Title

Genetic Algorithms for Genome Analysis
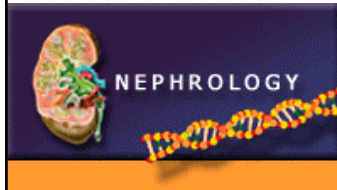
Paul Perco

Winterseminar, Bled
22-02-2005

Agenda

- Scientific field

- Project Overview (Microarrays)

- Genetic Algorithm for Promoter Module Detection

Scientific Field

Medical University of Vienna
Department of Nephrology

R. Oberbauer, Univ.Prof
P. Hauser, PhD
A. Kainz, PhD
C. Mitterbauer, MD

University of Innsbruck
Department of Nephrology

G. Mayer, Univ.Prof
M. Rudnicki, MD
C. Koppelstaetter, MD
J. Enrich, MTA

emergentec biodevelopment
Theoretical Chemistry Vienna

B. Mayer, PhD
A. Lukas, MSc

Stanford University
Department of Nephrology

  T. Meyer, Univ.Prof

Scientific Field

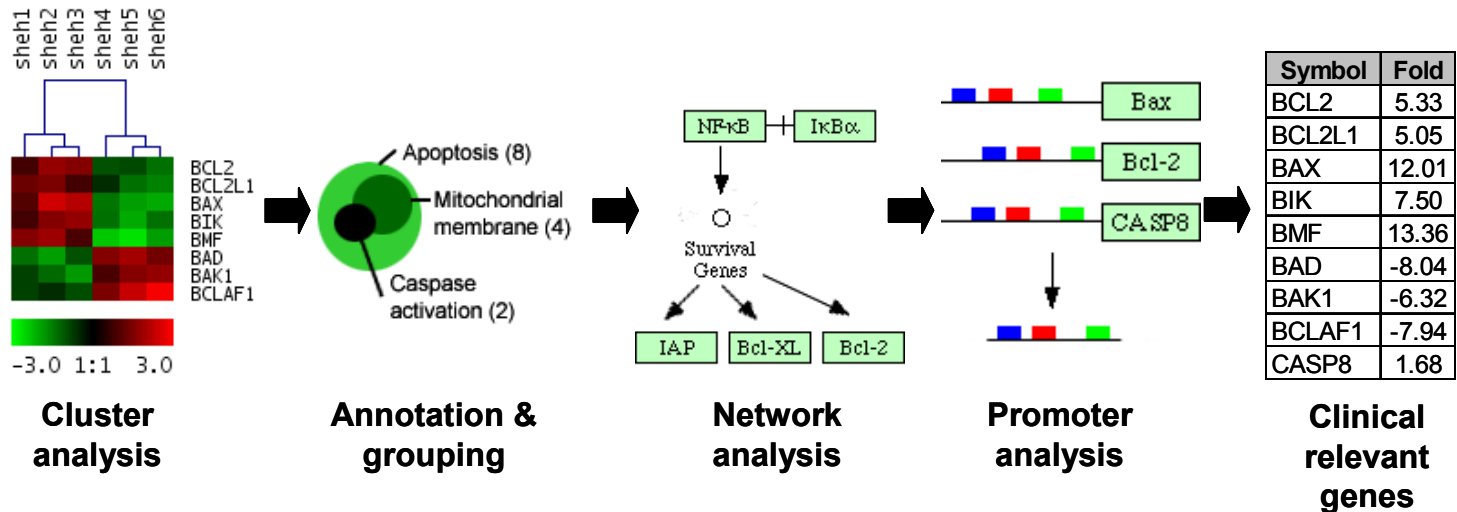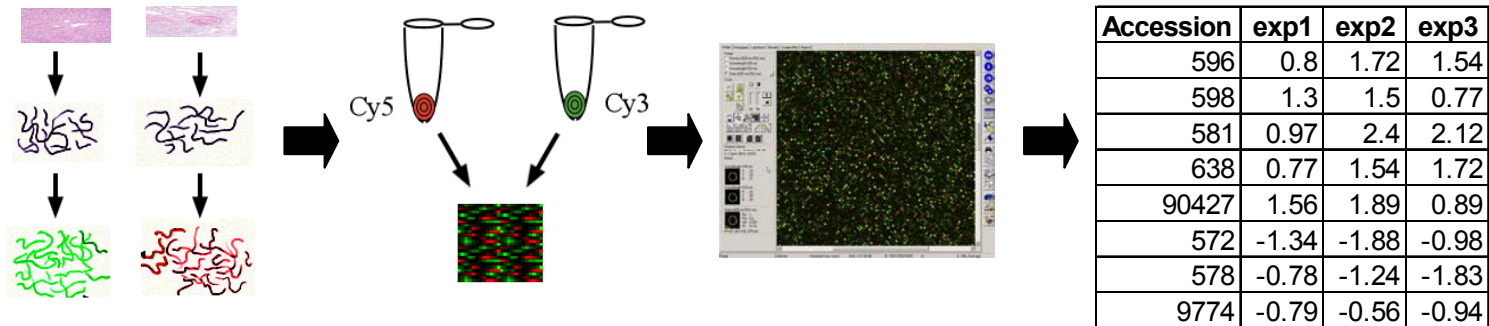Kidney transplantation

ARTF – acute renal transplant failure

       dialysis dependency within the first 7 days after transplantation
       in about 30% of cadaveric transplantations
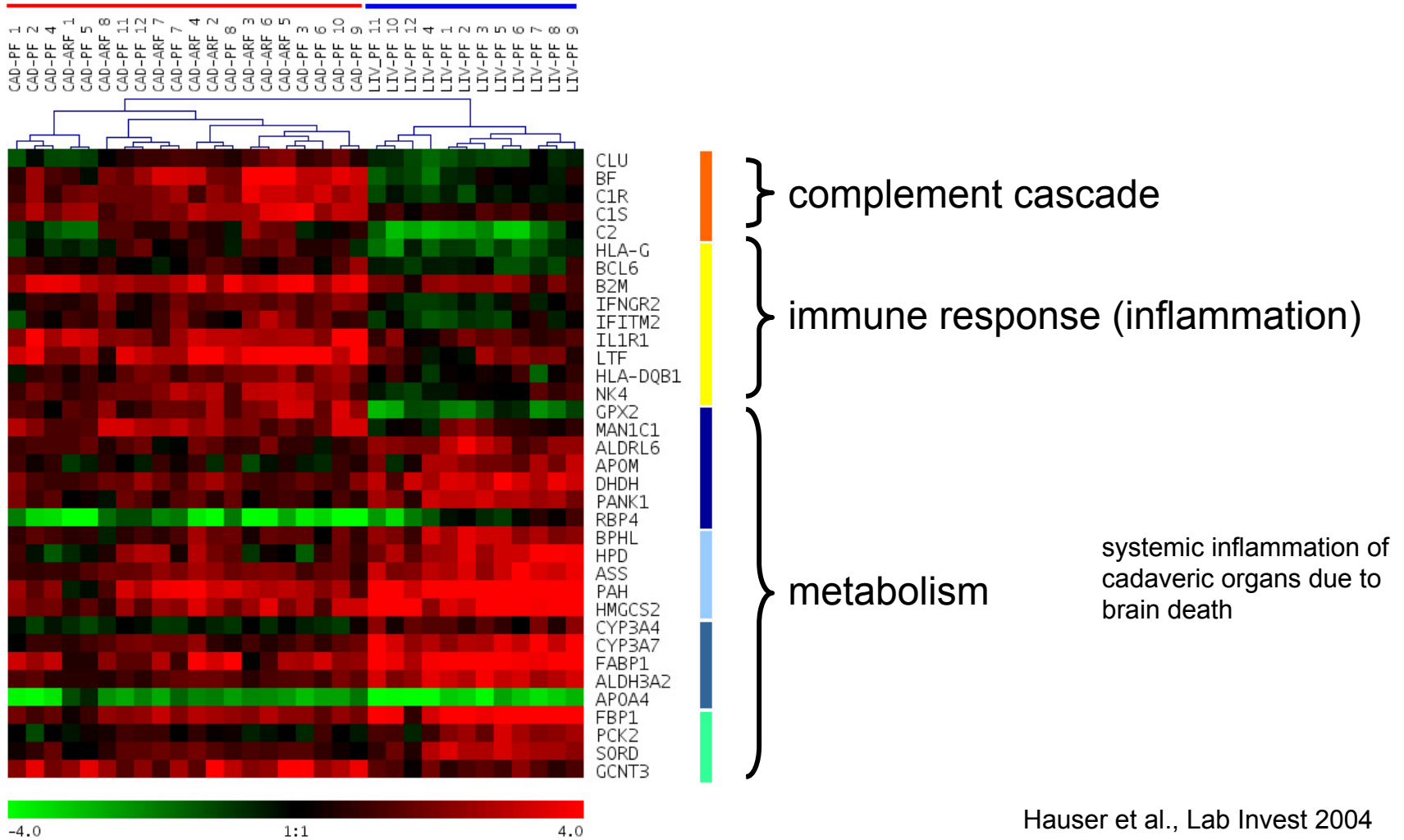   graft survival about 14 years (compared to 20)

       ARTF is the key risk factor for long-term graft survival

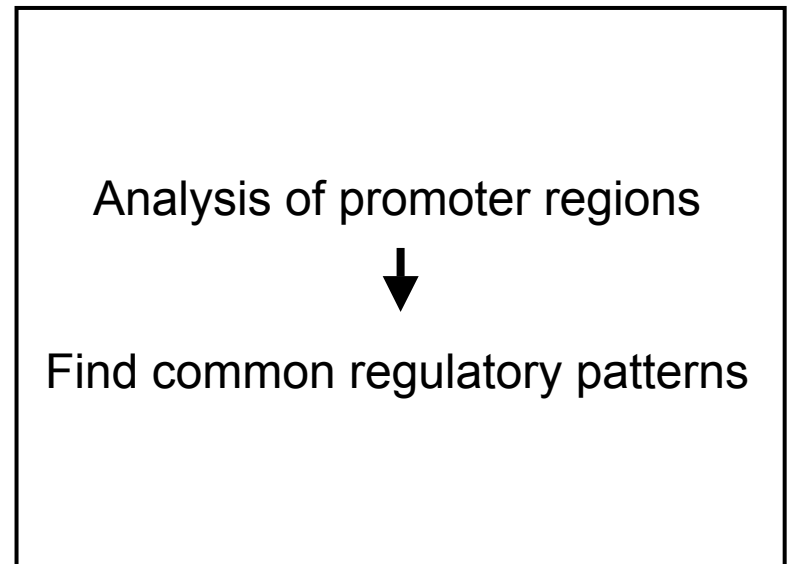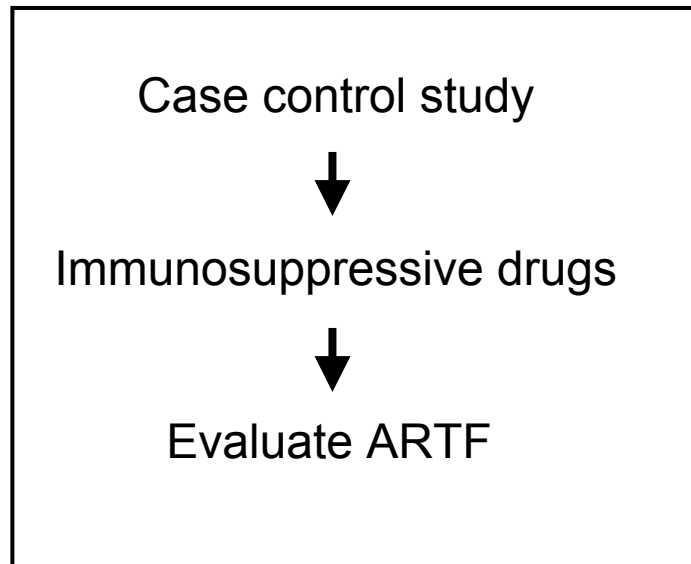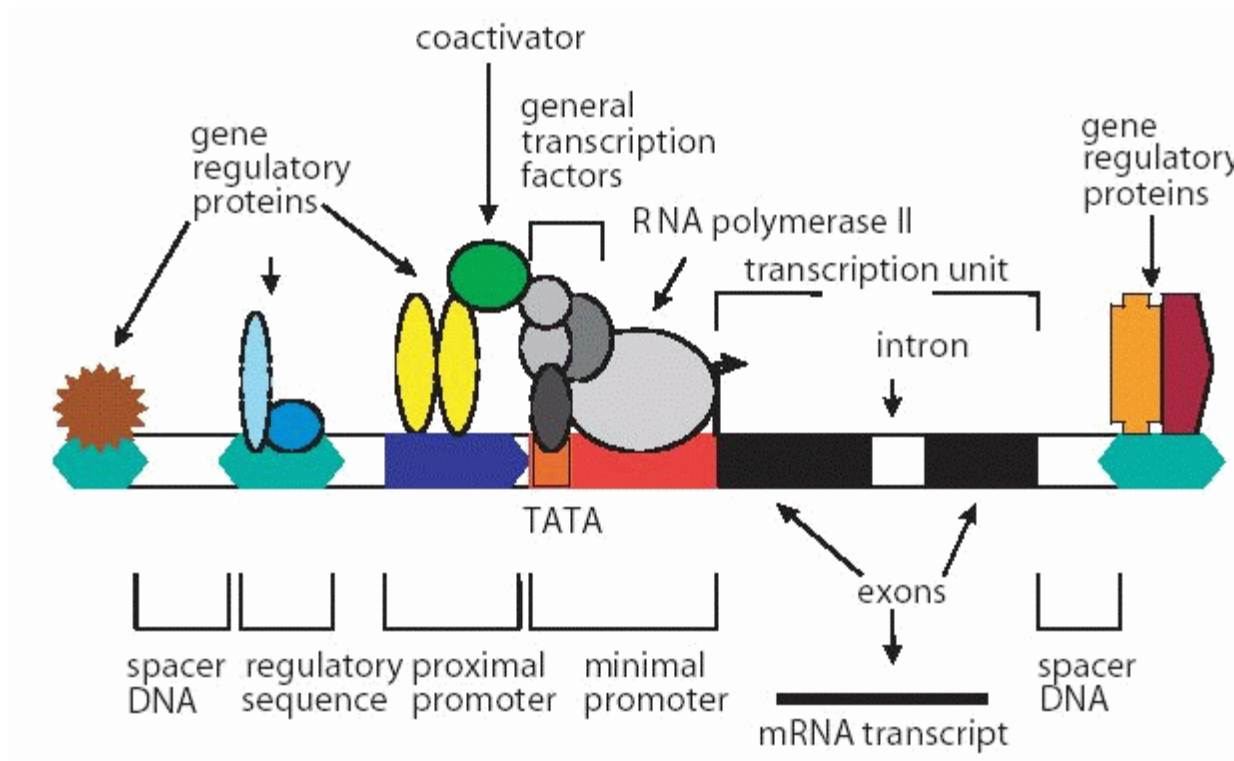Aim: decipher the molecular pathology of ARTF

Technology Overview

Accession	exp1	exp2	exp3
596	0.8	1.72	1.54
598	1.3	1.5	0.77
581	0.97	2.4	2.12
638	0.77	1.54	1.72
90427	1.56	1.89	0.89
572	-1.34	-1.88	-0.98
578	-0.78	-1.24	-1.83
9774	-0.79	-0.56	-0.94

Sample preparation   **Hybridization**   **Array image scanning**   **Raw expression data**

Symbol	Fold
BCL2	5.33
BCL2L1	5.05
BAX	12.01
BIK	7.50
BMF	13.36
BAD	-8.04
BAK1	-6.32
BCLAF1	-7.94
CASP8	1.68

Cluster analysis   **Annotation & grouping**   **Network analysis**   **Promoter analysis**   **Clinical relevant genes**

Transplant Kidneys

complement cascade

immune response (inflammation)

metabolism

systemic inflammation of cadaveric organs due to brain death

Hauser et al., Lab Invest 2004

Paul Perco   Bled 2005

Transplant Kidneys

Case control study

↓

Immunosuppressive drugs

↓

Evaluate ARTF

Analysis of promoter regions

↓

Find common regulatory patterns

Promoter Analysis

Villard et al., 2004

Promoter Analysis

Pattern Search algorithms
(Find known motifs stored in
diverse databases)

Pattern recognition algorithms

IUPAC
Consensus
Sequences

Position
Specific
Weight
Matrices
(PWMs)

Gibbs
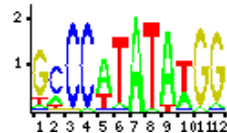Sampling

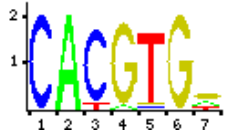Expectation
Maximization
(EM)
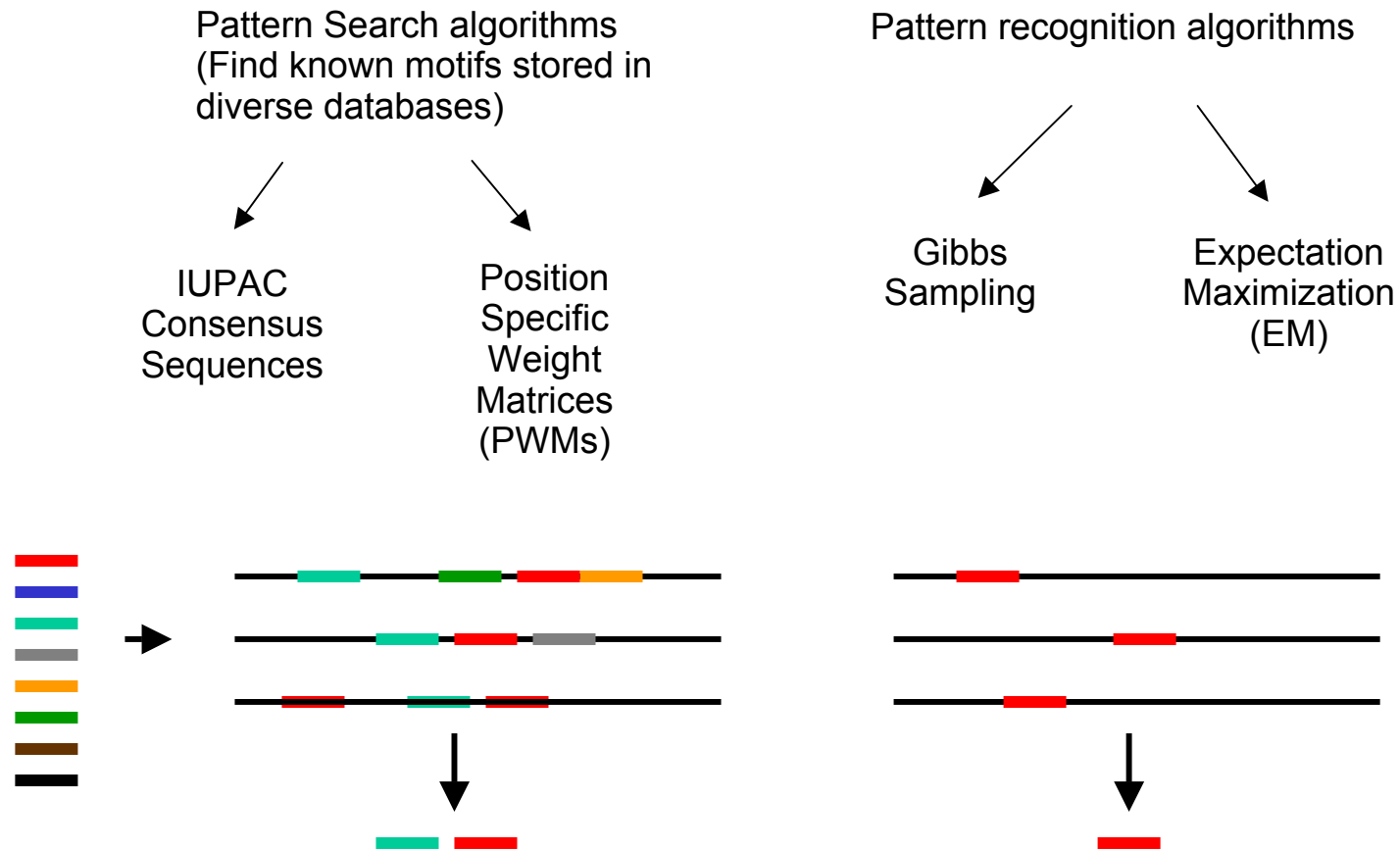
Genetic Algorithm   Aim

TFBS databases

TRANSFAC
JASPAR

USF
(upstream stimulating factor)
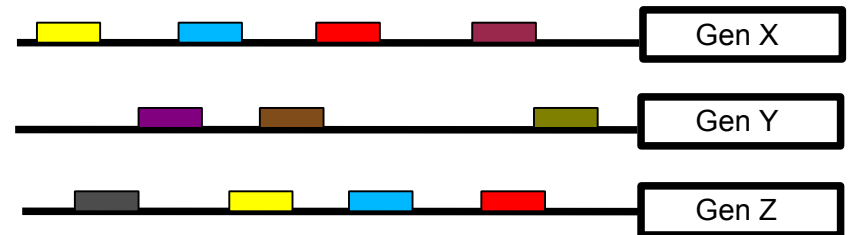
SRF
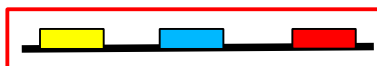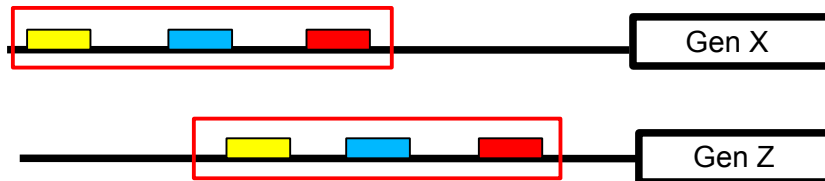(serum response factor)

① Search

motifs

Input sequences
Promoter sequences of coexpressed genes

Gen X

Gen Y

Gen Z

② Genetic algorithm (GA)

Find genes sharing a couple of TF-binding sites
and construct higher order "promoter modules"

Gen X

Gen Z

Promoter module indicating coregulation on transcriptional level

Genetic Algorithm   Input

	TFBS1	TFBS2	TFBS3	TFBS4	TFBS5 ...	TFBS N
Gen1	0	0	1	0	1	1
Gen2	1	1	1	1	1	0
Gen3	0	0	1	0	1	0
Gen4	1	1	1	0	0	0
Gen5	0	0	1	0	1	1
Gen6	0	0	1	0	1	1
Gen7	1	1	1	0	0	1
...						
Gen N	0	0	1	0	0	0

Pairwise Correlations:

Gen1	0	0	1	0	1	1
Gen2	1	1	1	1	1	0

Tanimoto Coefficient (Tan) = c/(a+b-c)

Genetic Algorithm   Input

	TFBS1	TFBS2	TFBS3	TFBS4	TFBS5	...	TFBS N
Gen1	0	0	1	0	1		1
Gen2	1	1	1	1	1		0
Gen3	0	0	1	0	1		0
Gen4	1	1	1	0	0		0
Gen5	0	0	1	0	1		1
Gen6	0	0	1	0	1		1
Gen7	1	1	1	0	0		1
...							
Gen N	0	0	1	0	0		0

	TFBS1	TFBS2	TFBS3
Gen2	1	1	1
Gen4	1	1	1
Gen7	1	1	1

	TFBS3	TFBS5	TFBS N
Gen1	1	1	1
Gen5	1	1	1
Gen6	1	1	1

1. Generation of an initial population I

I_1		1	0	1	1
I_2		1	0	1 0	
I_n	0	0	1 0	

```
0  1  0  0  0  0  0  0  0  0
0  1  0  1  0  1  0  1  1  1
1  0  0  0  0  0  1  1  0  1
.  .  .  .  .  .  .  .  .  .
.  .  .  .  .  .  .  .  .  .
1  1  1  0  0  0  0  1  1  0
0  0  0  1  0  1  0  1  1  0
1  1  1  0  1  0  1  0  0  1
```

2. Evaluation of each individual I_K of population I

I_1				1		0		1	1
I_2					1	0		1	0
I_n			0		0		1	0	

Tan_m^2	Tan_m^2	Tan_m^2											
0.11	0.00	0.00	0	1	0	0	0	0	0	0	0	0
9.00	1.00	0.00	0	1	0	1	0	1	0	1	1	1
1.78	0.00	0.11	1	0	0	0	0	0	1	1	0	1
			.	.	.	.	.	.	.	.	.	.
			.	.	.	.	.	.	.	.	.	.
1.78	0.25	0.00	1	1	1	0	0	0	0	1	1	0
0.11	1.78	0.00	0	0	0	1	0	1	0	1	1	0
1.00	0.00	0.25	1	1	1	0	1	0	1	0	0	1

F_1	F_2	F_n
13.78	3.03	0.36

P_1	P_2	P_n
0	0.78	0.97

3. Generation of a child population

No mutation

| I_1 | | 1 | 0 | 1 | 1 |

| I_1* | | 1 | 0 | 1 | 1 |

Bit flip flop (1 to 0)

4. Loop to step 2 until convergence criterion is met

| I_2 | | 1 | 0 | 1 | 0 |

| I_2* | | 0 | 0 | 1 | 0 |

Positional change

| I_3 | | 0 | 0 | 1 | 0 |

| I_3* | | 0 | 0 | 1 | 0 |
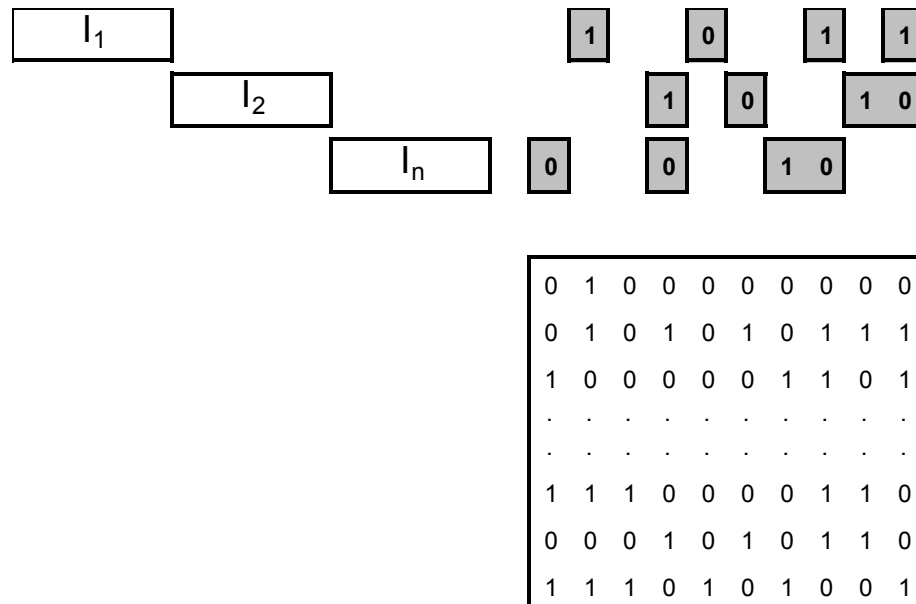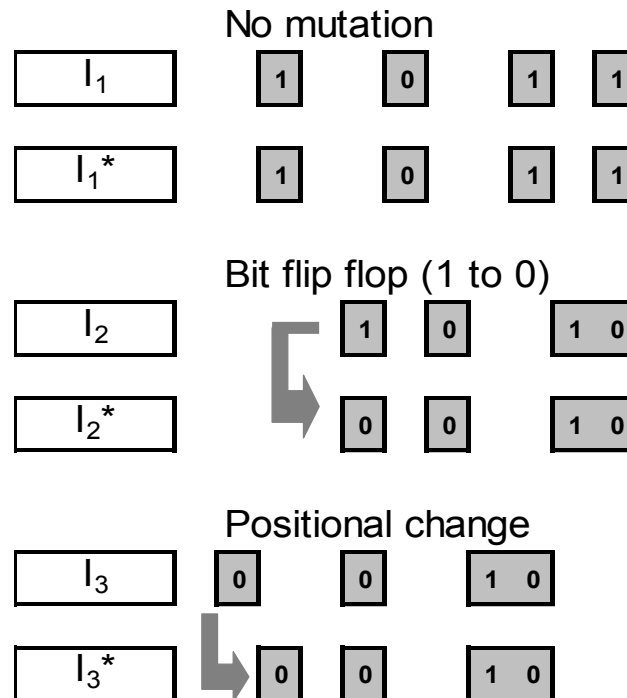
1. Generation of an initial population I

2. Evaluation of each individual I_K of population I
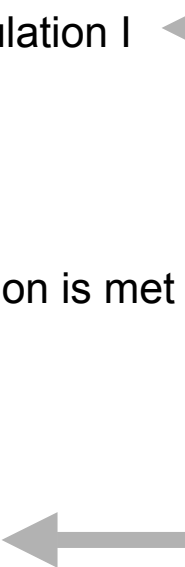
3. Generation of a child population

4. Loop to step 2 until convergence criterion is met

5. Stop GA and print out top individuals

Genetic Algorithm   Testing I

	Setup 1	Setup 2	Setup 3
String Length	8	6 - 10	8
Module Length	3 - 6	6	4
Module No.	1	1	3

B *B* *B*

7 ACTB genes from different organisms
13 randomly picked genes from the EPD

	V$E2F1	V$CAAT	V$NFY	V$MTATA	V$TFIIA	V$SRF	V$USF	V$LEF1	V$TATA	V$TCF4	V$HNF 4
GG_ACTB	1	1	1	1	1	1	1	1	1	1	0
HS_ACTB	1	1	1	1	0	1	0	1	1	1	0
RN_ACTB	1	1	1	1	1	1	0	1	1	1	0
CG_ACTB	0	1	1	1	1	1	0	1	1	1	0
MM_ACTB	0	1	1	0	1	1	0	0	1	0	1
MA_ACTB	0	1	1	1	0	1	0	1	1	1	1
OL_ACTB	1	1	1	1	1	1	1	0	1	1	1
HS_PU16	0	0	0	0	0	0	1	0	0	0	1
HS_RPS14	1	0	0	0	0	0	1	0	0	0	0
HS_TUBA1	1	0	0	0	0	0	1	0	0	0	0
HS_FN1	1	0	0	0	1	0	0	0	0	0	0
HS_TK1	1	1	1	1	1	0	1	1	0	0	0
HS_IL2	0	0	0	1	0	0	0	1	0	1	0
HS_TFRC	0	0	0	1	0	0	0	0	0	0	0
HS_ELK1	1	1	1	0	0	0	1	0	0	0	1
HS_CARS	1	0	0	0	0	0	0	0	0	0	0
HS_NONO	1	1	1	0	0	0	0	0	0	0	1
HS_INDO	0	1	0	0	1	0	0	0	0	0	0
HS_IK	1	0	0	0	0	0	0	0	0	0	0
HS_RAG1	0	1	1	0	0	0	0	0	0	0	0
Sum	12	12	11	9	8	7	7	7	7	7	6

Conclusion and Outlook

- Molecular signature of ARTF

- Genetic Algorithm is able to identify promoter modules

- Combine Genetic Algorithm with phylogenetic footprinting

End

Thanks for your attention!