



The Way Ahead.™

# Analysis of Gene Regulatory Networks by ChIP-on-chip

*Dominic Schmidt*

*Max Planck Institute for Molecular Genetics*

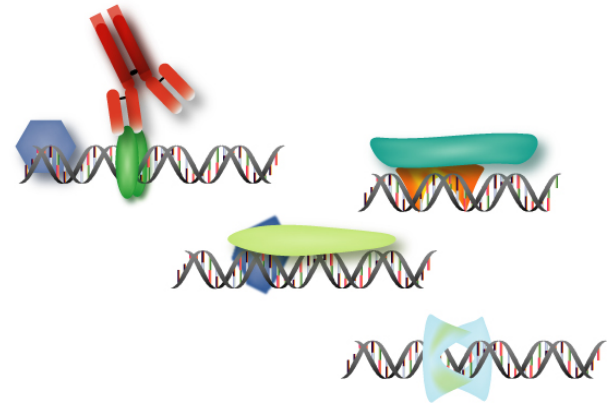
1st European ChIP-on-chip Scientific Forum





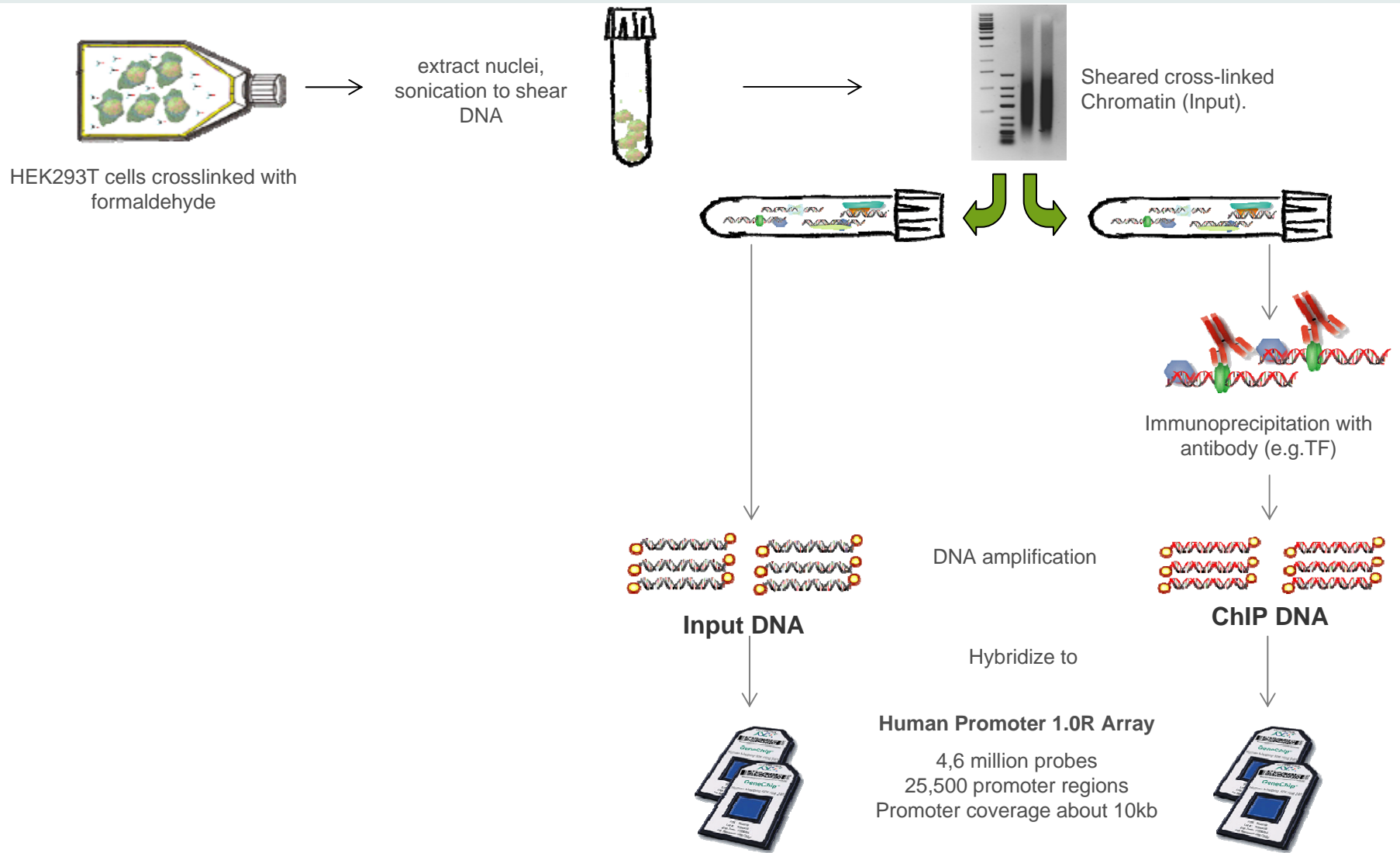
# Outline

- Selection of biological system
- Optimization of the assay
- Basic data-analysis
- Analysis of ChIP-chip data
- Interpretation of binding events





# Assay overview





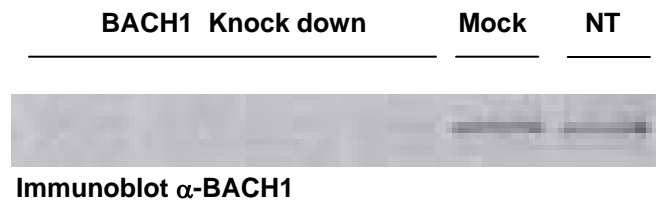
# Selection of biological system

- Substantial number of cells  $\sim 5 \cdot 10^7$ 
  - HEK293T17 cell line
- Transcription Factor of interest should be expressed and show nuclear localization
  - mRNA array data, Western blot, Immunofluorescence
- Need of high quality antibody for your TF
  - Previously used in ChIP or perform small scale ChIPs to test several antibodies
- Availability of other Data for your system and TF
  - Integration with other projects and public available Data

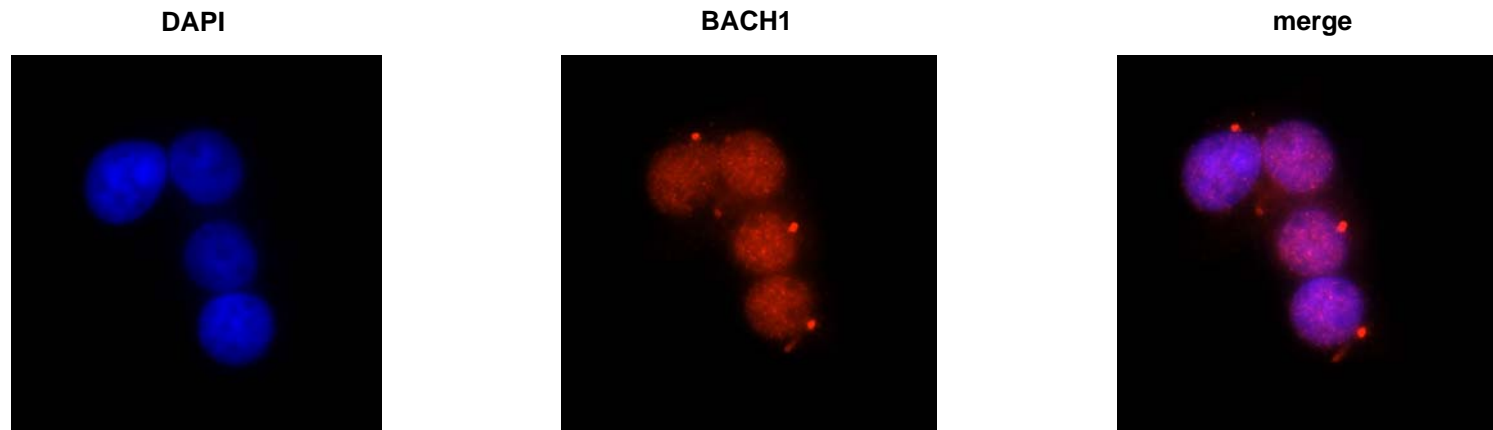




# Selection of biological system



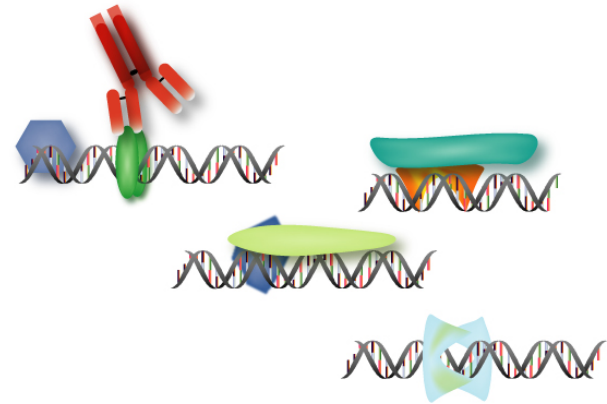
- Western blot
- Immunofluorescence





# Outline

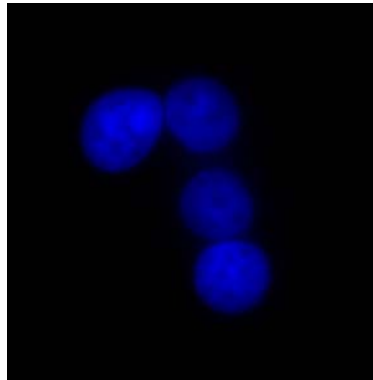
- Selection of biological system
- Optimization of the assay
- Basic data-analysis
- Analysis of ChIP-chip data
- Interpretation of binding events



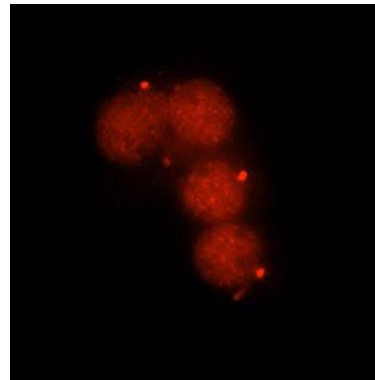


# Optimization

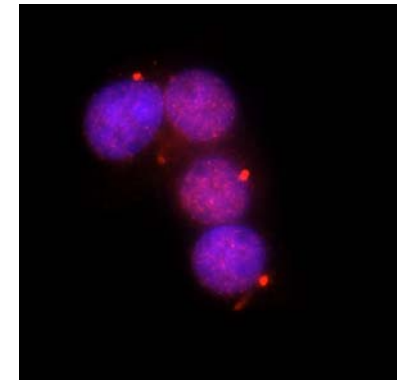
DAPI



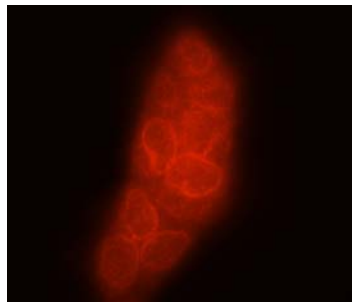
BACH1



merge



(BACH1)



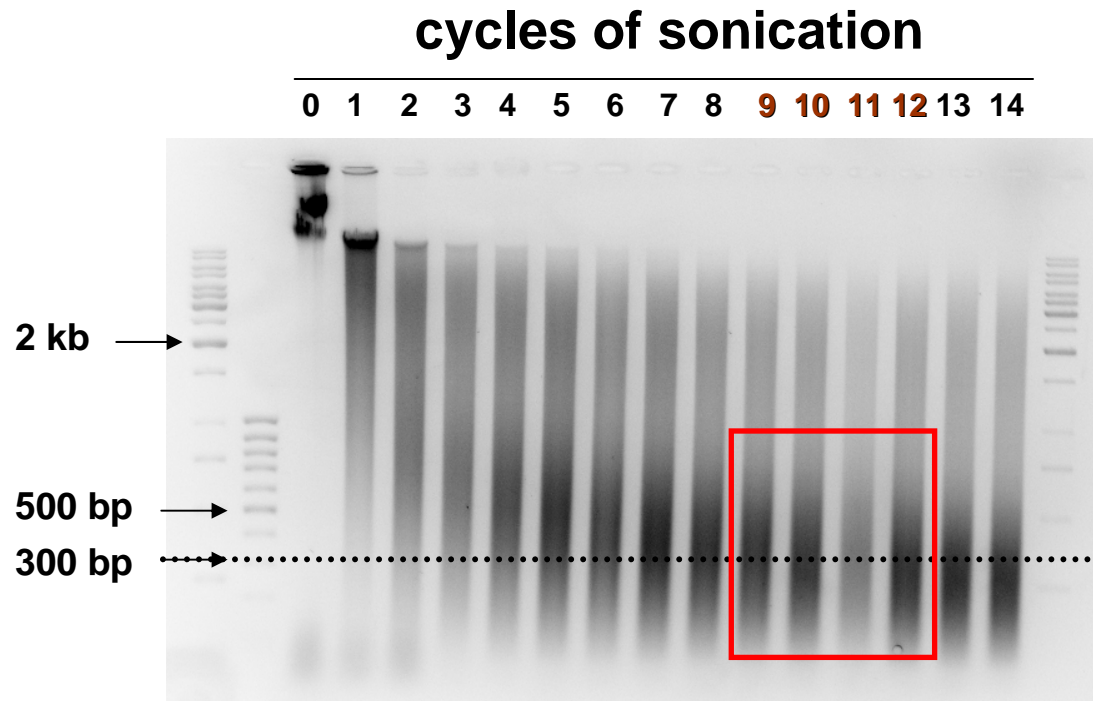
(BACH1)



- Antibody performance in IF is no guarantee, but may indicate ChIP performance



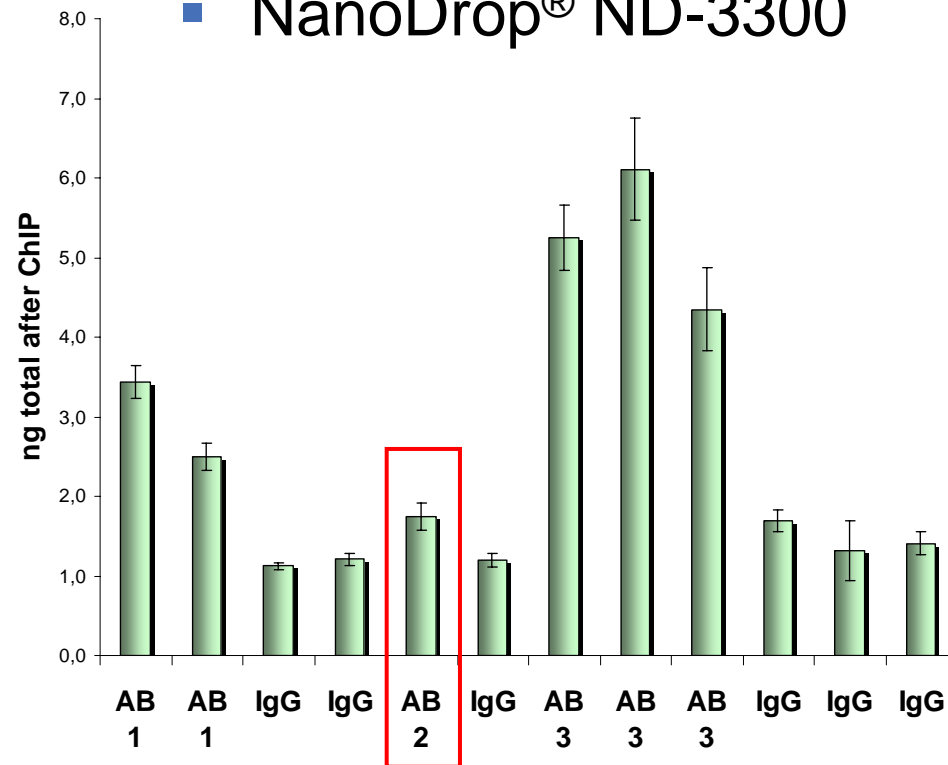
# Optimization of Sonication





# Amounts of Precipitated DNA in the ChIP

- PicoGreen<sup>®</sup>
- NanoDrop<sup>®</sup> ND-3300

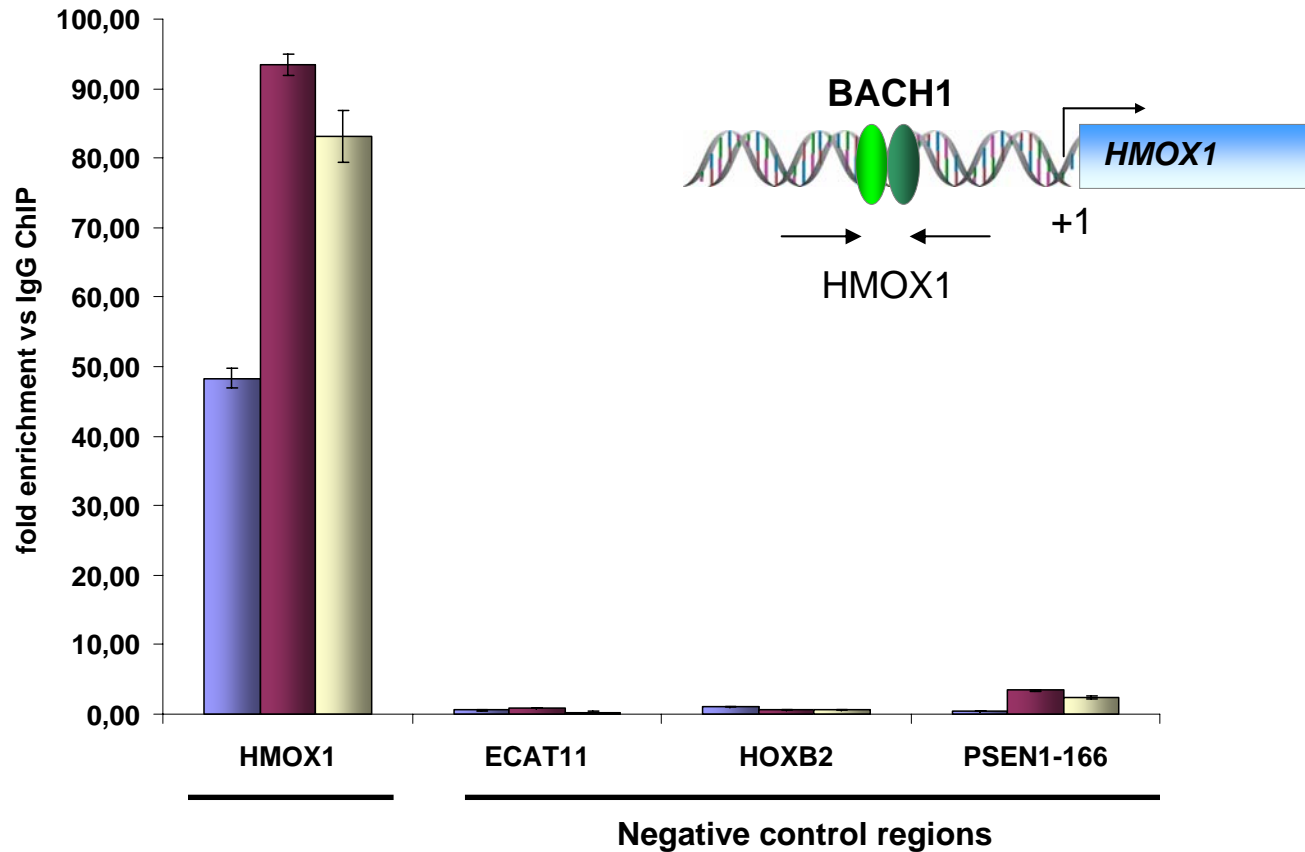


**Antibody (2) did not work! Confirmed with qRT-PCR.**



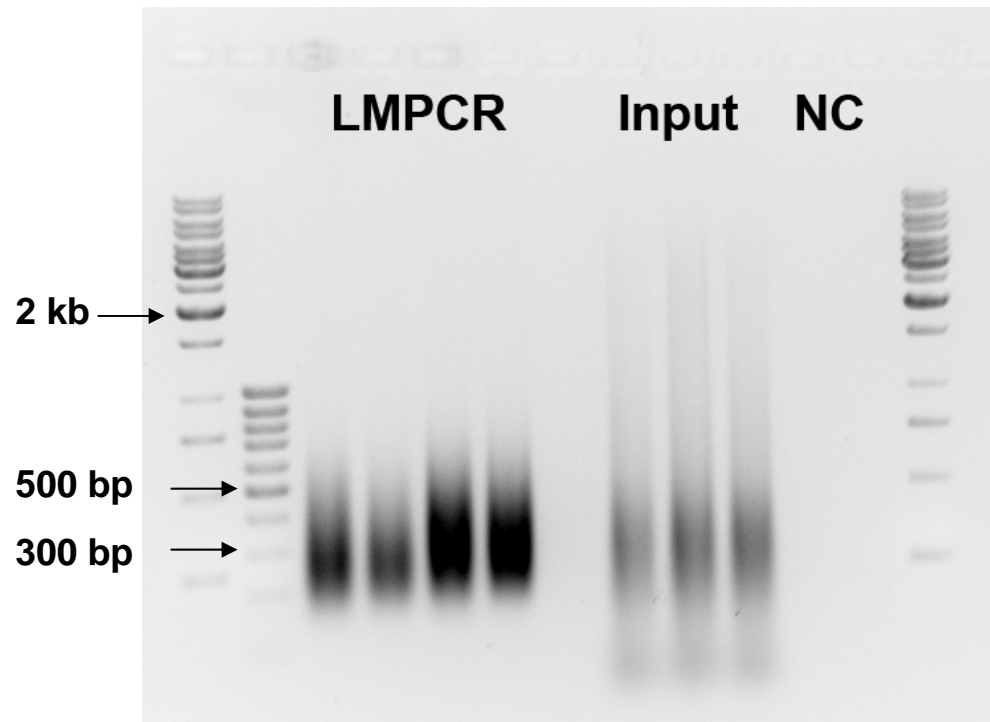
# Optimization of IP and Amplification

*HMOX-1 promoter*





# Optimization of IP and Amplification

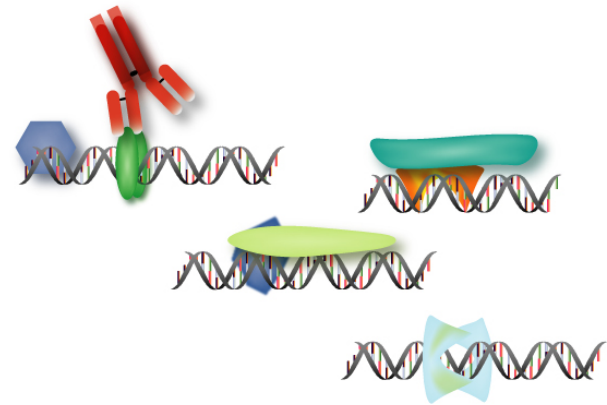


- Size range of Input and amplified samples



# Outline

- Selection of biological system
- Optimization of the assay
- **Basic data-analysis**
- Analysis of ChIP-chip data
- Interpretation of binding events





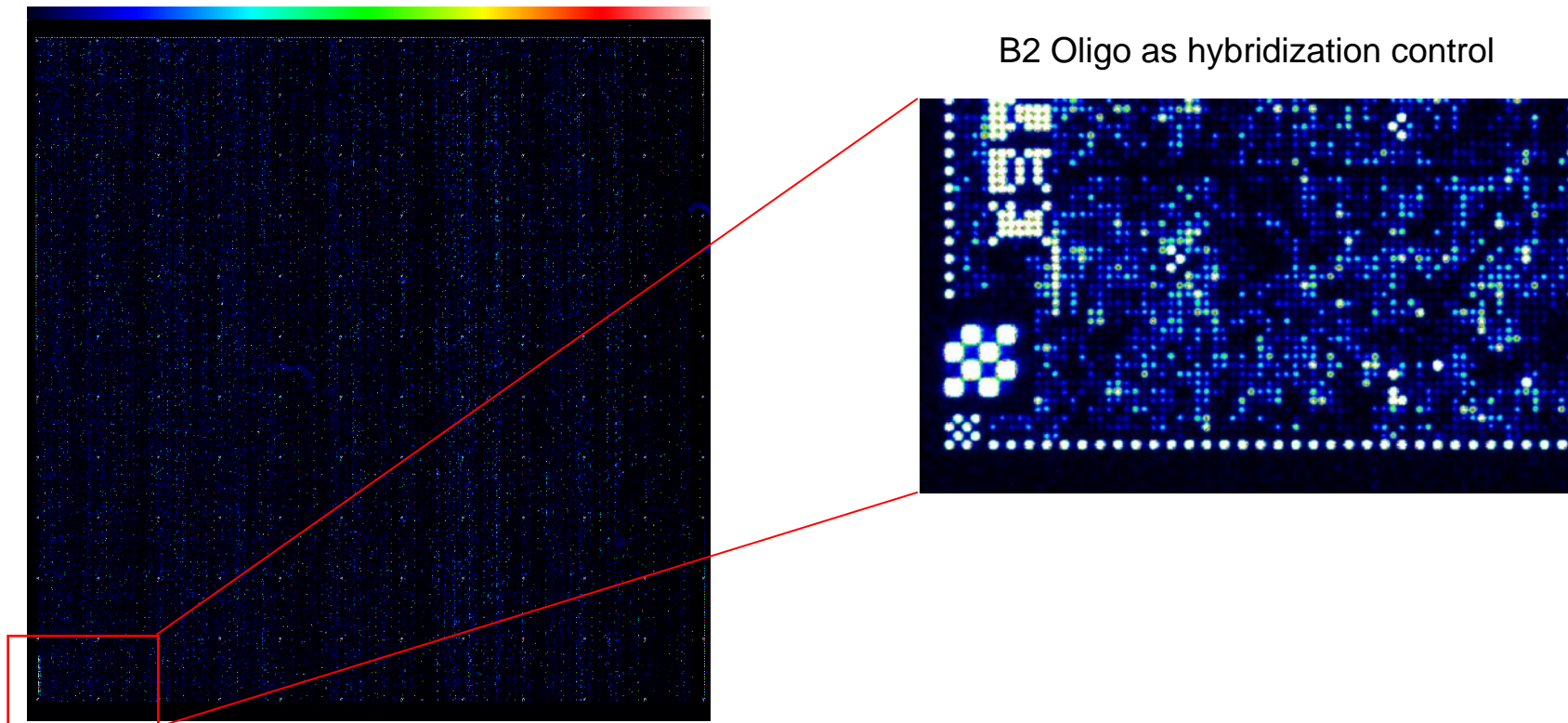
# Basic Data Analysis (QC)

- Array images
  - .CEL files or .DAT files
    - Affy DATViewer, Bioconductor, MBR
  
- MVA-Plots
  - Affy-Tiling Analysis Software (TAS)
  
- Positive and negative control regions
  - Positive: binding sites used for qRT-PCR
  - Negative: control regions used for qRT-PCR
  - If female genotype: Y-Chr as negative control



# Array images

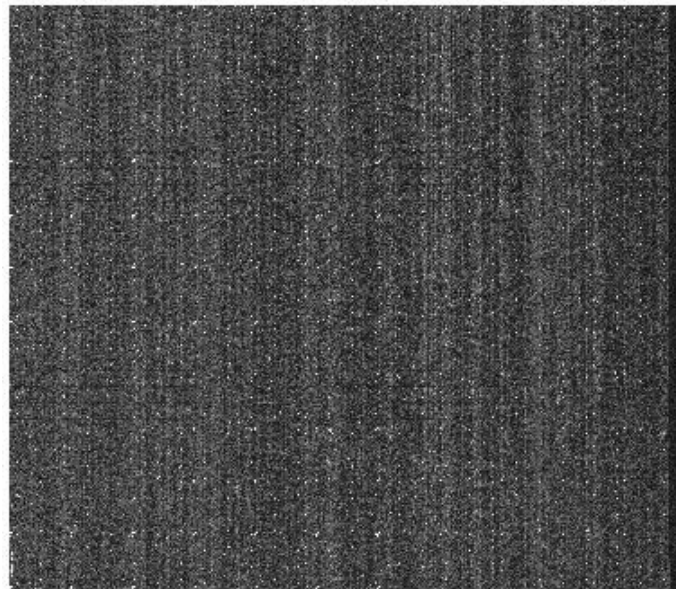
- Inspection of Array image with Affy DATViewer





# Array images

- Rebuilding of image from .CEL file with Bioconductor
  - [www.bioconductor.org/](http://www.bioconductor.org/)





# Array images

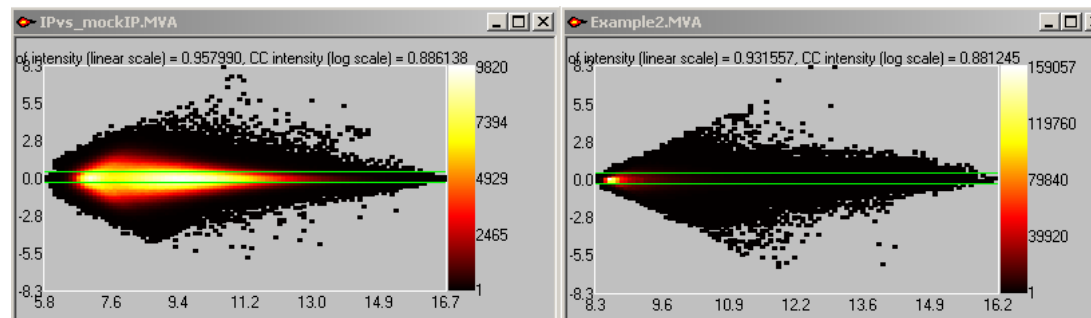
- Microarray Blob Remover (MBR)  
**A microarray JAVA tool which allows rapid visualization, detection, and removal of blob-like defects as an initial quality control step.**
  - <http://liulab.dfci.harvard.edu/Software/MBR/MBR.htm>





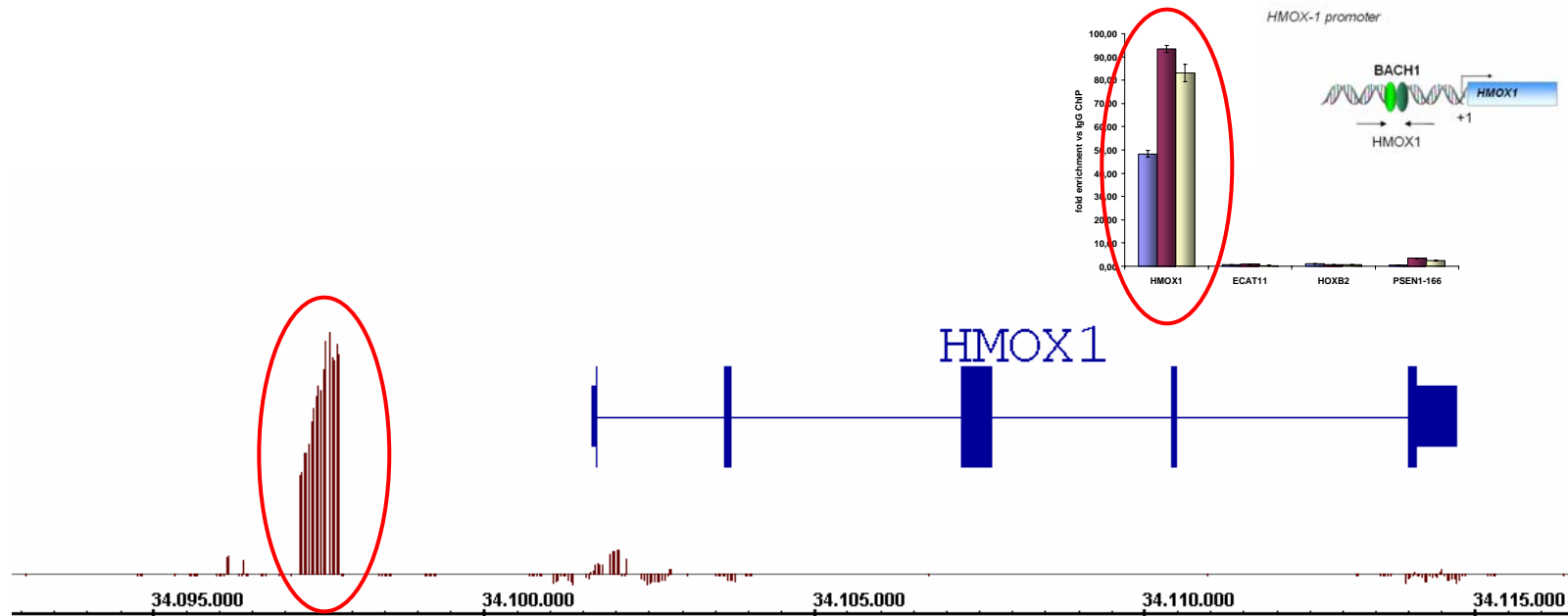
# MVA-Plots

- MVA-Plots
  - Typical correlations >90%



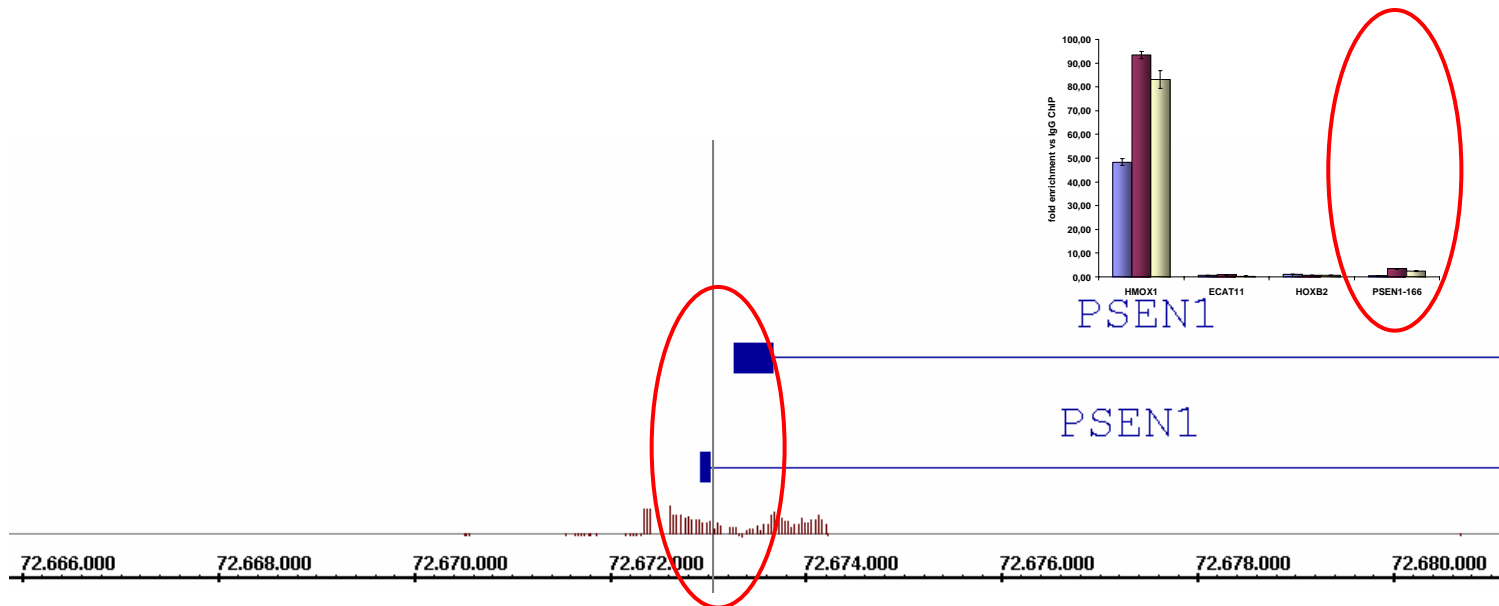


# Positive and negative control regions



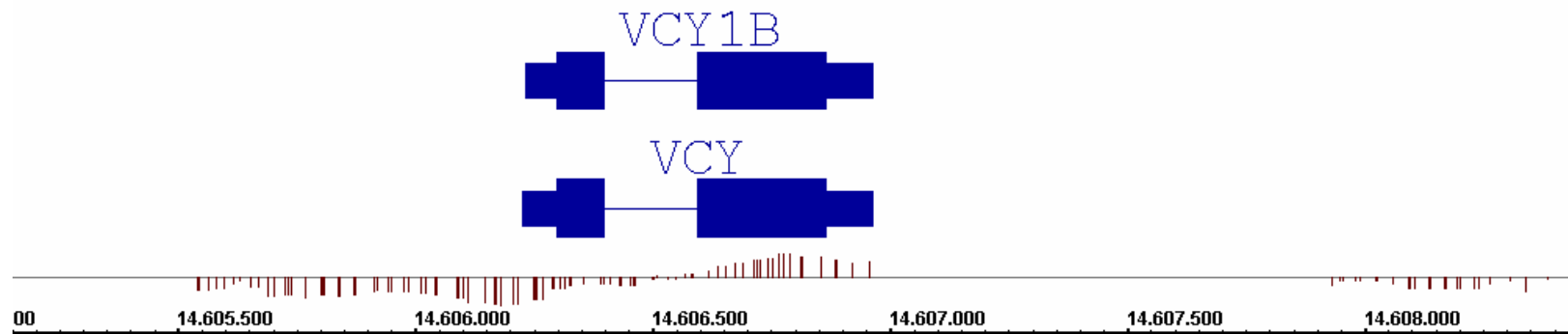


# Positive and negative control regions





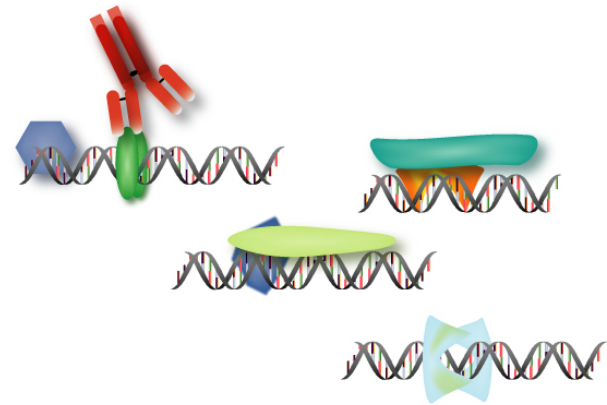
# Positive and negative control regions





# Outline

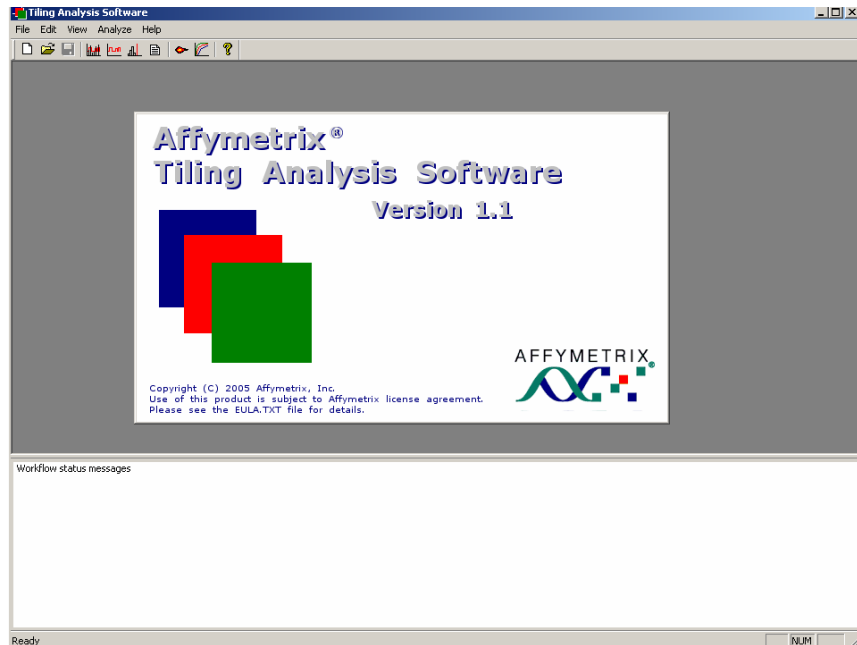
- Selection of biological system
- Optimization of the assay
- Basic data-analysis
- **Analysis of ChIP-chip data**
- Interpretation of binding events





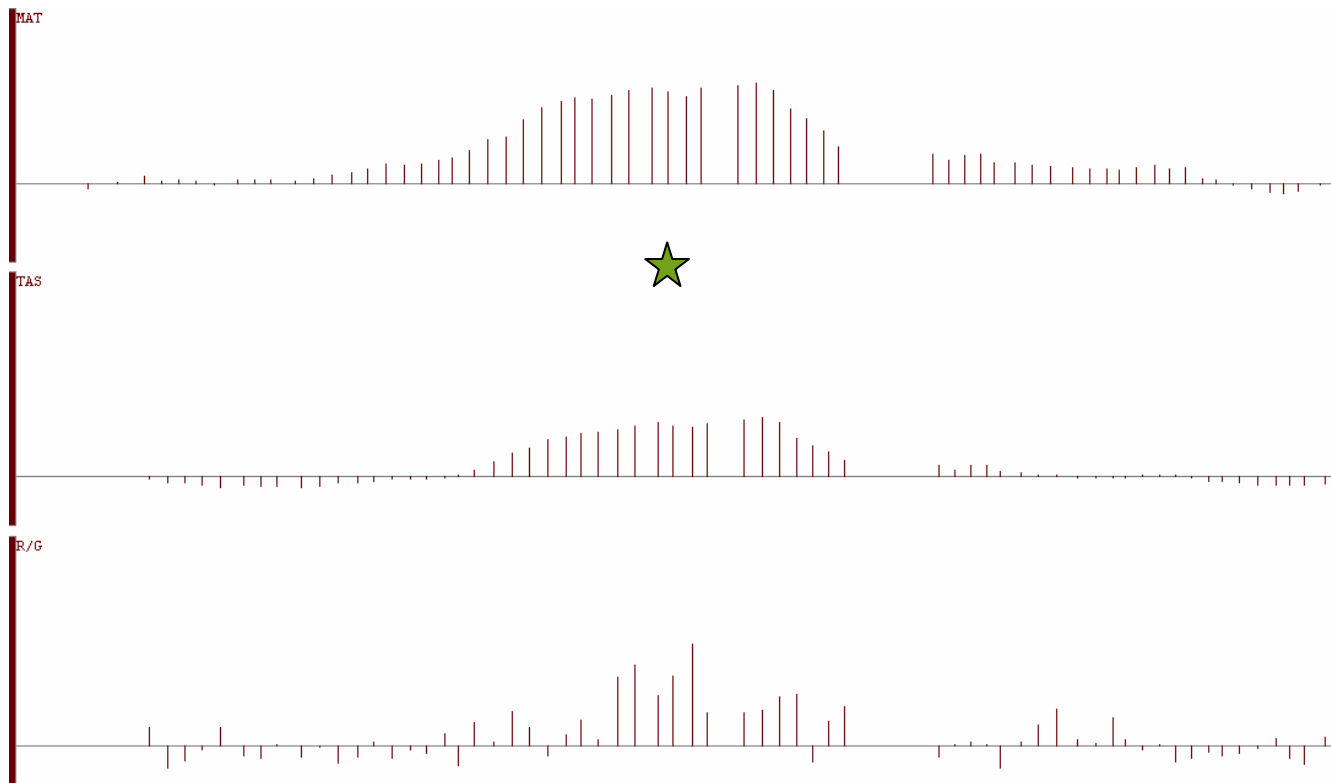
# Affymetrix Tiling Analysis Software (TAS)

- Quantile normalization
- Calculation of log Ratios
- $\log_2(\text{PMChIP})/\log_2(\text{PMControl})$
- Smear values along ~600-bp
  - rank-sum test
  - sliding window





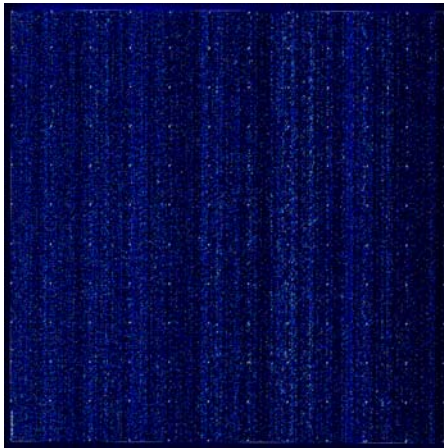
# Sliding window approach



# Model-based analysis of tiling-arrays (MAT)

- Instead of quantile normalization:
  - Probe Behavior Model Estimate and Probe Standardization:
    - <http://chip.dfci.harvard.edu/~wli/MAT/>

$$\log(PM_i) = \alpha n_{iT} + \sum_{j=1}^{25} \sum_{k \in \{A,C,G\}} \beta_{jk} I_{ijk} + \sum_{k \in \{A,C,G,T\}} \gamma_k n_{ik}^2 + \delta \log(c_i) + \varepsilon_i,$$



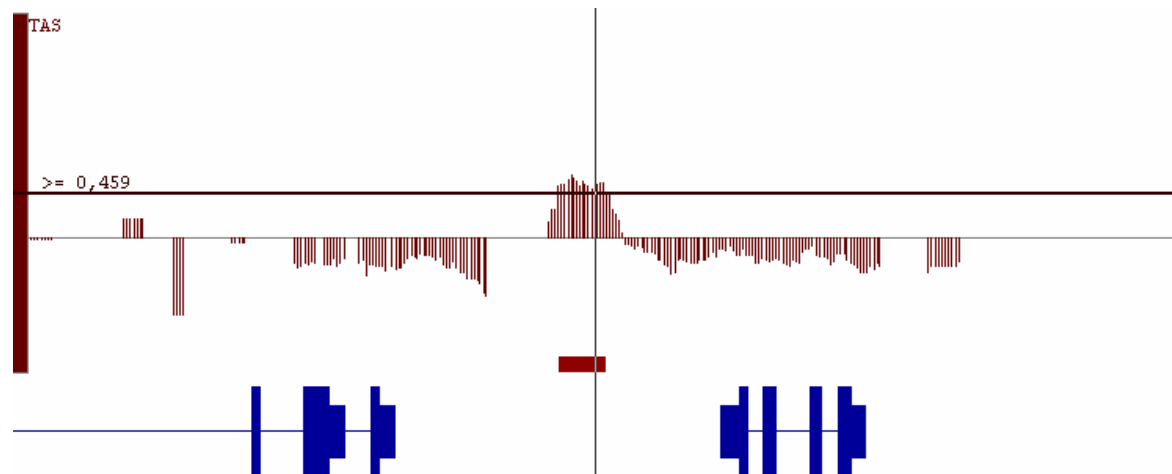
$$t_i = \frac{\log(PM_i) - \hat{m}_i}{S_i \text{ affinity bin}},$$

PNAS 2006;103;12457-12462; originally published online Aug 8, 2006;

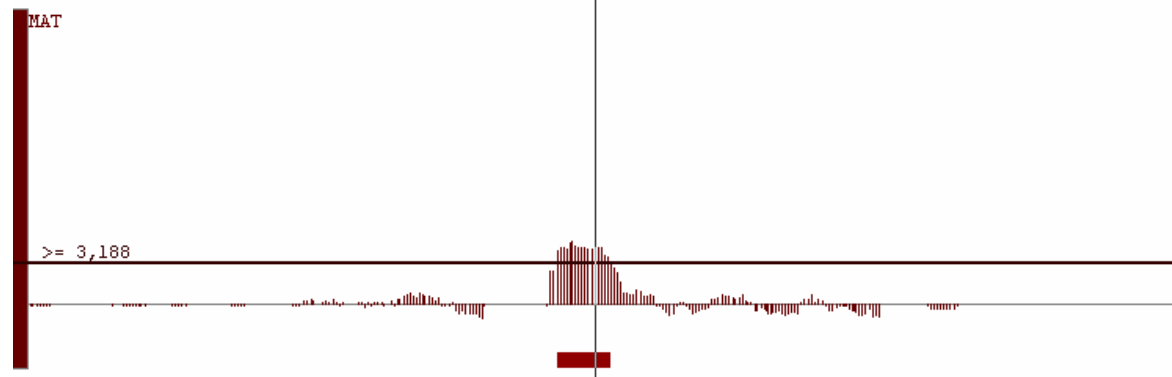


# Comparison TAS vs MAT

**TAS**

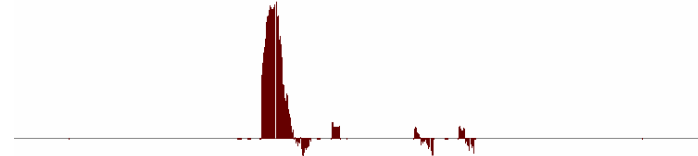
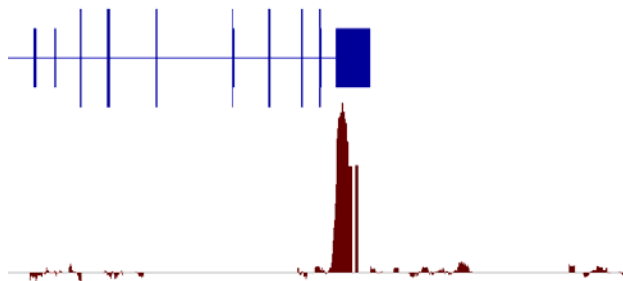
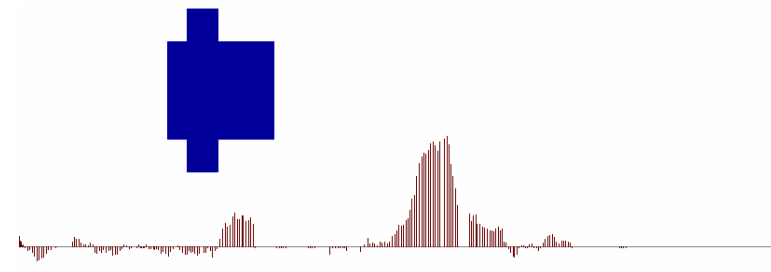
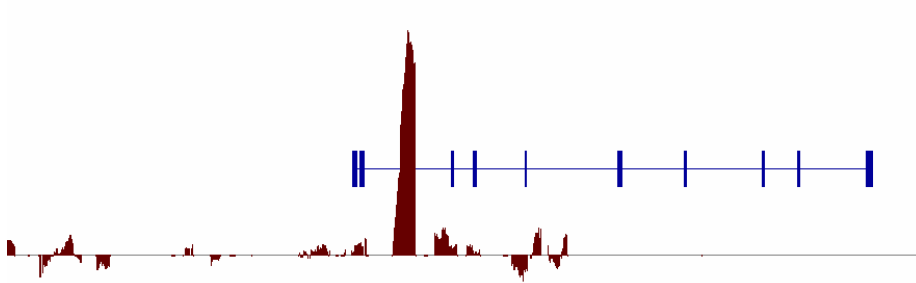


**MAT**





# MAT Peaks





# *De novo motif discovery*

- Meme
  - <sup>1</sup><http://meme.sdsc.edu/meme/intro.html>
  - Visualization with Weblogo
    - <sup>2</sup><http://weblogo.berkeley.edu/>

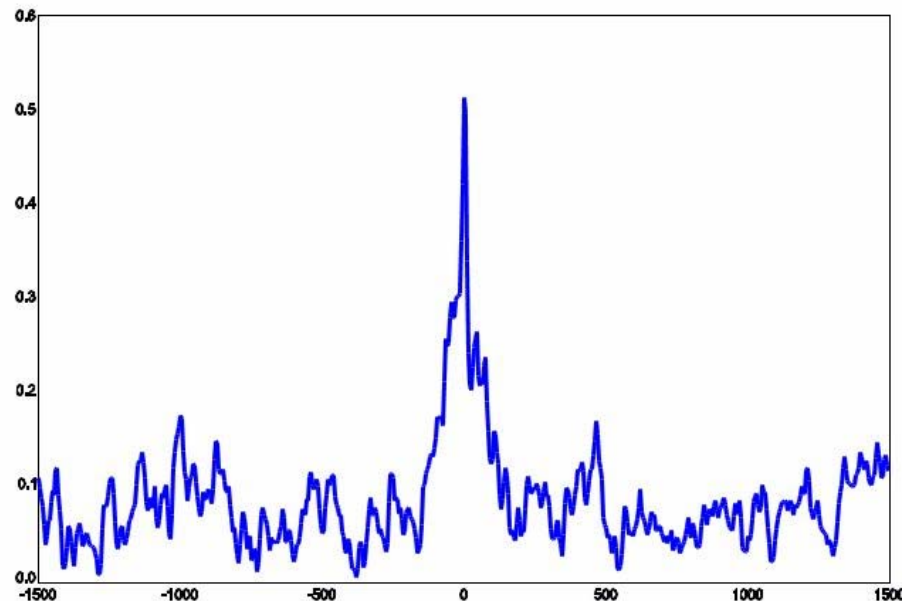
<sup>1</sup> Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

<sup>2</sup> Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, Genome Research, 14:1188-1190, (2004)



# Conservation of motif

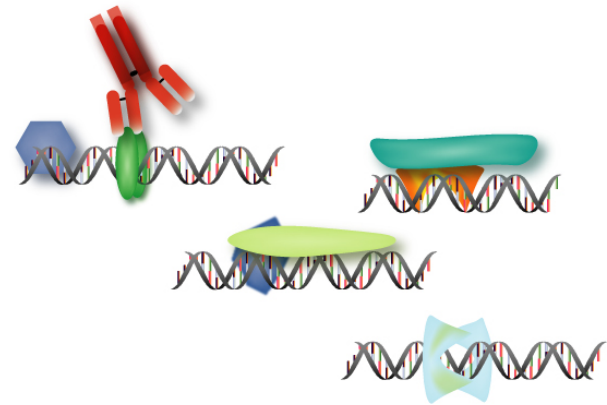
- High conservation score for the sites containing the motif
  - Cis-regulatory Element Annotation System (CEAS)
    - <http://ceas.cbi.pku.edu.cn/>





# Outline

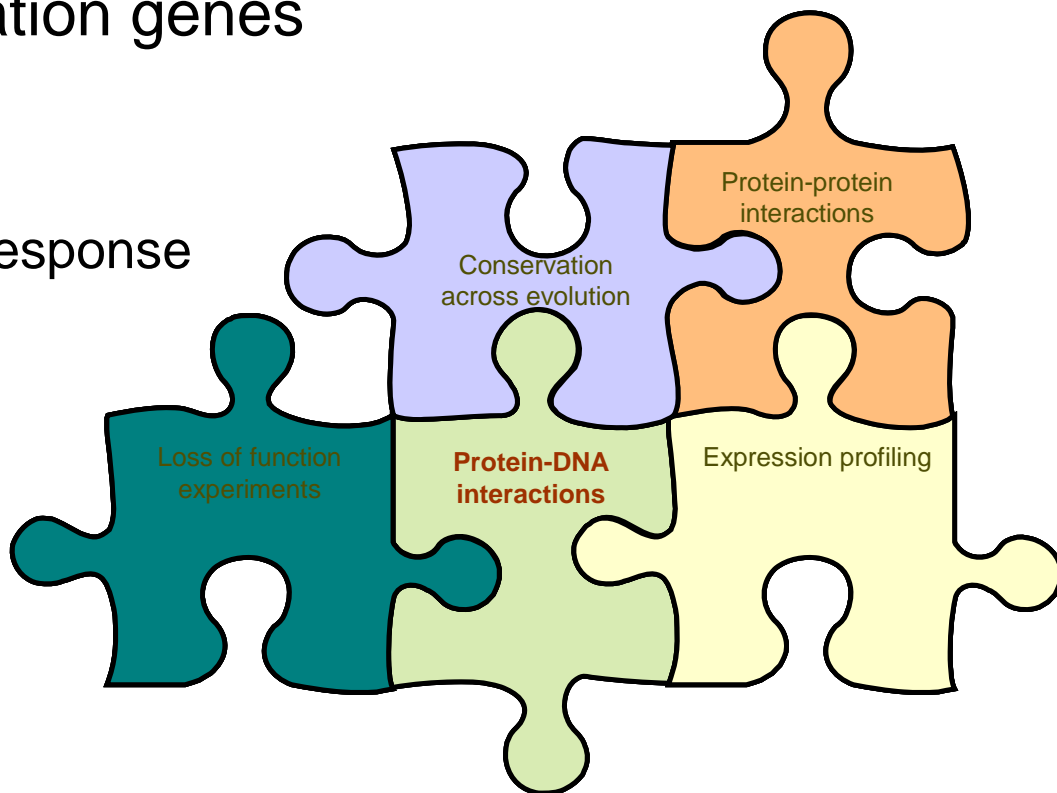
- Selection of biological system
- Optimization of the assay
- Basic data-analysis
- Analysis of ChIP-chip data
- Interpretation of binding events





# Biological interpretation of binding sites

- HMOX1 and  $\beta$ -Globin
- Phase II detoxification genes
  - Iron homeostasis
  - Cell cycle
  - Oxidative Stress response





Vertebrate Genomics dept

**Hans Lehrach**

Chromosome 21 group

Ilaria Piccini

Daniela Balzereit

Marc Sultan <http://chr21.molgen.mpg.de/>

Robert Querfurth

Alon Magen

**Marie-Laure Yaspo**

School of Computer Science- Tel Aviv University

Ron Shamir, Igor Ulitsky

