BIOINFORMATIK II
Übung 2

MEDIZINISCHE
UNIVERSITÄT
INNSBRUCK

Dietmar Rieder

http://icbi.at/bioinf2

Übung II

Einführung

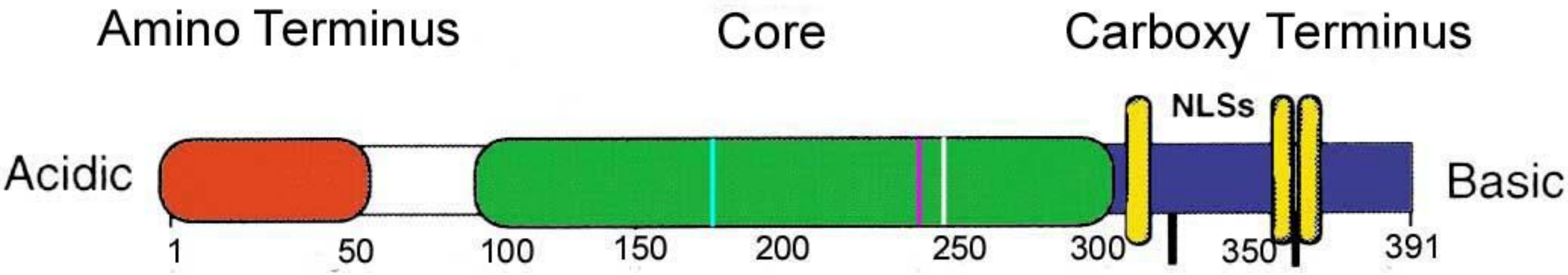
Transktiptionelle Regulation
TFBS

Transcription Factors

- These proteins promote transcription by
  1. Binding to DNA
  2. Activate Transcription

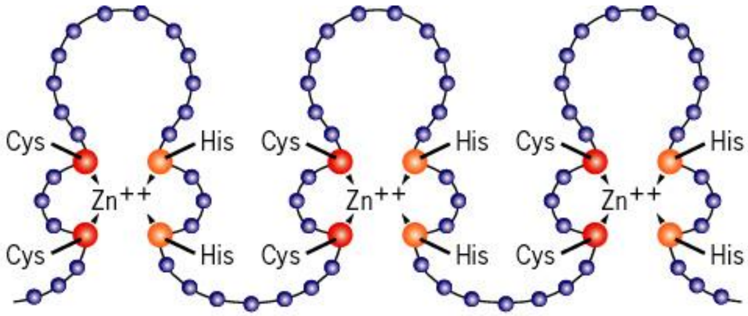
- These two functions usually reside on separate structural domains of the Transcription Factor protein.
- Activation occurs via interactions with other transcription factors and /or RNA polymerase

TF domains



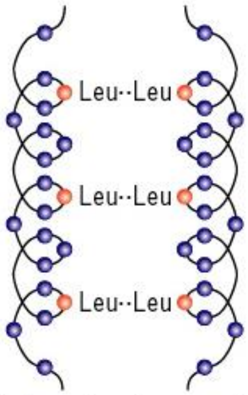
Activation domain **DNA binding domain** **tetramerizazion domain**

DNA Binding Domains – Conserved Motifs



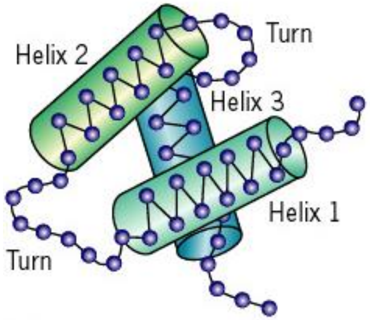
(*a*) Zinc finger motif.

Zinc Finger Motif



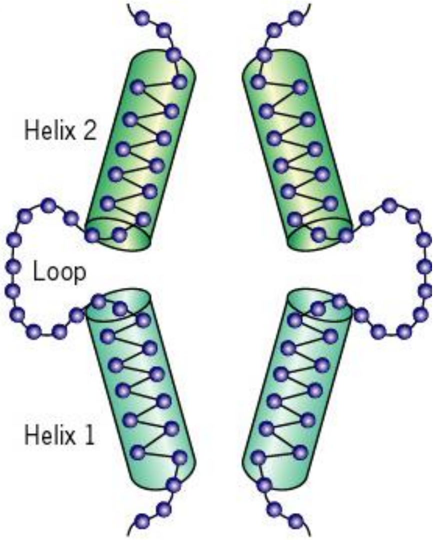
(*c*) Leucine zipper motif.

Leucine Zipper Motif



(*b*) Helix-turn-helix motif.

Helix-turn-Helix Motif

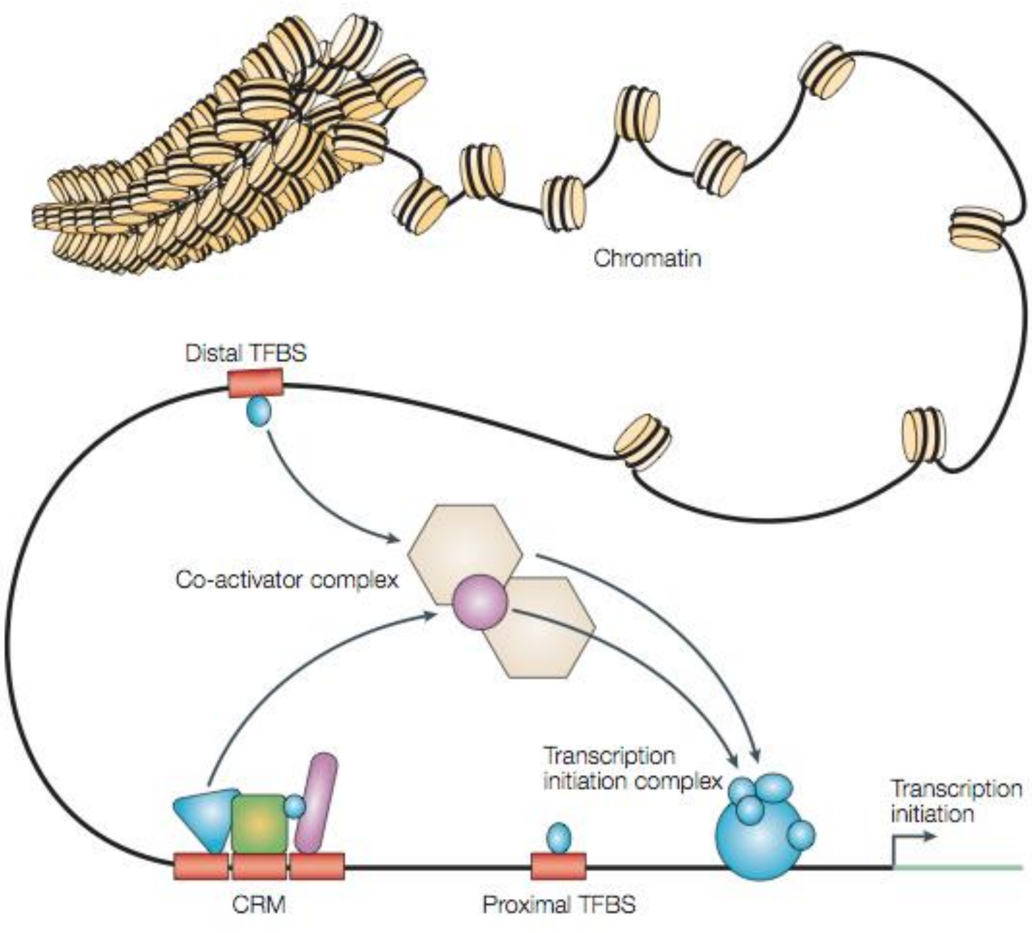


Helix-loop-Helix Motif

(*d*) Helix-loop-helix motif.

Structural motifs within different types of transcription factors.

Promoters

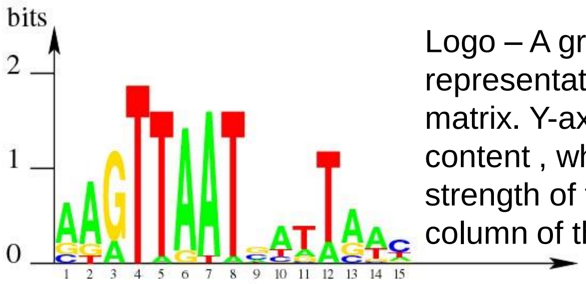


- Promoters are DNA segments upstream of genes where transcription factors can bind and initiate transcription
- Promoter *attracts* RNA Polymerase to the transcription start site
- Regulatory Regions Proximal/Distal to TSR
  - May be positive (enhancing) or negative (repressing)
  - Modules – Sets of TFBS within a region that function together

Representing Binding Sites for a TF

- A single site
  - `AAGTTAATGA`
- A set of sites represented as a consensus
  - `VDRTWRWWSHD (IUPAC degenerate DNA)`
- A matrix describing a set of sites:

A	14	16	4	0	1	19	20	1	4	13	4	4	13	12	3
C	3	0	0	0	0	0	0	0	7	3	1	0	3	1	12
G	4	3	17	0	0	2	0	0	9	1	3	0	5	2	2
T	0	2	0	21	20	0	1	20	1	4	13	17	0	6	4

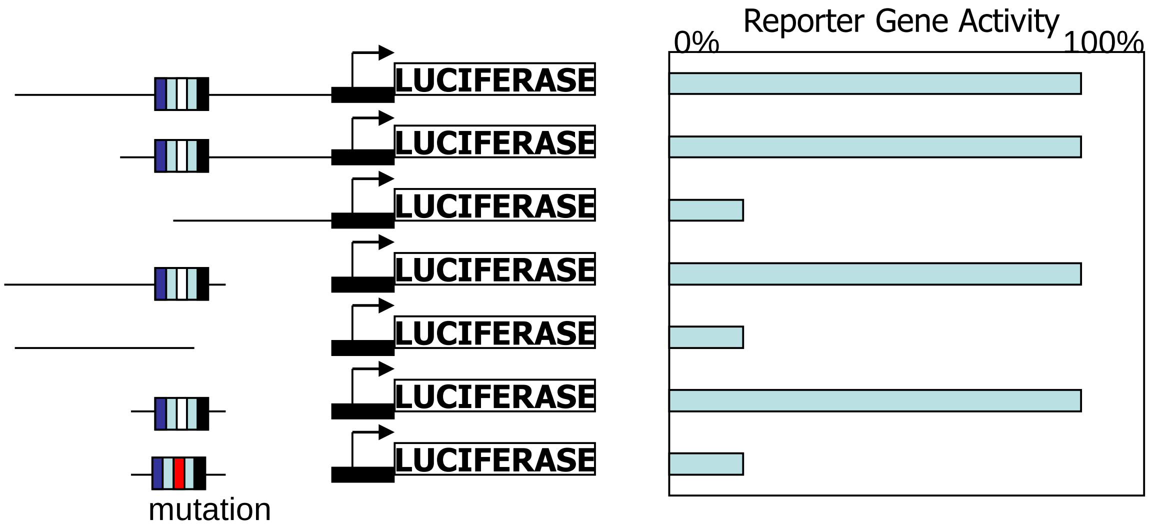


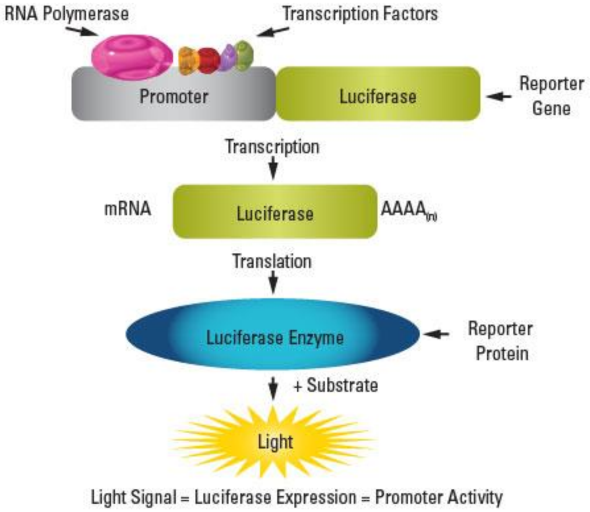
Logo – A graphical representation of frequency matrix. Y-axis is information content , which reflects the strength of the pattern in each column of the matrix

Set of binding sites

AAGTTAATGA
CAGTTAATAA
GAGTTAAACA
CAGTTAATTA
GAGTTAATAA
CAGTTATTCA
GAGTTAATAA
CAGTTAATCA
AGATTAAAGA
AAGTTAACGA
AGGTTAACGA
ATGTTGATGA
AAGTTAATGA
AAGTTAACGA
AAATTAATGA
GAGTTAATGA
AAGTTAATCA
AAGTTGATGA
AAATTAATGA
ATGTTAATGA
AAGTAAATGA
AAGTTAATGA
AAGTTAATGA
AAATTAATGA
AAGTTAATGA
AAGTTAATGA
AAGTTAATGA
AAGTTAATGA

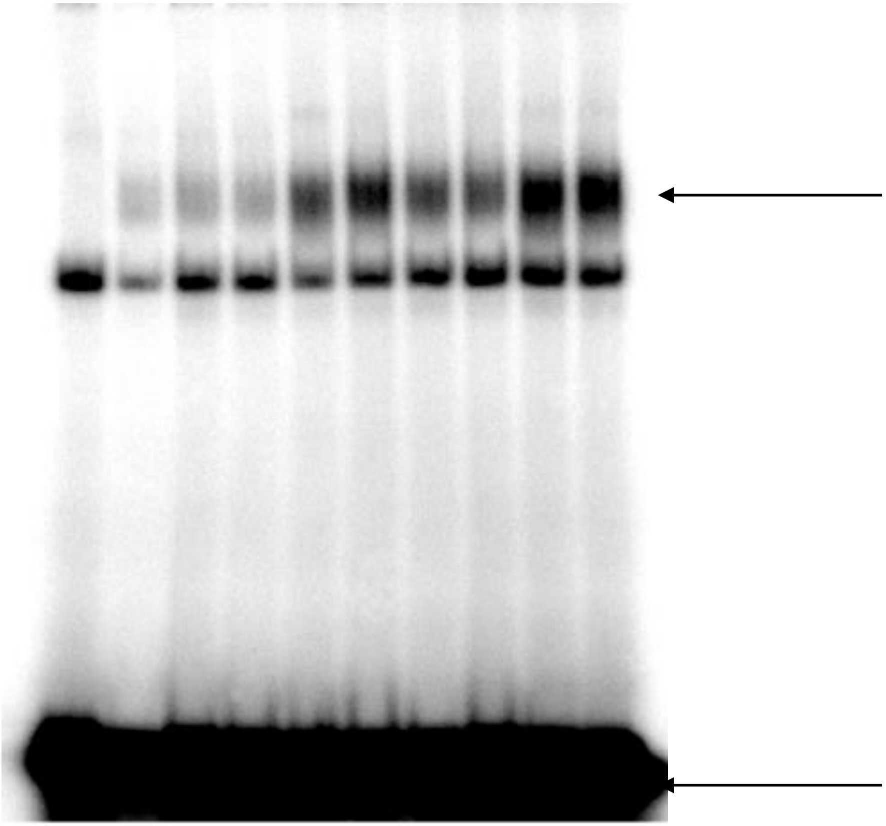
Lab Discovery of TF Binding Sites





Identify functional regulatory region within a sequence and delineate specific TFBS through mutagenesis (and in vitro binding studies)

EMSA/Gel Shift Assays to Identify Binding Proteins



TF + DNA

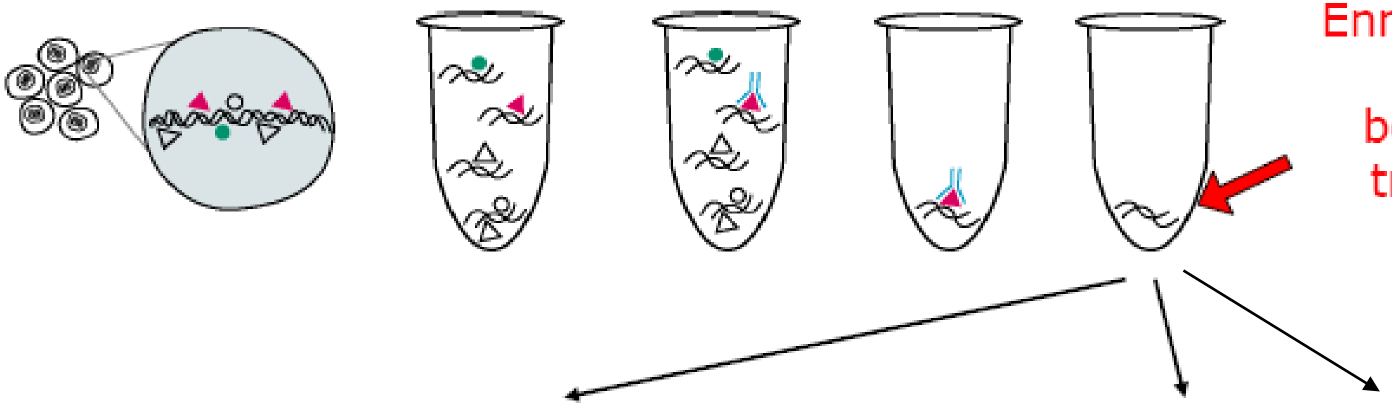
DNA

http://www.biomedcentral.com/content/figures/1741-7015-4-28-8.jpg

High-throughput Methods

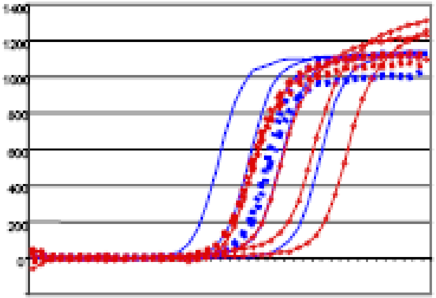
- <u>SELEX</u>
  - mix random ds DNA oligonucleotides with TF protein, recover TF-DNA complexes and sequence DNA
- <u>Protein Binding Arrays</u>
  - prepare arrays with ds DNA attached, label protein with a fluorescent mark and observe DNA bound by protein
- <u>ChIP</u>
  - covalently link proteins to DNA in cell, shear DNA, recover protein-DNA complexes and identify DNA (PCR, array or sequencing)

ChIP

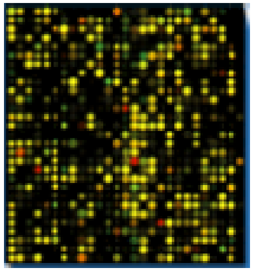


Enriches for DNA fragments bound by the transcription factor

ChIP-QPCR



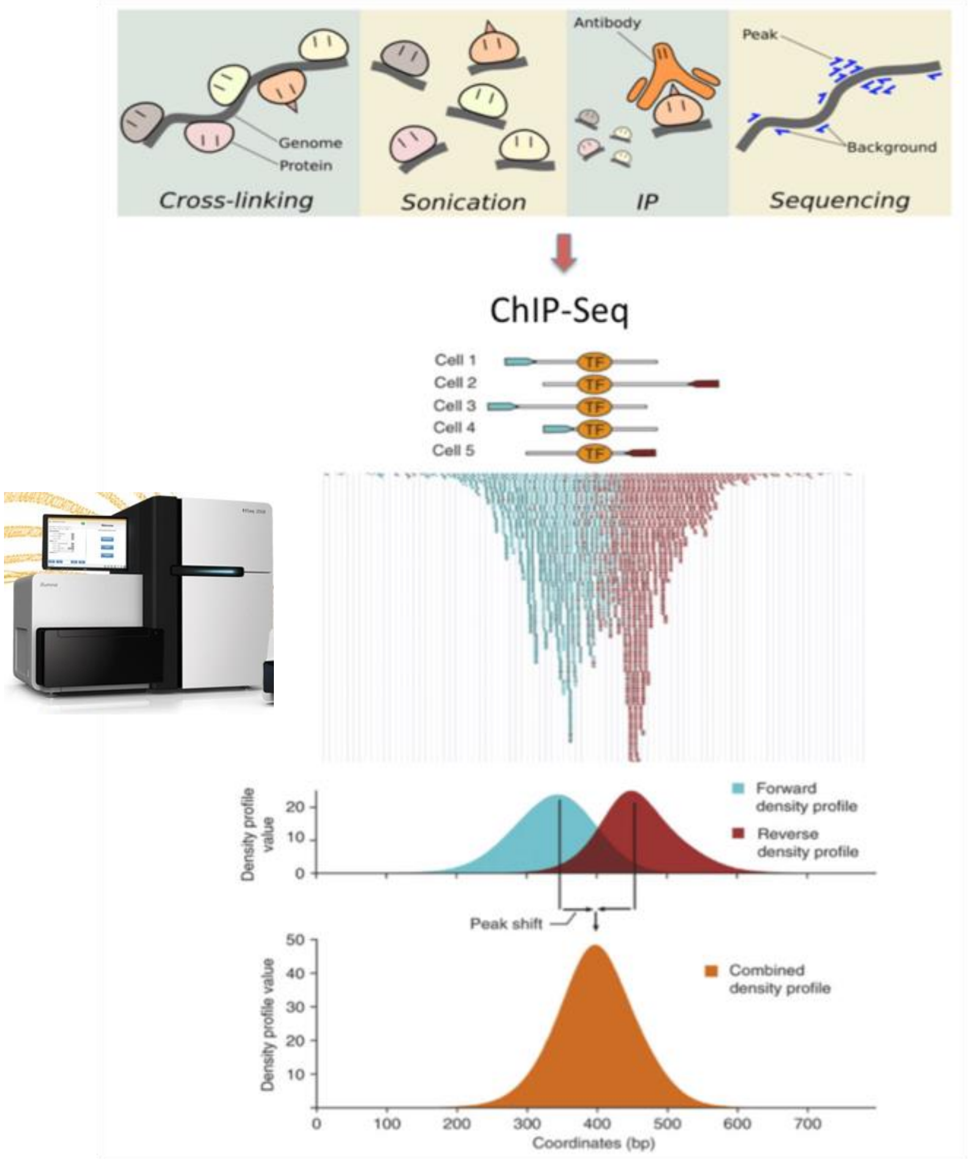
ChIP-chip



ChIP-seq

Resolution is about ~500bp

ChIP-seq


Antibody
Peak
Genome
Protein
Background
Cross-linking
Sonication
IP
Sequencing
ChIP-Seq
Cell 1
Cell 2
Cell 3
Cell 4
Cell 5
TF
Density profile value
20
10
0
Forward density profile
Reverse density profile
Peak shift
50
40
30
20
10
Combined density profile
0
100
200
300
400
500
600
700
Coordinates (bp)

Databases and Servers

- Sequence logo Servers
  - http://icbi.at/logo/
  - http://weblogo.berkeley.edu/logo.cgi
- TF databases
  - TRANSFAC: http://www.gene-regulation.com
  - JASPAR: http://jaspar.genereg.net/ (free)
  - Genomatix: http://www.genomatix.de/ (free trial)
- Servers for analyses
  - MEME: http://meme-suite.org/ (mirror: http://meme.i-med.ac.at)
  - RSAT: http://www.rsat.eu/

Exercise II

1. Generate a PFM (position frequency matrix) and the corresponding sequence logo from a set of experimentally verified TFBS

2. Retrieve promoters (-700 … +300) for the genes in geneListMatrixTest.txt and use the PFM from the previous task to search for potential TFBS (transcription factor binding sites). Compare with PSM/PFMs from Jaspar and Transfac.

3. Map the p53 peak regions from a ChIPseq result file to the human genome. And display the region around some target genes. Extract the sequence and search the exact TFBS with the p53 PFM. Map them back to the genome browser.

4. Extract the peak sequences from the top 300 ChIPseq peaks (peak file from task 3). Discover a TFBS motif with MEME-ChIP and/or RSAT using the sequences from the top 300 peaks. Do a motif comparison.