-Alternative Splicing- Discovery of Splicing Regulatory Elements

Stephanie Keller

Prof. Mihaela Zavolan
Biozentrum, University of Basel & Swiss Institute of Bioinformatics

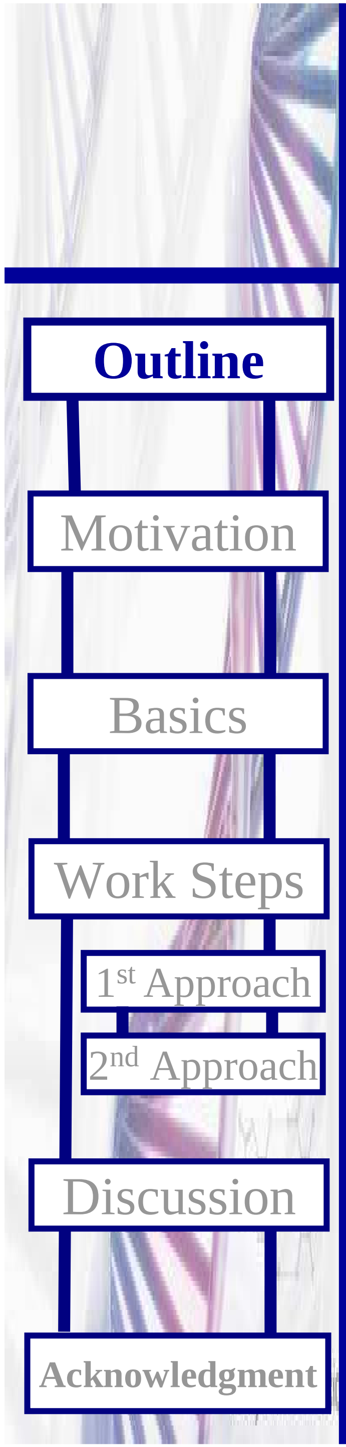
Bled, February 2007

Outline

- Motivation
- Basics
- Work Steps
  - 1^{st} approach

    Inference of binding specificity of splicing regulatory factors from known binding sites
  - 2^{nd} approach

    Inference of regulatory elements using phylogeny-sensitive methods and comparative genomic data
- Discussion
- Acknowledgment



Outline
Motivation
Basics
Work Steps
1^{st} Approach
2^{nd} Approach
Discussion
Acknowledgment

Motivation

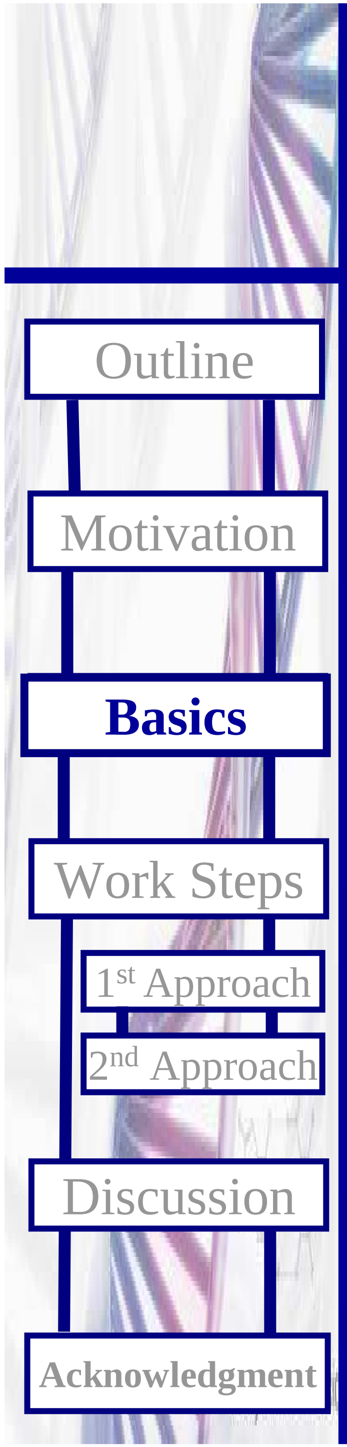
Outline
Motivation
Basics
Work Steps
1st Approach
2nd Approach
Discussion
Acknowledgment

- What?
  - Molecular recognition is necessary for the regulation of biological processes

    e.g. specific recognition of sites for replication, initiation, termination of transcription by proteins

  →Find such regulating sequences to which proteins can bind

- Why?
  - Motifs correspond to changes in development or environment
  - Understand how complex processes are regulated in specific cellular context
  - Indicate relationships and ancestry between different species
  - Treatment of ailments (i.e. research in gene therapy)



Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Basics

- Alternative Splicing
- Splice-regulating Sequences and corresponding RNA-binding Proteins
  - ESE (SRp)
  - ESS (hnRNP)

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

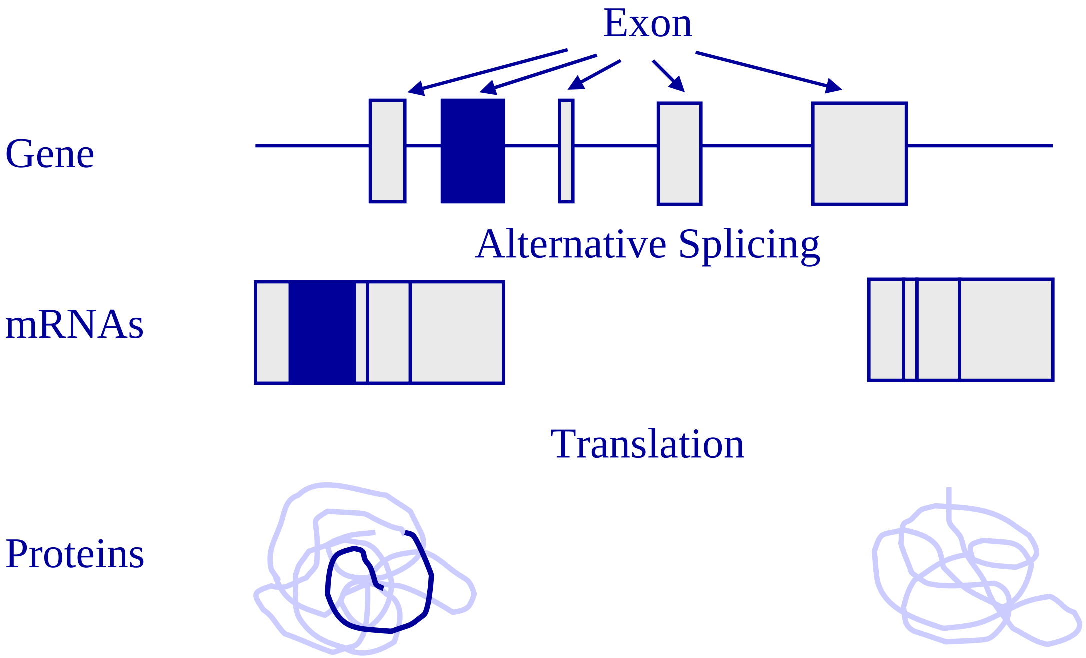
Discussion

Acknowledgment

Basics

» Alternative Splicing

- Individual genes produce multiple protein isoforms
- Alternative use of exons or exon parts within pre-mRNA transcript
- Can be specific to a tissue, developmental stage or a condition
- ~40-60% of human genes are alternatively spliced



Outline
Motivation
Basics
Work Steps
1st Approach
2nd Approach
Discussion
Acknowledgment

Basics

» Splice-regulating sequences & RNA-binding Proteins

- Discrete and highly variable sequences within exons
- Important in defining constitutive and alternative exons
- Control splice site choice

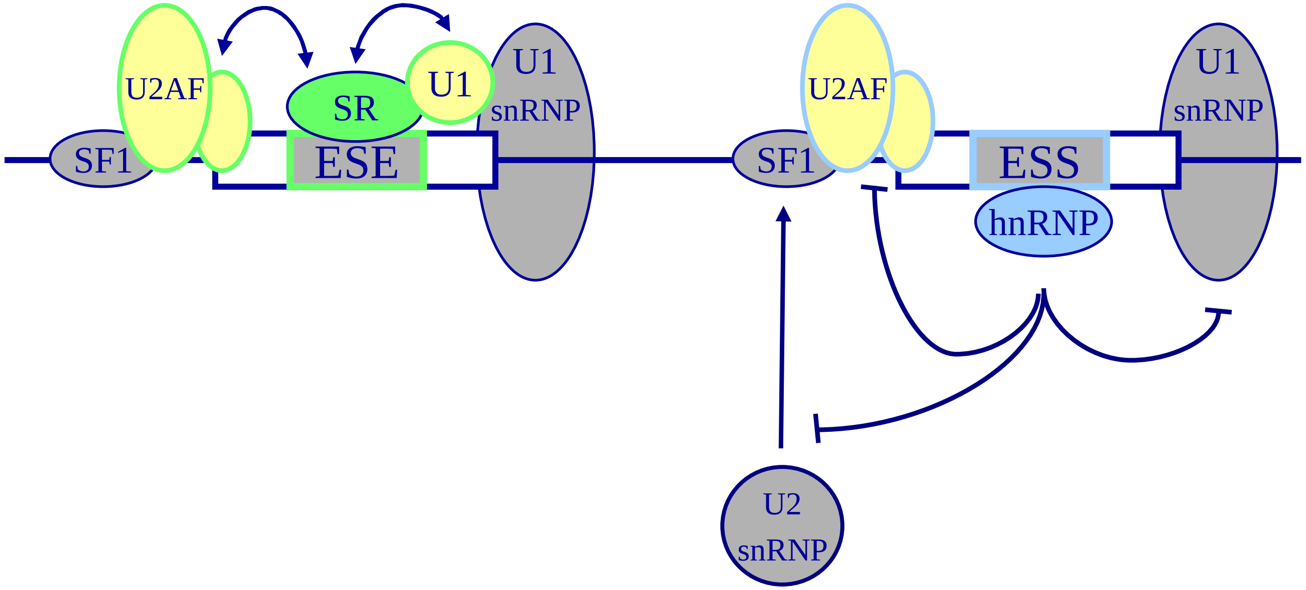
ESE (Exonic Splicing Enhancer)

- Activity involves their binding by members of a family of splicing regulators (often SRp - serine-arginine-rich proteins)
- Promote use of weak or regulated splice sites

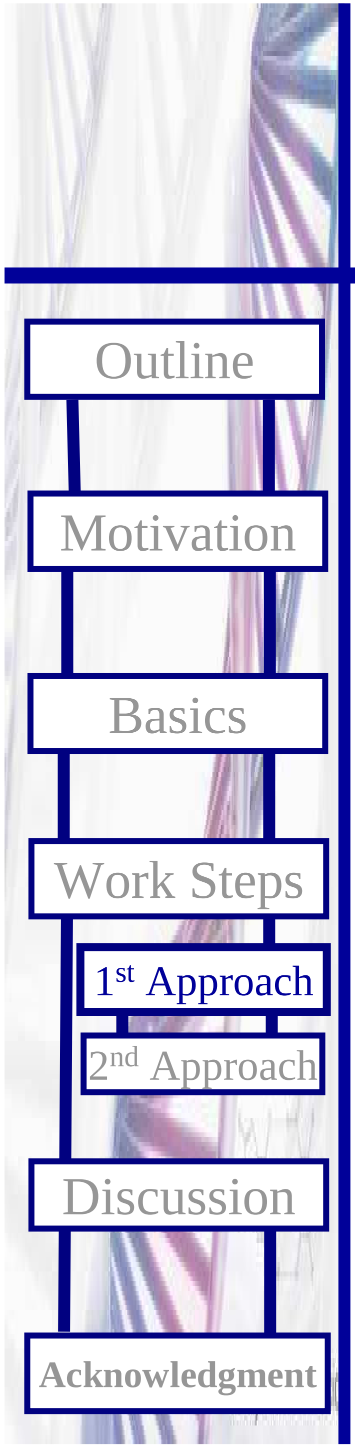
ESS (Exonic Splicing Silencer)

- Build complex with splice regulatory factors (often hnRNP - heterogeneous nuclear ribonucleoproteins)
- Repress use of splice sites

Basics



SRps and hnRNPs regulate the recognition of splice sites and the definition of intron and exon sequences



Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Work Steps

► *First Approach*

Using known binding sites of splicing regulatory factors

- Second Approach

Work Steps

» First Approach (I)

- Inference of binding specificity of splicing regulatory factors from known binding sites (SELEX data)
- Extract sequences from [Singh & Valcárcel, 2005]

→ Preferred binding sites from splice regulatory factors

hnRNP	SRp
A1	9G8
C1/C2	ASF/SF2
E1/E2	SC35
H/H‘/F	SRp30c
I	SRp40c
K	SRp55
M	Tra2β
SXL	

Work Steps

» First Approach (II)

Outline
Motivation
Basics
Work Steps
1st Approach
2nd Approach
Discussion
Acknowledgment

- Cluster sequences of each protein with PROCSE

→ Weight matrices (WMs) of possible motifs with a length 6 up to 10 nt

- Extract most representative WMs

- Create profile with PROFILER
  - Background model using hg18
  - Random sequence (hg18, chr18, position 748411 to 763327)

  to calculate z-score along the sequence

- Run separately for WMs of hnRNP and SRp
  (window length of 3 at each site, 100 nt enhancer length)

Work Steps

» First Approach (III)

Outline

Motivation

Basics

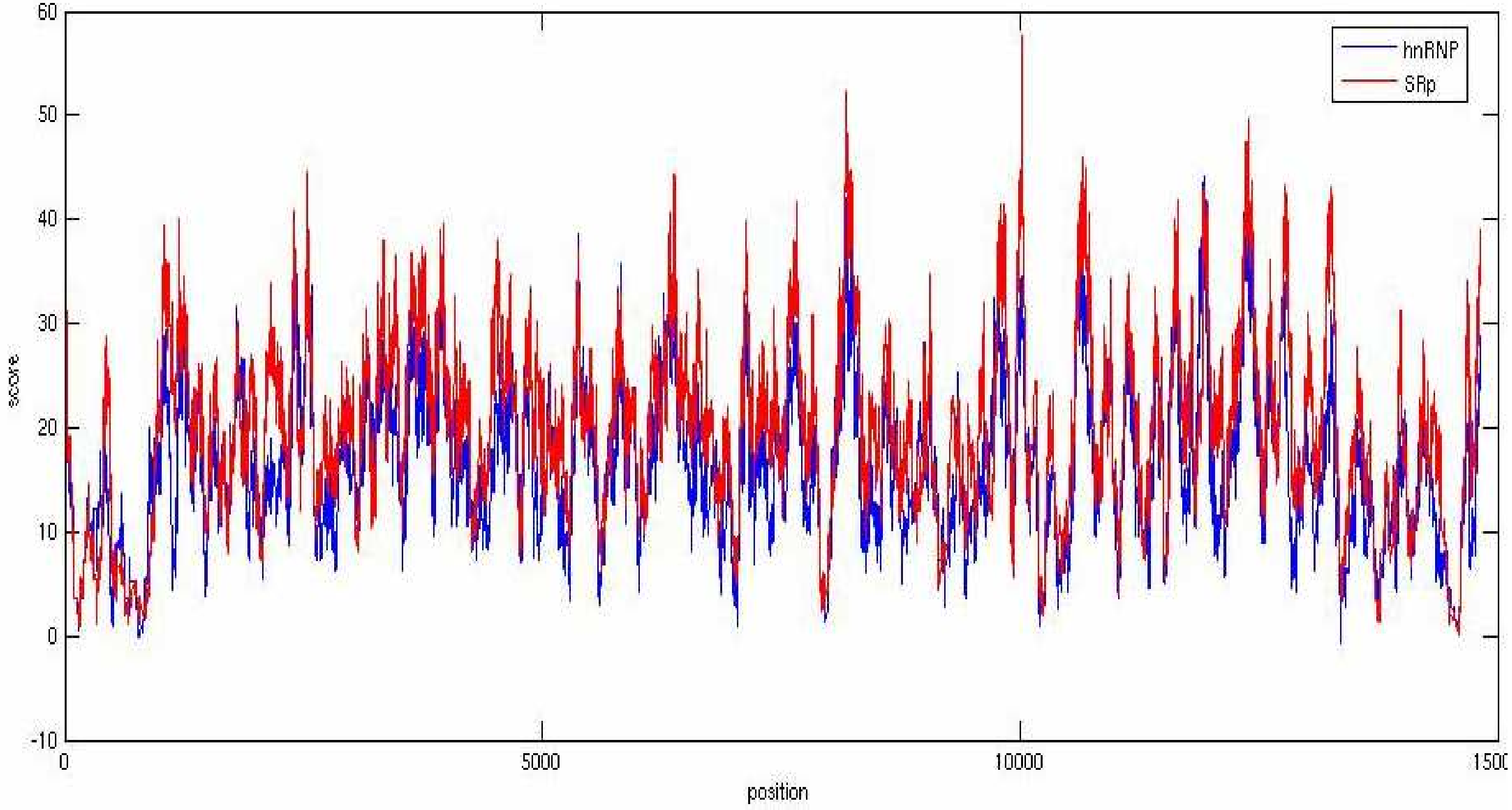
Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment



Work Steps

» First Approach (IV)

Outline

Motivation

Basics

Work Steps

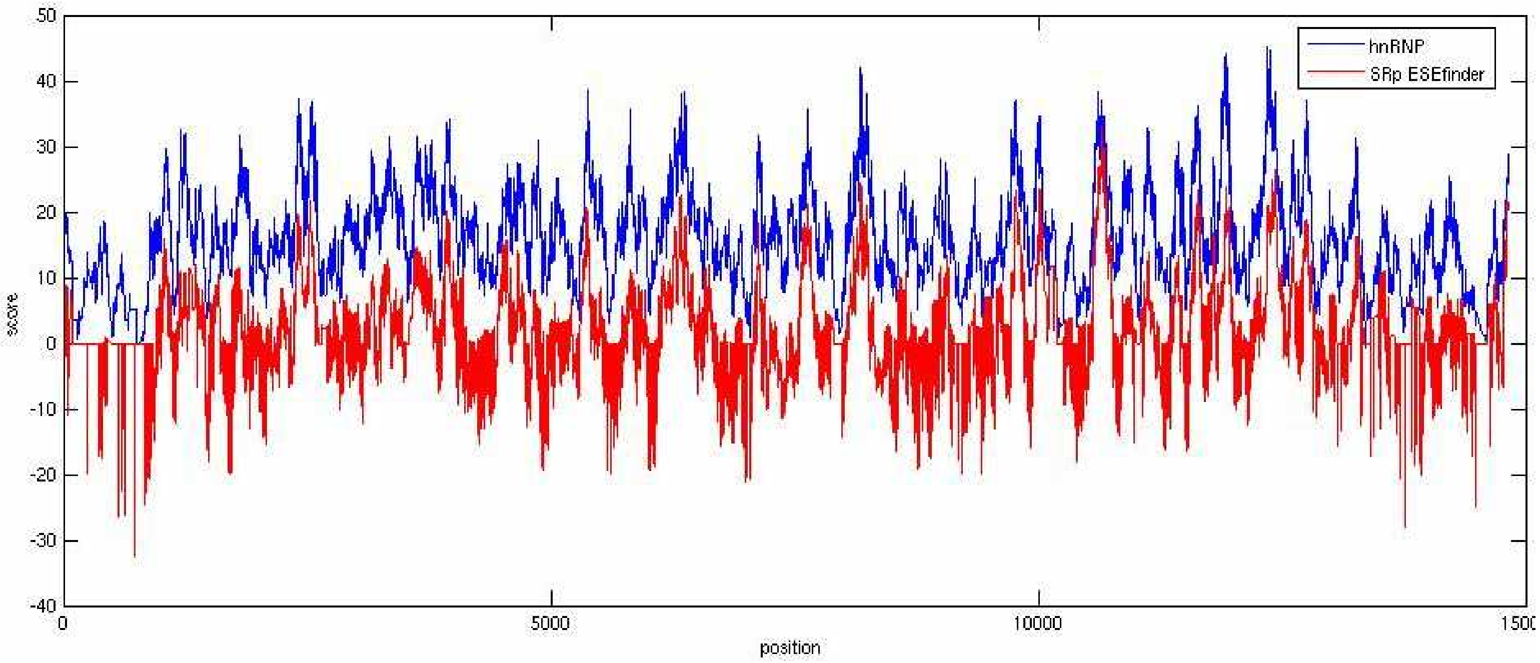
1st Approach

2nd Approach

Discussion

Acknowledgment

- Take WMs of ASF/SF2, SC35, SRp40, SRp55 from [Cartegni et al., 2003]
- Run PROFILER
- Compare output of hnRNP with output of these four WMs



Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Work Steps

- First Approach

► *Second Approach*

Using comparative genomic data

Work Steps

» Second Approach (I)

Outline
Motivation
Basics
Work Steps
1st Approach
2nd Approach
Discussion
Acknowledgment

- Inference of regulatory elements using phylogeny-sensitive methods and comparative genomic data
- Get exons which are internal and non-coding using fantom3DB
- Get pairwise alignments of
  - rhesus (rhemac1)
  - chimp (pantro1)
  - cow (bostau2)
  - human (hg17)
  - rat (rn3)

  aligned to mouse (mm7) from UCSC Genome Browser

Outline
Motivation
Basics
Work Steps
1st Approach
2nd Approach
Discussion
Acknowledgment

Work Steps

» Second Approach (II)

- For each pairwise alignment:
  - MultiZSearch to get all sequences from mm7 and aligned ones which occur in the wanted exon regions
- Modify output for alignments on same exon
  - If overlapping
    - Remove alignments
  - If difference of boundaries > 10% difference of exon sites
    - Remove alignments

    Else concatenate alignments
- For each internal, non-coding exon create a FASTA file containing exon sequence and corresponding pairwise alignments, remove gaps
- Realign sequences for each FASTA with ClustalW

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

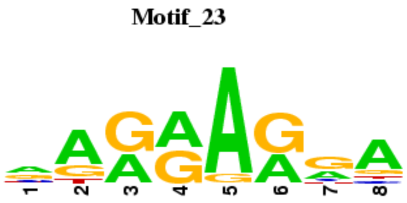
Acknowledgment

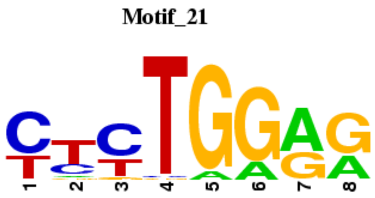
Work Steps

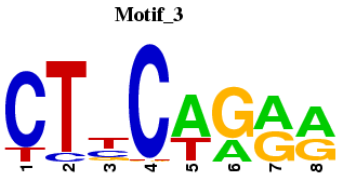
» Second Approach (III)

- Run PHYLOGIBBS
  - Comparative analysis of orthologues intergenic regions of related species

  → Identifies binding sites for regulatory proteins
  - Inputdata:
    - Aligned sequences splitted in 500 blocks per file
    - 10 motifs with length of 8 nt
  - Cluster WMs and create sequence logos with WEBLOGO





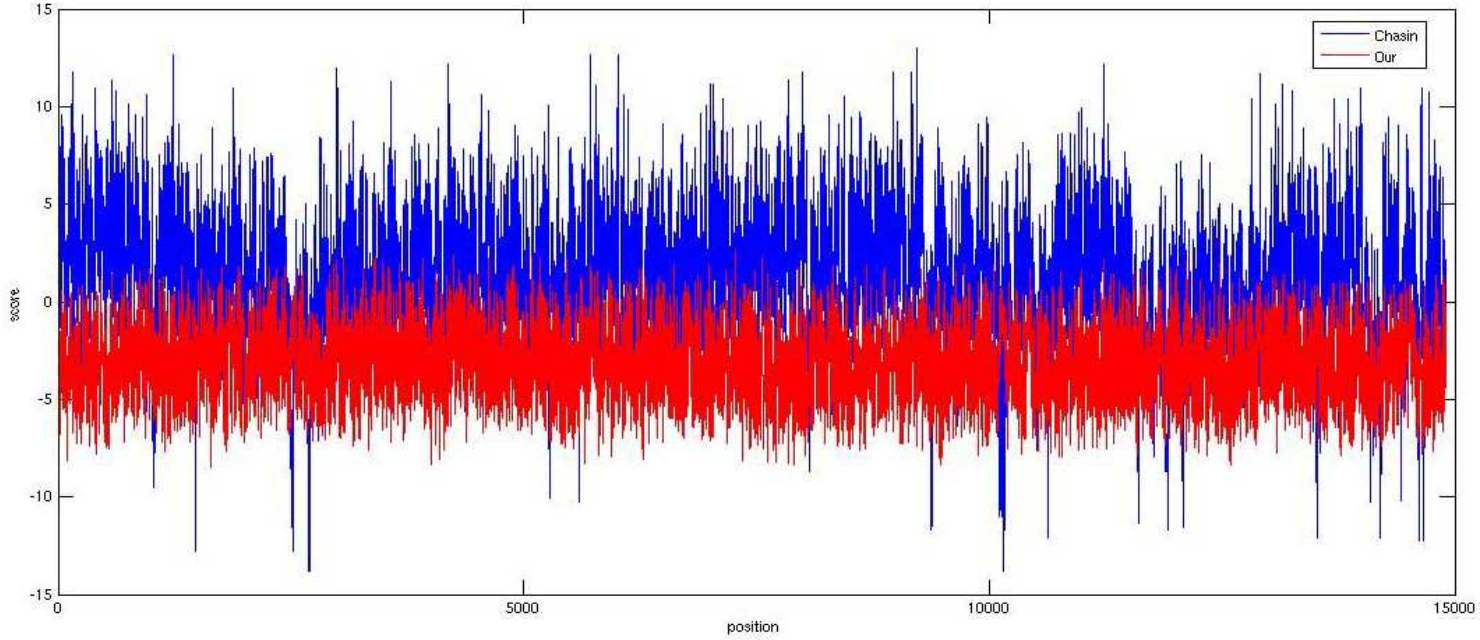


(out of 50 motifs)

Work Steps

» Second Approach (IV)

- Octamers with scores (if ES or EE) from [Chasin et al., 2004]
- Calculate max. log-likelihood over all WMs
- Make a profile for both scores with a random sequences (hg18, chr18, position 748411 to 763327)



Work Steps

» Second Approach (V)

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

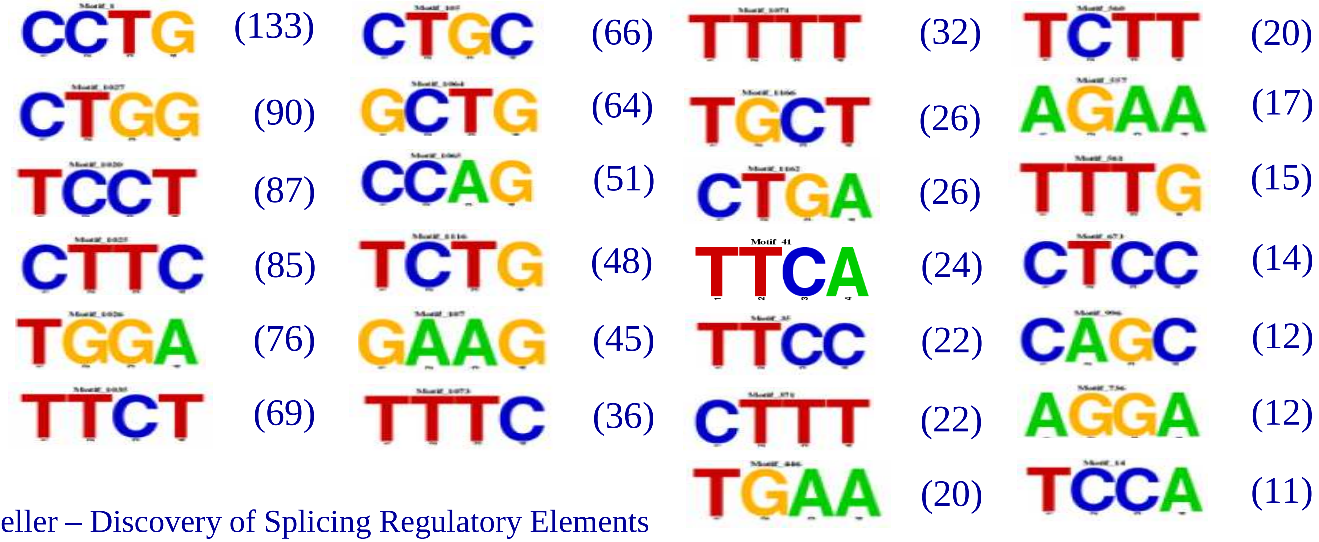
Discussion

Acknowledgment

- Final input data for PHYLOGIBBS
  - 100 datasets of aligned sequences splitted randomly in 500 blocks; 5 motifs with length 4 nt
- Analyse output of PHYLOGIBBS
  - Sequence logos of all WMs to look for conserved motifs

  → Many motifs which occurred several times

  → Extract WMs of motifs with occurrence > 10 (26/1246)



Work Steps

» Second Approach (VI)

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

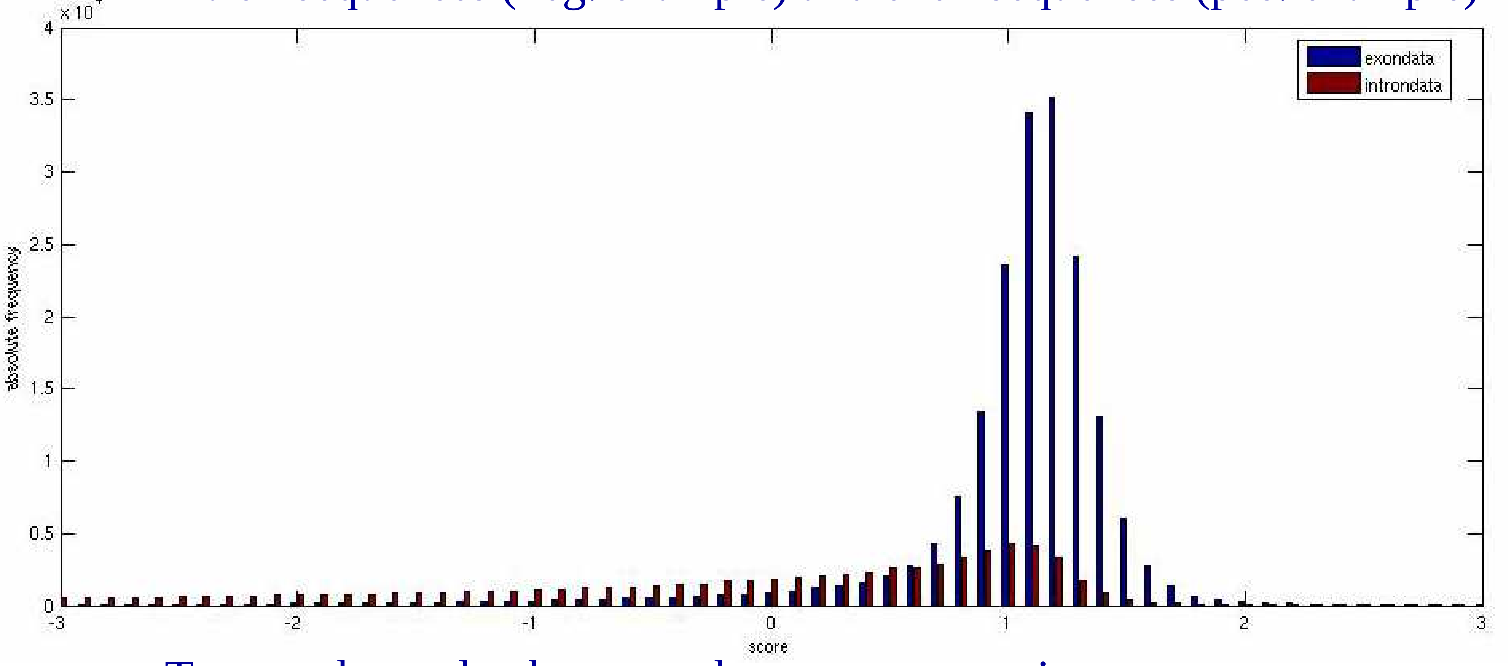
Acknowledgment

- Find out in which regions the motifs can be found
  - Appear as individuals
  - Appear in cluster
  - Predominantly in exon sequences
    - Whole sequences
    - Splice sites (10%)
  - Predominantly in intron sequences
    - Whole sequence
    - Flanking introns (size of exon sequence)

Work Steps

» Second Approach (VII)

- SVM^{light} from [Joachims, 2002] to decide with the motifs that have been found if a sequence is an exon or intron sequence
  - Create training sample for the known motifs
  - Intron sequences (neg. example) and exon sequences (pos. example)



  - Too much overlap between these two categories

    → Badly chosen motifs to define intron sequences ?

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Discussion

- Using SELEX data to find regulatory elements did not work out
- Using comparative genomic data with internal, non-coding exons
  - – Do the same for introns getting data from fantom3DB and compare the output of $\text{SVM}^{\text{light}}$ of exons
- Choose different motifs
- Use motifs with a different length
- Find other ways to calculate motifs

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Acknowledgment

Mihaela Zavolan and Group Members

*(Piotr Balwierzr, Philipp Berninger, Tzu-Ming Chern, Viktoria Dorfer,
Dimosthenis Gaidatzis, Jean Hausser, Nicodème Paul)*

BIOZENTRUM

SIB

Outline

Motivation

Basics

Work Steps

1st Approach

2nd Approach

Discussion

Acknowledgment

Thank you for your Attention!