons is retained in mRNAs after splicing, but not both.

**Alternative donor site:** An alternative 5' splice junction (donor site) is used, changing the 3' boundary of the upstream exon.

**Alternative acceptor site:** An alternative 3' splice junction (acceptor site) is used, changing the 5' boundary of the downstream exon.

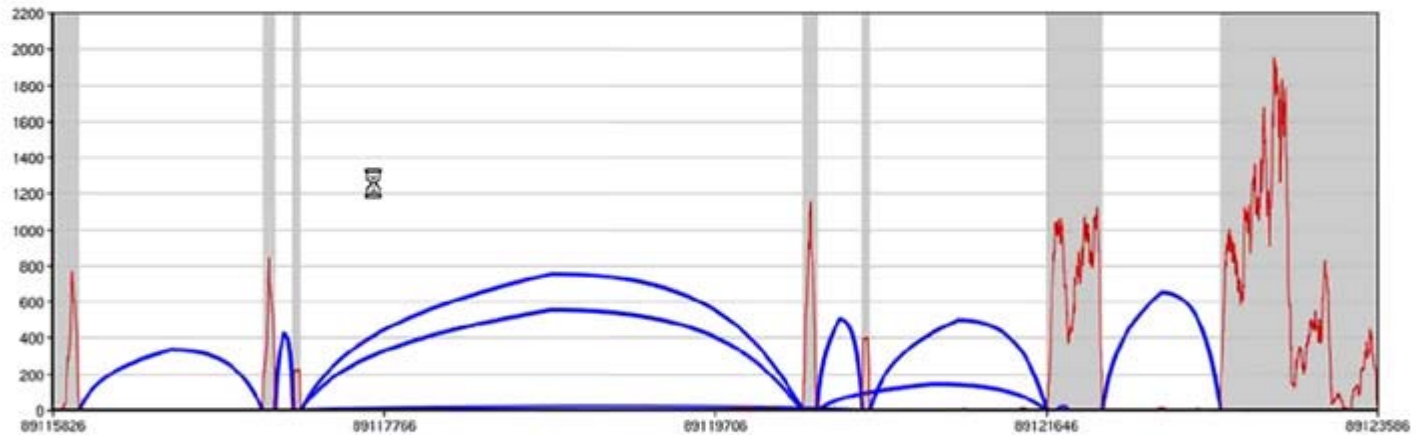
**Intron retention:** A sequence may be spliced out as an intron or simply retained. This is distinguished from exon skipping because the retained sequence is not flanked by introns. If the retained intron is in the coding region, the intron must encode amino acids in frame with the neighboring exons, or a stop codon or a shift in the reading frame will cause the protein to be non-functional. This is the rarest mode in mammals.

# Next generation sequencing

## Read Mapping with Splice Junctions

### Secreted phosphoprotein 1 (SPP1)

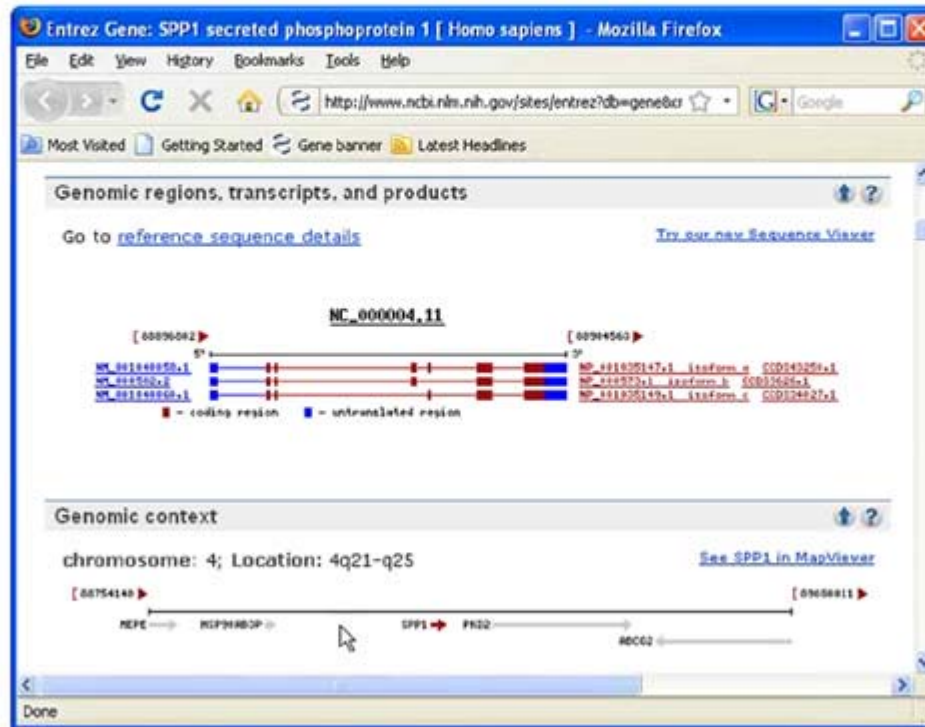
6696



# Next generation sequencing

## Gene Mapping with Splice Junctions

### Secreted phosphoprotein 1 (SPP1)



# Next generation sequencing

## Sequences to Biological Interpretation – Gene Level Expression Values



```
>2357_2012_1445_F3
T21102021110100012222022010110002001
>2357_2012_1665_F3
T21111033030212120001202212232122111
>2357_2012_1761_F3
T21310211201210302111012122022023222
>2357_2012_1816_F3
T10022001210300222220222100011122201
>2357_2012_1835_F3
T10112311210021222202201220101321322
...
```

### Secondary Analysis

Alignments  
Annotation  
Summaries

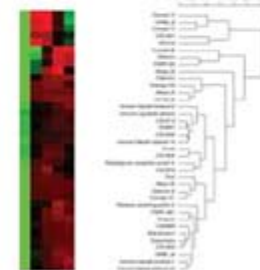


# Reads	Reference Definition
929872	hsa.mf.117a MMAT0002802 Homo sapiens mR-517a, g2c.mf.117a
109913	hsa.mf.43i MMAT0003771 Galba galba mR-43i, cap.mf.43i MMAT0003021 Onchocerca volvulus anabius mR-43i
137507	hsa.mf.113.3a MMAT0003937 Homo sapiens mR-131-3a, g2c.mf.113
209623	hsa.mf.12j MMAT0002954 Homo sapiens mR-12j, g2c.mf.12j
140329	hsa.mf.23c MMAT0000681 Homo sapiens mR-23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c, g2c.mf.23c
131498	hsa.mf.117i MMAT0001888 Homo sapiens mR-117i, g2c.mf.117a



### Tertiary Analysis


Comparison Statistics  
Clustering  
Pathway Analysis





# Next generation sequencing

## GeneSifter – Differentially Expressed Genes

Main (login: gzae\_demo\_4) > Analysis > Pairwise > Results 

Pairwise Analysis: SOLID [Reports: [Ontology](#) | [KEGG](#) | [Chromosome](#) | [Scatter Plot](#) | [Volcano Plot](#)] [Results: [Export](#) | [Save](#)]

	Group 1	Group 2
Conditions:	Wild-type oocyte	Dicer- KO oocyte
Experiments:	77100, 77101	77102, 77103
Significance:	1.5, t-test	
Normalization:	None	
Quality Cutoff:	1	
Data Transformation:	Log Transformed	

Show:  Sort By:  p Cutoff:  Threshold:   (478 results found) [1 - 20] [21 - 40]

No.	Ratio	p-value	Identifier	Gene Name
1	▲ 1883.43	0.00043	XM_001476651	PREDICTED: hypothetical LOC672511 (LOC672511), mRNA
2	▲ 1883.43	0.00043	XM_001477846	PREDICTED: Mus musculus hypothetical LOC672511 (LOC672511), mRNA
3	▲ 1566.95	0.00216	NM_146026	ATPase family, AAA domain containing 4, mRNA (cDNA clone MGC:37571 IMAGE:4988054
4	▲ 1149.36	0.00279	XM_001475719	RIKEN cDNA C130026121 gene, mRNA (cDNA clone MGC:8305 IMAGE:3593825)
5	▲ 906.53	1.98e-05	NM_024269	ADP-ribosylation factor-like 2 binding protein (Arl2bp), transcript variant 1, m
6	▲ 782.87	0.00049	NM_007574	Complement component 1, q subcomponent, C chain (C1qc), mRNA
7	▲ 751.60	0.00297	NM_054088	Patatin-like phospholipase domain containing 3, mRNA (cDNA clone MGC:41626 IMAGE
8	▲ 617.79	0.00074	NM_007572	Complement component 1, q subcomponent, alpha polypeptide, mRNA (cDNA clone MGC:
9	▲ 447.37	9.91e-05	XM_001475462	PREDICTED: similar to Kelch-like 2, Mayven (Drosophila) (LOC100040903), mRNA
10	▼ 295.34	0.00330	XM_001001076	PREDICTED: immunoglobulin heavy chain (gamma polypeptide), transcript variant 1
11	▲ 285.46	0.00291	NM_175219	RIKEN cDNA C130026121 gene (C130026121Rik), transcript variant 2, mRNA
12	▲ 262.07	0.00045	NM_016850	Interferon regulatory factor 7 (Irf7), mRNA
13	▲ 242.89	0.00666	NM_001025377	Rho GTPase activating protein 15 (Arhgap15), transcript variant 1, mRNA
14	▲ 231.91	5.54e-05	NM_001037909	RIKEN cDNA C130026121 gene (C130026121Rik), transcript variant 2, mRNA
15	▲ 217.33	0.00201	NM_011671	Uncoupling protein 2 (mitochondrial, proton carrier), mRNA (cDNA clone MGC:13955
16	▲ 209.69	3.68e-05	NM_011662	TYRO protein tyrosine kinase binding protein (Tyrobp), mRNA
17	▲ 181.17	0.00052	NM_010185	Fc receptor, IgE, high affinity I, gamma polypeptide, mRNA (cDNA clone MGC:36077
18	▲ 166.00	0.00784	NM_007815	Cytochrome P450, family 2, subfamily c, polypeptide 29, mRNA (cDNA clone MGC:299
19	▼ 165.62	0.00339	XM_001473733	PREDICTED: Mus musculus hypothetical protein LOC100039892 (LOC100039892), mRNA
20	▲ 161.72	0.00040	NM_027836	Membrane-spanning 4-domains, subfamily A, member 7, mRNA (cDNA clone MGC:36243 I

Show:  Sort By:  p Cutoff:  Threshold:   (478 results found) [1 - 20] [21 - 40]



# Next generation sequencing

## Differential Splicing

Main (login: eric\_test) > Analysis > Pairwise > Results > Gene Summary

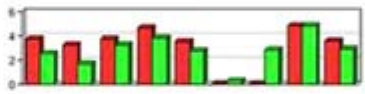
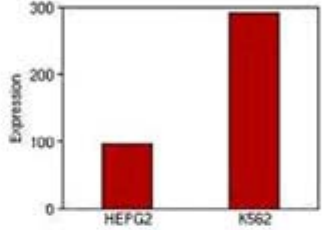
### » Gene Summary: Cathepsin C

• **By Group**

Group	Condition	N	Mean	SEM	SEM/Mean	Quality Mean
1	HEPG2	1	95.7900	-	-	95.7900
2	K562	1	291.460	-	-	291.460

• **By Target**

Group	Sample	Expression	Quality
1	s_002067_sequence	95.7900	95.79
2	s_007332_sequence	291.460	291.46



### » One-Click Gene Summary™

Accession No.: [NM\\_148170](#)  
Cluster ID: [Hs.128065](#)  
UG Title: [Cathepsin C](#)  
Gene ID: [CTSC](#)  
Homologene: [Mm.322945](#), [Rn.122504](#)  
Chromosome: [11](#)  
Cytoband: [11q14.1-q14.3](#)  
Seq Count: [948](#)  
Entrez Gene: [1075](#)  
Gene Name: [cathepsin C](#)  
Synonyms: [CPPI](#) | [DPP1](#) | [DPPI](#) | [HMS](#) | [JP](#) | [JPO](#) | [PALS](#) | [PLS](#)  
ONIM: [170650](#)  
RefSeq mRNA: [NM\\_148170](#) (FASTA)  
RefSeq Prot: [NP\\_680475](#) (FASTA)  
Summary: The protein encoded by this gene, a member of the peptidase C1 family, is a lysosomal cysteine proteinase that appears to be a central coordinator for

**Gene Ontologies:**

**Biological Process**

- immune response
- proteolysis

**Molecular Function**

- chloride ion binding
- cysteine-type endopeptidase activity
- cysteine-type peptidase activity
- peptidase activity

**Cellular Component**

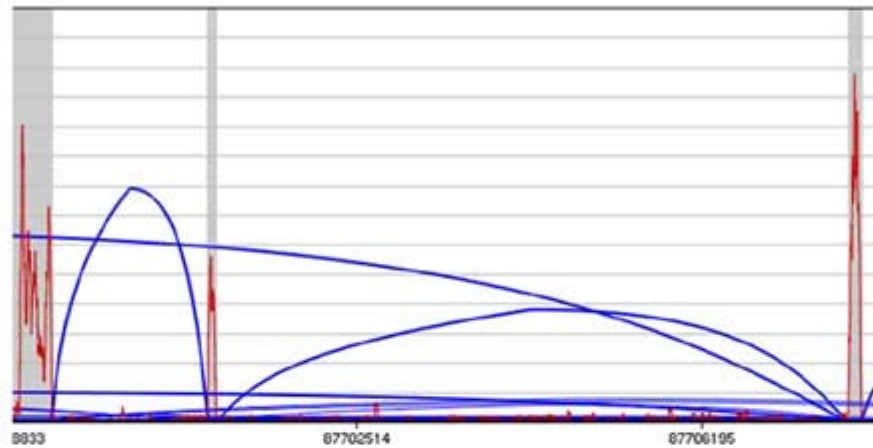
- lysosome



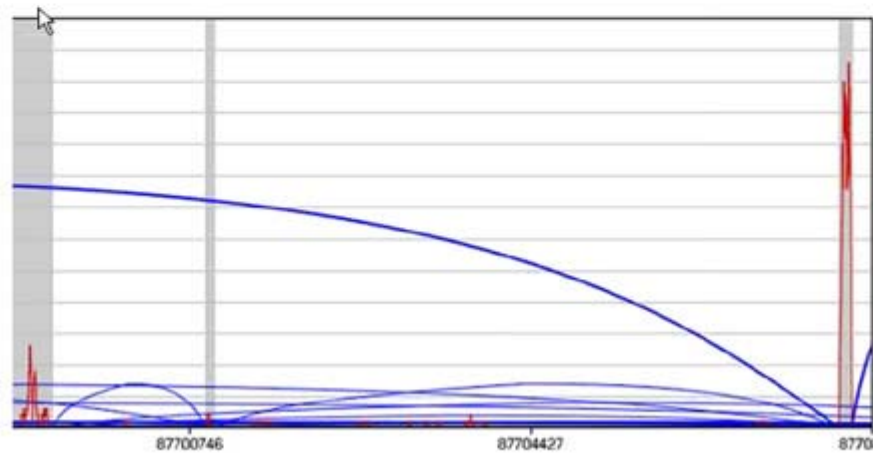
# Next generation sequencing

## Differential Splicing – Cathepsin C

HEPG2



K562



# Next generation sequencing

- Alternative to microarrays
- Can be used to profile mRNA as well as non-protein coding RNAs without prior information
- Can survey the entire transcriptome, including novel, un-annotated regions
- Can determine gene structure and isoform levels by mapping read density across transcript or gene and can use reads to map splice junctions
- Can identify and quantify both rare and common transcripts, with over six orders of magnitude of dynamic range.

# Conclusions

techniques of molecular biology can be used to:

- identify (PCR) and characterize genes whose transcription is controlled by specific visual experience (defocus) (northern blot, quantitative PCR)
- identify gene products that could provide potential targets for pharmacological intervention of myopia (microarray)
- survey the entire transcriptome, including novel, un-annotated regions (next generation sequencing)

*(not complete)*