Hands on session:
Basic promoter analysis

Thomas Werner
CEO&CSO
Genomatix Software GmbH
Landsberger Strasse 6, D-80339 München
http://www.genomatix.de

A real life project

A project fully integrating in silico & in vitro methods

This application example was taken from a scientific project that had lasted for about five years and included experimental verification of every prediction made by bioinformatics analysis.

The results of the combined *in silico* and *in vitro* apporaches were published in two articles:

Benkhart E., Siedler M., Wedel A., Werner T., Ziegler-Heitbrock H.W.L. (2000) Role of Stat3 in Lipopolysaccharide-Induced IL-10 Gene Expression. J. Immunol. 165, 1612 - 1617.

Ziegler-Heitbrock H.W.L., Lötzerich M., Schaefer A., Werner T., Frankenberger M., Benkart E. (2003). IFN-a Induces Human IL-10 Gene by Recruiting Both IFN Regulatory Factor I and Stat3. J. Immunol. 171, 285 - 290.

Outline of tasks

What you will be doing during the next 2 hours

- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (***ElDorado***)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

Genomatix

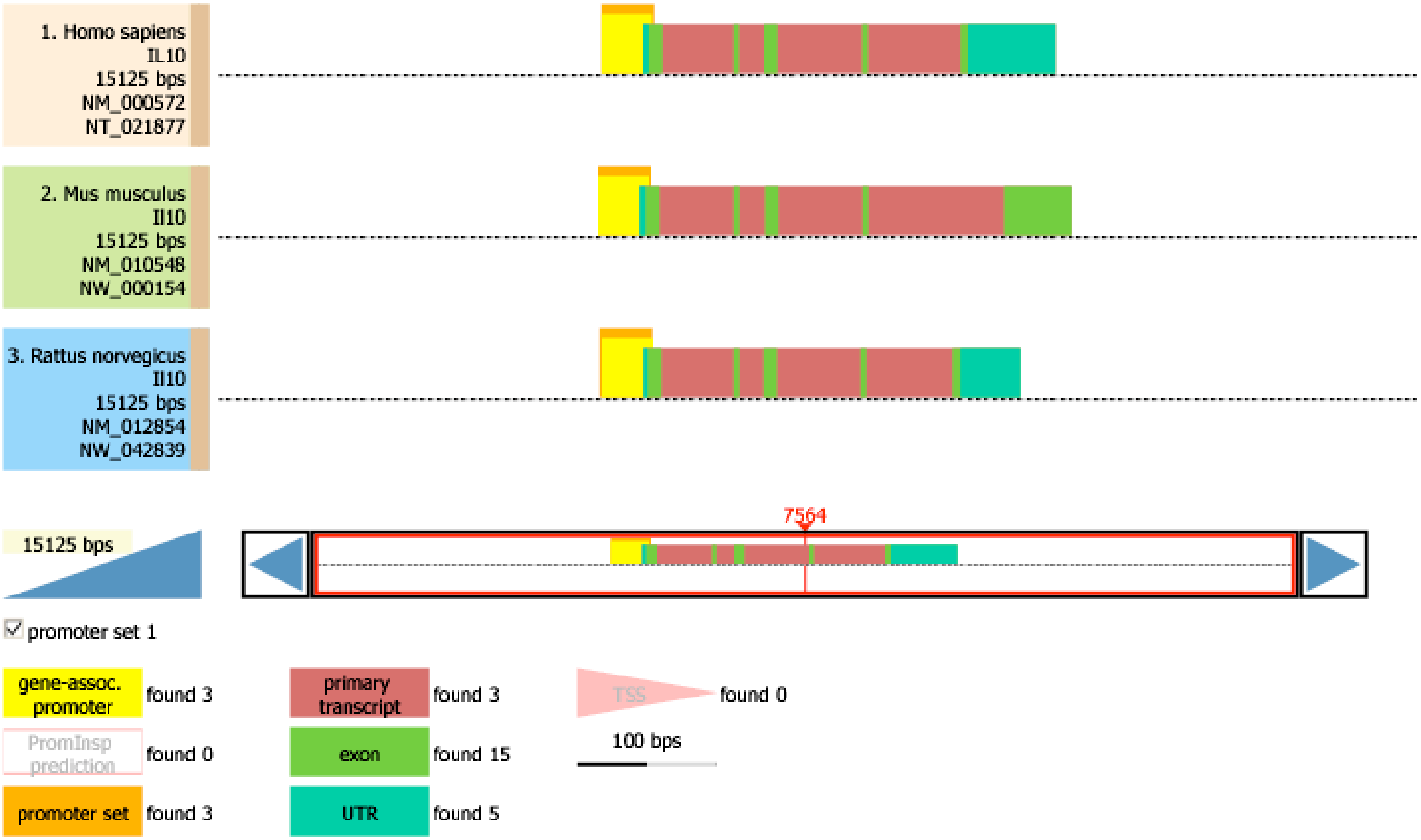
Task 1

What you will be doing next

- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (**ElDorado**)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

Genomatix ElDorado is a unique promoter source

ElDorado provides human, mouse, and rat IL-10 promoters



1. Homo sapiens
IL10
15125 bps
NM_000572
NT_021877

2. Mus musculus
Il10
15125 bps
NM_010548
NW_000154

3. Rattus norvegicus
Il10
15125 bps
NM_012854
NW_042839

☑ promoter set 1

gene-assoc. promoter	found 3	primary transcript	found 3	TSS	found 0
PromInsp prediction	found 0	exon	found 15		
promoter set	found 3	UTR	found 5		

Task 2

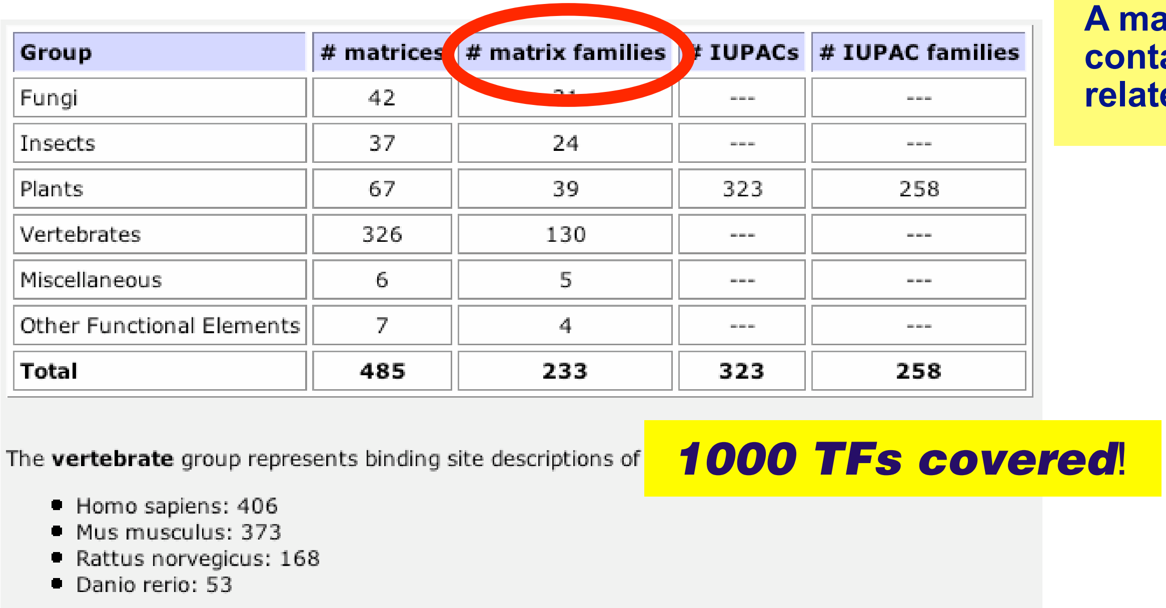
What you will be doing next

- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (***ElDorado***)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

Genomatix IL-10 project initial in silico analysis

***MatInspector*: Finding TF binding sites by weight matrices**

- A IUPAC consensus sequence cannot tolerate mismatches
- A weight matrix can tolerate mismatches



Group	# matrices	# matrix families	# IUPACs	# IUPAC families
Fungi	42	[illegible]	---	---
Insects	37	24	---	---
Plants	67	39	323	258
Vertebrates	326	130	---	---
Miscellaneous	6	5	---	---
Other Functional Elements	7	4	---	---
Total	**485**	**233**	**323**	**258**

The **vertebrate** group represents binding site descriptions of

- Homo sapiens: 406
- Mus musculus: 373
- Rattus norvegicus: 168
- Danio rerio: 53

***1000 TFs covered*!**

A matrix family contains several related matrices

MatInspector can rapidly identify candidate binding sites and factors

IL-10 project TFBS analysis

MatInspector analysis of the human IL-10 promoter


132 matches found.

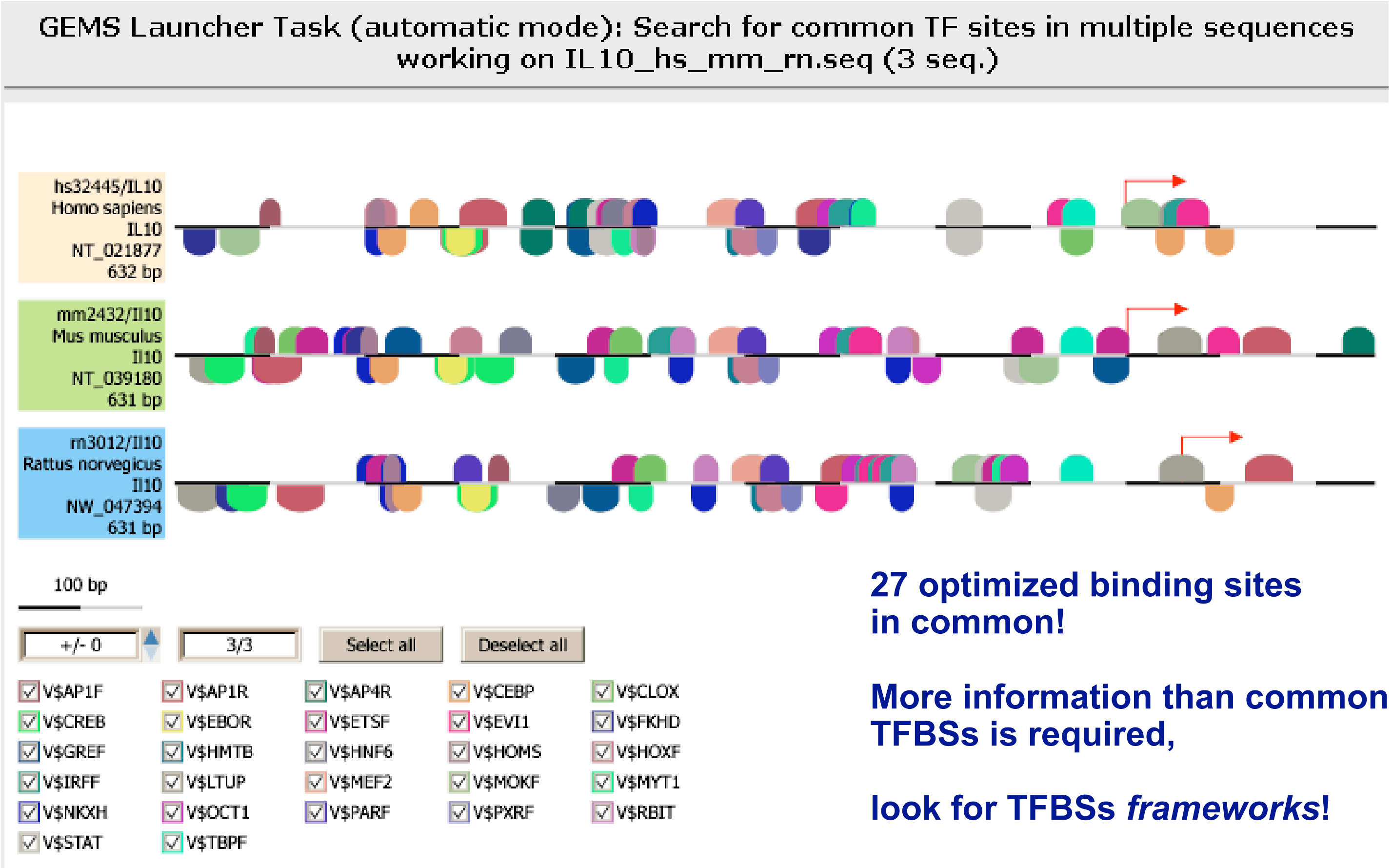
Human IL-10 promoter

```
TCCAGCCTCCATGGAATCCTGACTTCTTTTCCTTGTTATTTCAACTTCTTCCACCCCATCTTTTAAACTTTAGACTCCAG
CCACAGAAGCTTACAACTAAAAGAAACTCTAAGGCCAATTTAATCCAAGGTTTCATTCTATGTGCTGGAGATGGTGTACA
GTAGGGTGAGGAAACCAAATTCTCAGTTGGCACTGGTGTACCCTTGTACAGGTGATGTAACATCTCTGTGCCTCAGTTTG
CTCACTATAAAATAGAGACGGTAGGGGTCATGGTGAGCACTACCTGACTAGCATATAAGAAGCTTTCAGCAAGTGCAGAC
TACTCTTACCCACTTCCCCCAAGCACAGTTGGGGTGGGGGACAGCTGAAGAGGTGGAAACATGTGCCTGAGAATCCTAAT
GAAATCGGGGTAAAGGAGCCTGGAACACATCCTGTGACCCCGCCTGTCCTGTAGGAAGCCAGTCTCTGGAAAGTAAAATG
GAAGGGCTGCTTGGGAACTTTGAGGATATTTAGCCCACCCCCTCATTTTTACTTGGGGAAACTAAGGCCCAGAGACCTAA
GGTGACTGCCTAAGTTAGCAAGGAGAAGTCTTGGGTATTCATCCCAGGTTGGGGGGACCCAATTATTTCTCAATCCCATT
GTATTCTGGAATGGGCAATTTGTCCACGTCACTGTGACCTAGGAACACGCGAATGAGAACCCACAGCTGAGGGCCTCTGC
GCACAGAACAGCTGTTCTCCCAGGAAATCAACCTTTTTTTAATTGAGAAGCTAAAAAATTATTCTAAGAGAGGTAGCCCA
TCCTAAAAATAGCTGTAATGCAGAAGTTCATGTTCAACCAATCATTTTTGCTTACCATCCAAAAAATTGAAAACTAAGTTT
ATTAGAGAGGTTAGAGAAGGAGGAGCTCTAAGCAGAAAAATCCTGTGCCGGGAAACCTTGATTGTGGCTTTTTAATGCAA
TGAAGAGGCCTCCCTGAGCTTACAATATAAAAGGGGGACAGAGAGGTGAAGGTCTACACATCAGGGGCTTGCTCTTGCAA
AACCAAACCACAAGACAGACTTGCAAAAGAAGGC
```

Genomatix

Beyond simple TFBS analysis

TFBSs shared between human mouse and rat promoter



GEMS Launcher Task (automatic mode): Search for common TF sites in multiple sequences
working on IL10_hs_mm_rn.seq (3 seq.)

☑ V$AP1F ☑ V$AP1R ☑ V$AP4R ☑ V$CEBP ☑ V$CLOX
☑ V$CREB ☑ V$EBOR ☑ V$ETSF ☑ V$EVI1 ☑ V$FKHD
☑ V$GREF ☑ V$HMTB ☑ V$HNF6 ☑ V$HOMS ☑ V$HOXF
☑ V$IRFF ☑ V$LTUP ☑ V$MEF2 ☑ V$MOKF ☑ V$MYT1
☑ V$NKXH ☑ V$OCT1 ☑ V$PARF ☑ V$PXRF ☑ V$RBIT
☑ V$STAT ☑ V$TBPF

27 optimized binding sites in common!

More information than common TFBSs is required,

look for TFBSs *frameworks*!

Genomatix

Task 3

What you will be doing next

- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (***ElDorado***)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

Genomatix

IL-10 promoter frameworks

Exploit knowledge about the IL-10 promoters

- IL-10 is very well covered in the literature (more than 1000 papers!)
- At this point we are only interested in TFs linked to IL-10
- ***BiblioSphere*** automatically analyzes the whole PubMed for TF links
- Get candidates from the IL-10 ***BiblioSphere***!
- Start with the TFs with BSs in the IL-10 promoter and cocited with IL-10

Focus the framework analysis on likely candidate TFBSs already known

Genomatix

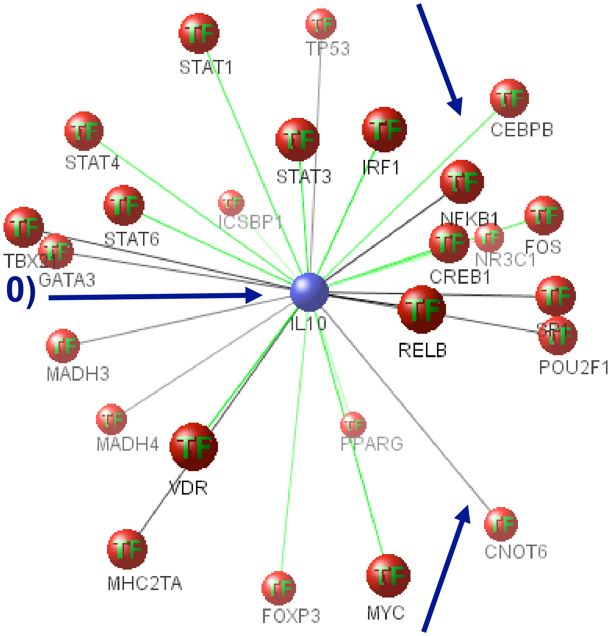
IL-10 promoter frameworks

Transcriptional context of IL-10 identified by ***BiblioSphere***

- IL-10 is cocited with 1209 genes
- 96 are transcription factors
- ***BiblioSphere*** determined TFBS for 10 TFs in the IL-10 promoter
  - **AP1**
  - **CEBP**
  - **CREB**
  - **EBOX**
  - **FKHD**
  - **GRE**
  - **IRF**
  - **PERO**
  - **RXRF**
  - **STAT**

Green connection:
Binding site confirmed
In promoter sequence

Reference gene (IL-10)



Gray connection:
Binding site not found
In promoter sequence

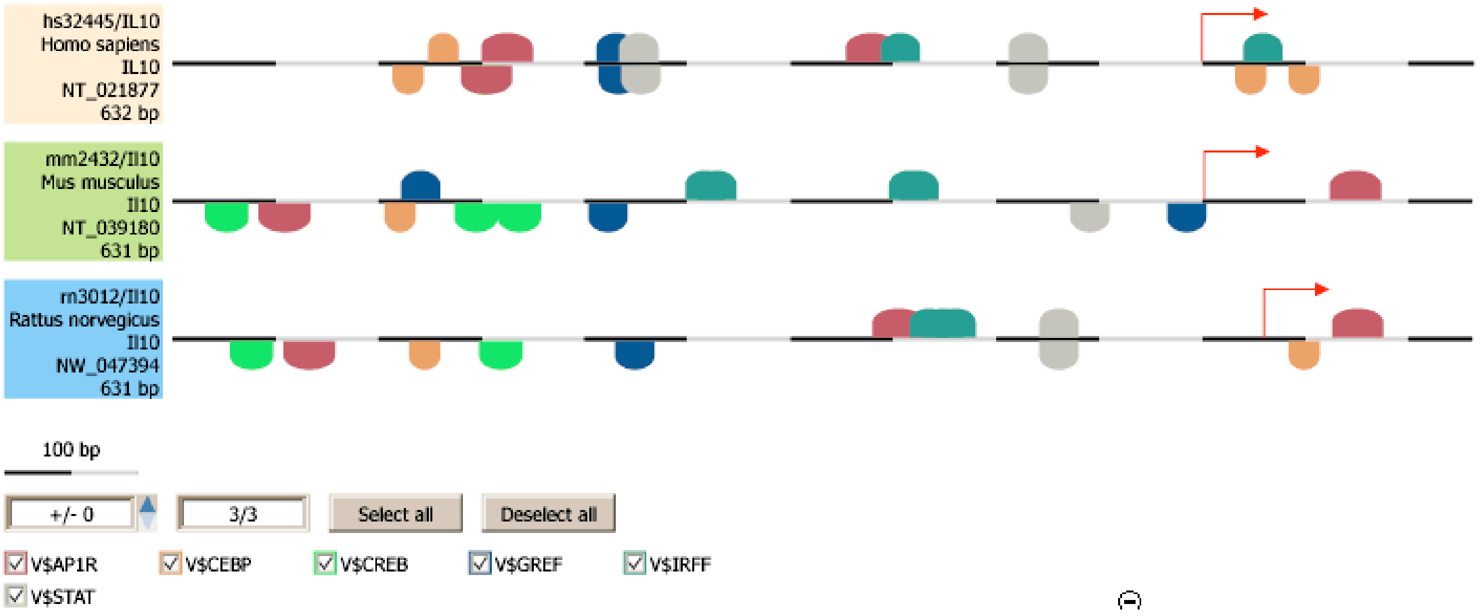
Parameters (default)
TFs, cocited on sentence level
minimum 2 cocitations

Genomatix

IL-10 promoter frameworks

Common TFs of human and mouse IL-10 promoter

- Only the BS for the 10 TFs identified by ***BiblioSphere*** are shown



+/- 0 | 3/3

☑ V$AP1R ☑ V$CEBP ☑ V$CREB ☑ V$GREF ☑ V$IRFF
☑ V$STAT

Frameworks are not obvious even with only 6 TFBSs to be considered!

Task 4

What you will be doing next

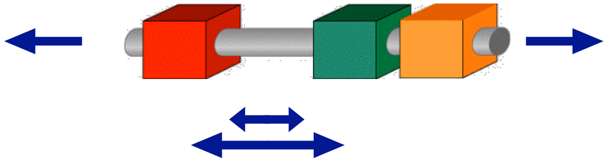
- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (***ElDorado***)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

Genomatix

FrameWorks and Modules

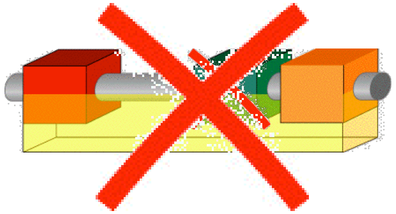
Describing promoter/enhancer TFBSs organization

- A *framework* describes a conserved set of transcription elements



- The distances of TFBSs within a *framework* are variable as is the position of the whole framework within different sequences

- A *module* describes experimentally verified mutually dependent functional transcription elements



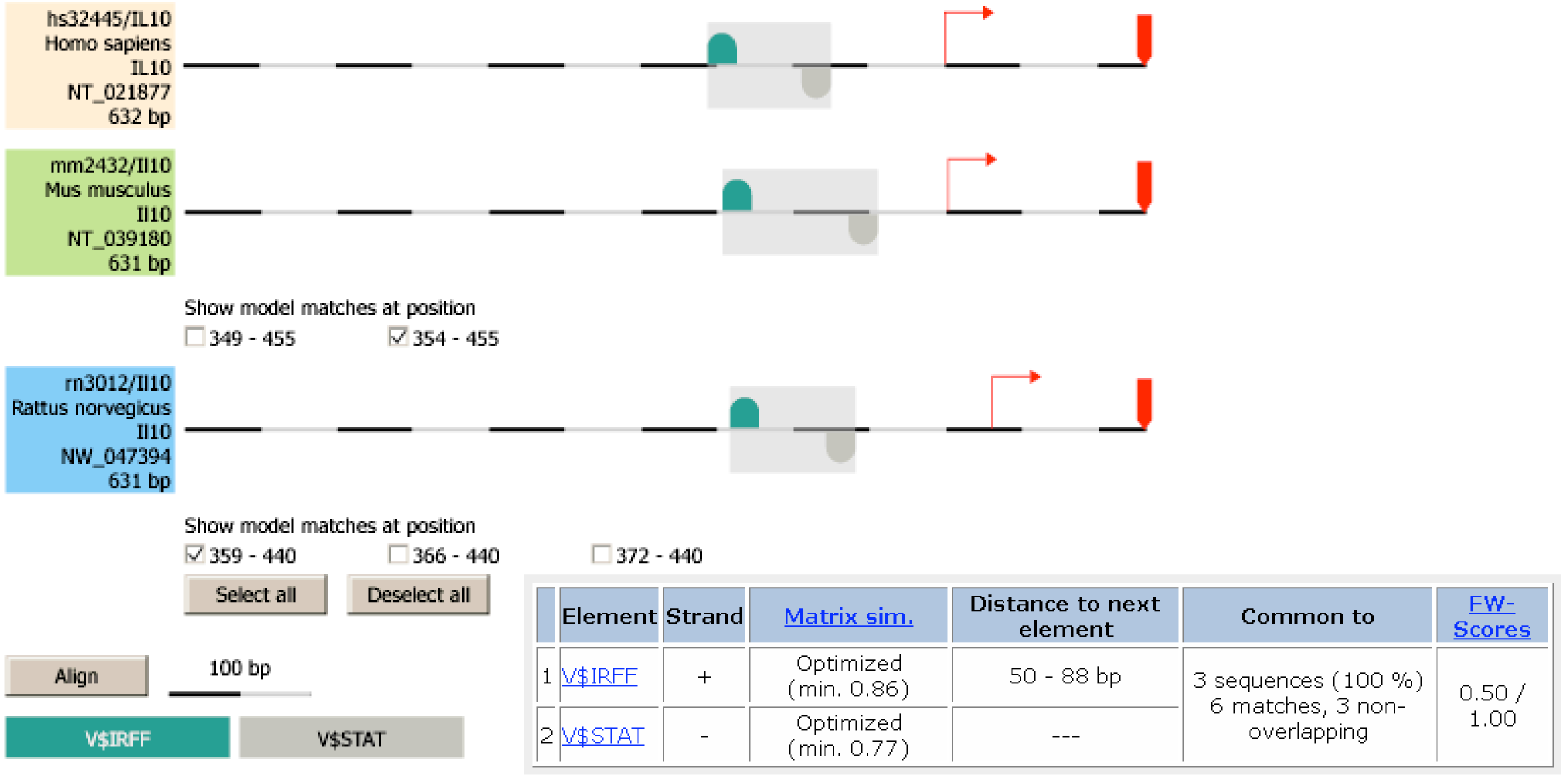
Delete a single site - lose function of all

Modules definitely have a biological function, frameworks may have a function

Genomatix

IL-10 promoter frameworks

FrameWorker identified a single framework including the functional STAT binding site



Show model matches at position
☐ 349 - 455 ☑ 354 - 455

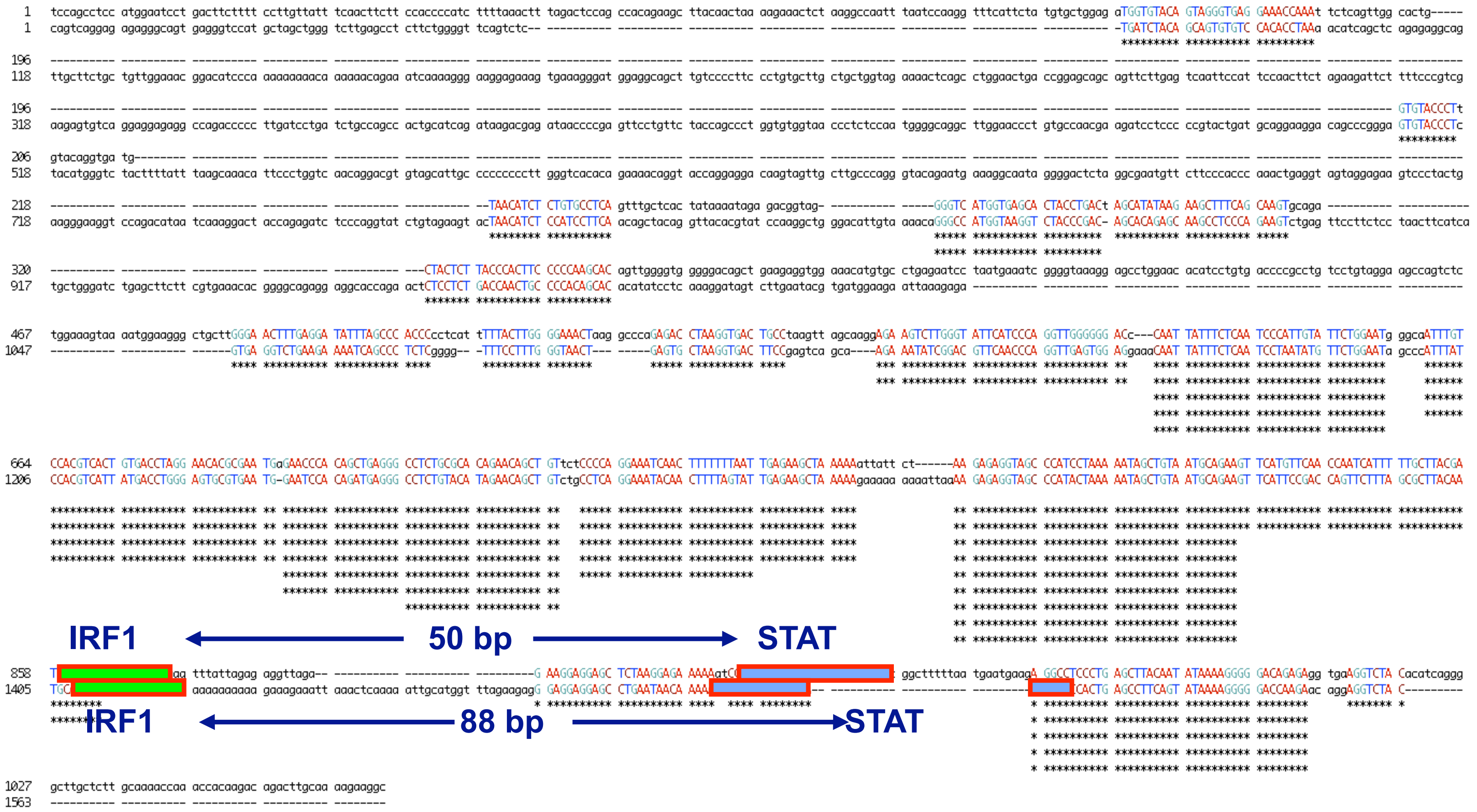
Show model matches at position
☑ 359 - 440 ☐ 366 - 440 ☐ 372 - 440

	Element	Strand	Matrix sim.	Distance to next element	Common to	FW-Scores
1	V$IRFF	+	Optimized (min. 0.86)	50 - 88 bp	3 sequences (100 %) 6 matches, 3 non-overlapping	0.50 / 1.00
2	V$STAT	-	Optimized (min. 0.77)	---		

Genomatix

IL-10 promoter frameworks

The IRF1-STAT TFBS pair is conserved, the sequence is not!



```
   1  tccagcctcc atggaatcct gacttctttt ccttgttatt tcaacttctt ccaccccatc ttttaaactt tagactccag ccacagaagc ttacaactaa aagaaactct aaggccaatt taatccaagg tttcattcta tgtgctggag aTGGTGTACA GTAGGGTGAG GAAACCAAAt tctcagttgg cactg-----
   1  cagtcaggag agagggcagt gagggtccat gctagctggg tcttgagcct cttctggggt tcagtctc-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -TGATCTACA GCAGTGTGTC CACACCTAAa acatcagctc agagaggcag
                                                                                                                                                                                    ********* ********** *********

 196  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
 118  ttgcttctgc tgttggaaac ggacatccca aaaaaaaaca aaaaacagaa atcaaaaggg aaggagaaag tgaaagggat ggaggcagct tgtccccttc cctgtgcttg ctgctggtag aaaactcagc ctggaactga ccggagcagc agttcttgag tcaattccat tccaacttct agaagattct tttcccgtcg

 196  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- GTGTACCCTt
 318  aagagtgtca ggaggagagg ccagaccccc ttgatcctga tctgccagcc actgcatcag ataagacgag ataaccccga gttcctgttc taccagccct ggtgtggtaa ccctctccaa tggggcaggc ttggaaccct gtgccaacga agatcctccc ccgtactgat gcaggaagga cagcccggga GTGTACCCTc
                                                                                                                                                                                                                       *********

 206  gtacaggtga tg-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
 518  tacatgggtc tacttttatt taagcaaaca ttccctggtc aacaggacgt gtagcattgc cccccccctt gggtcacaca gaaaacaggt accaggagga caagtagttg cttgcccagg gtacagaatg aaaggcaata ggggactcta ggcgaatgtt cttcccaccc aaactgaggt agtaggagaa gtccctactg

 218  ---------- ---------- ---------- ---------- ---------- ---------- --TAACATCT CTGTGCCTCA gtttgctcac tataaataga gacggtag- ---------- -----GGGTC ATGGTGAGCA CTACCTGACt AGCATATAAG AAGCTTTCAG CAAGTgcaga ---------- ----------
 718  aagggaaggt ccagacataa tcaaaggact accagagatc tcccaggtat ctgtagaagt acTAACATCT CCATCCTTCA acagctacag gttacacgtat ccaaggctg ggacattgta aaacaGGGCC ATGGTAAGGT CTACCCGAC- AGCACAGAGC AAGCCTCCCA GAAGTctgag ttccttctcc taacttcatca
                                                                          ******** **********                                                  ***** ********** *********  ********** ********** *****
                                                                                                                                               ***** ********** *********

 320  ---------- ---------- ---------- ---------- ---------- ---CTACTCT TACCCACTTC CCCCAAGCAC agttggggtg ggggacagct gaagaggtgg aaacatgtgc ctgagaatcc taatgaaatc ggggtaaagg agcctggaac acatcctgtg accccgcctg tcctgtagga agccagtctc
 917  tgctgggatc tgagcttctt cgtgaaacac ggggcagagg aggcaccaga actCTCCTCT GACCAACTGC CCCACAGCAC acatatcctc aaaggatagt cttgaatacg tgatggaaga attaaagaga ---------- ---------- ---------- ---------- ---------- ---------- ----------
                                                                ******* ********** **********

 467  tggaaagtaa aatggaaggg ctgcttGGGA ACTTTGAGGA TATTTAGCCC ACCCcctcat tTTTACTTGG GGAAACTaag gcccaGAGAC CTAAGGTGAC TGCCtaagtt agcaaggAGA AGTCTTGGGT ATTCATCCCA GGTTGGGGGG ACc---CAAT TATTTCTCAA TCCCATTGTA TTCTGGAATg ggcaATTTGT
1047  ---------- ---------- ------GTGA GGTCTGAAGA AAATCAGCCC TCTCgggg-- -TTTCCTTTG GGTAACT--- -----GAGTG CTAAGGTGAC TTCCgagtca gca----AGA AATATCGGAC GTTCAACCCA GGTTGAGTGG AGgaaaCAAT TATTTCTCAA TCCTAATATG TTCTGGAATa gcccATTTAT
                                  **** ********** ********** ****        ********* *******         ***** ********** ****              *** ********** ********** ********** **    **** ********** ********** *********      ******
                                                                                                                                      *** ********** ********** ********** **    **** ********** ********** *********      ******
                                                                                                                                                                                 **** ********** ********** *********      ******
                                                                                                                                                                                 **** ********** ********** *********      ******
                                                                                                                                                                                 **** ********** ********** *********

 664  CCACGTCACT GTGACCTAGG AACACGCGAA TGaGAACCCA CAGCTGAGGG CCTCTGCGCA CAGAACAGCT GTtctCCCCA GGAAATCAAC TTTTTTTAAT TGAGAAGCTA AAAAattatt ct------AA GAGAGGTAGC CCATCCTAAA AATAGCTGTA ATGCAGAAGT TCATGTTCAA CCAATCATTT TTGCTTACGA
1206  CCACGTCATT ATGACCTGGG AGTGCGTGAA TG-GAATCCA CAGATGAGGG CCTCTGTACA TAGAACAGCT GTctgCCTCA GGAAATACAA CTTTTAGTAT TGAGAAGCTA AAAAgaaaaa aaaattaaAA GAGAGGTAGC CCATACTAAA AATAGCTGTA ATGCAGAAGT TCATTCCGAC CAGTTCTTTA GCGCTTACAA

      ********** ********** ********** ** ******* ********** ********** ********** **   ***** ********** ********** ********** ****               ** ********** ********** ********** ********** ********** ********** **********
      ********** ********** ********** ** ******* ********** ********** ********** **   ***** ********** ********** ********** ****               ** ********** ********** ********** ********** ********** ********** **********
      ********** ********** ********** ** ******* ********** ********** ********** **   ***** ********** ********** ********** ****               ** ********** ********** ********** ********
      ********** ********** ********** ** ******* ********** ********** ********** **   ***** ********** ********** ********** ****               ** ********** ********** ********** ********
                                          ******* ********** ********** ********** **   ***** ********** **********                               ** ********** ********** ********** ********
                                          ******* ********** ********** ********** **                                                             ** ********** ********** ********** ********
                                                             ********** ********** **                                                             ** ********** ********** ********** ********
                                                                                                                                                  ** ********** ********** ********** ********
                                                                                                                                                  ** ********** ********** ********** ********

 858  T                  ag tttattagag aggttaga-- ---------- ---------- ---------G AAGGAGGAGC TCTAAGGAGA AAAAatC                       ggctttttaa tgaatgaagA GGCCTCCCTG AGCTTACAAT ATAAAAGGGG GACAGAGAgg tgaAGGTCTA Cacatcaggg
1405  TGC                   aaaaaaaaaa gaaagaaatt aaactcaaaa attgcatggt ttagaagagG GAGGAGGAGC CTGAATAACA AAA                  -- ---------- ---------- --------       CACTG AGCCTTCAGT ATAAAAGGGG GACCAAGAac aggAGGTCTA C---------
      ********                                                                   * ********** ********** ****  **** ********                                  * ********** ********** ********** ********      ******* *
      *******                                                                                                                                                 * ********** ********** ********** ********
                                                                                                                                                              * ********** ********** ********** ********
                                                                                                                                                              * ********** ********** ********** ********
                                                                                                                                                              * ********** ********** ********** ********

1027  gcttgctctt gcaaaaccaa accacaagac agacttgcaa aagaaggc
1563  ---------- ---------- ---------- ---------- --------
```

Task 5

What you will be doing next

- General aim: Functional analysis of the IL-10 promoter
- Find and extract IL-10 promoters (***ElDorado***)
- Analyze IL-10 promoters for transcription factor binding sites (TFBS) (***MatInspector*** from ***GEMS*** package)
- Analyze literature for IL-10 Promoter (***BiblioSphere***)
- Analyze IL-10 promoters for TFBS frameworks (***FrameWorker*** from ***GEMS*** package)
- Database search with frameworks (***ModelInspector*** from ***GEMS*** package)

IL-10 promoter frameworks

Database search with the IL-10 IRF-STAT framework

- 221 matches located in the Eukaryotic Promoter Database (EPD, 4.6%)
- Among those were:

- IFNA13
- IFNA6
- IFNA5
- IFNA7

- IFNA16
- IL-2
- IL1B
- IL2RG

- MHCI
- HLA-D
- HLA-E

- and several other Interferon target genes

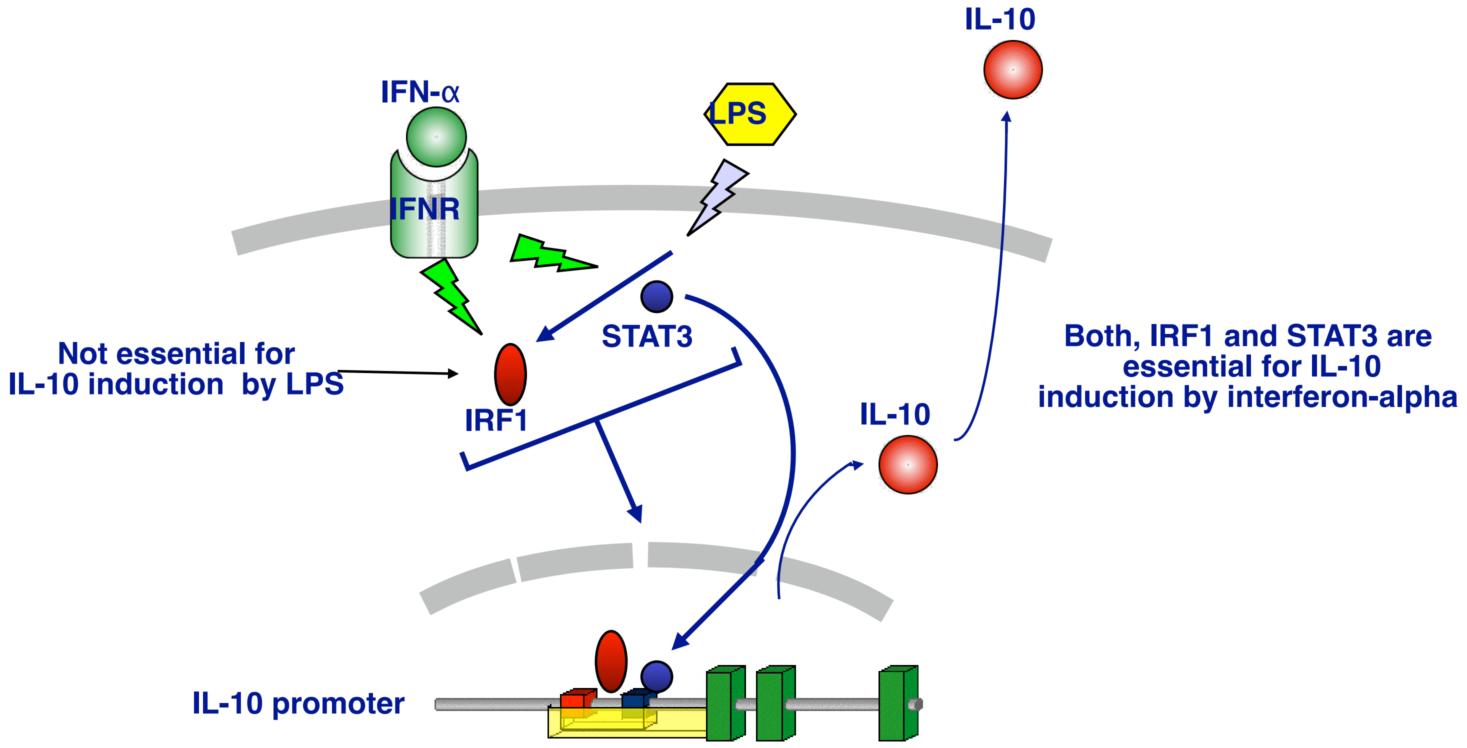
- All together 11 of 221 matches were directly IFN/IRF/STAT related

The IL-10 IRF-STAT module identified other potential IFN-alpha targets

Genomatix

IL-10 promoter summary

The IL-10 promoter as signal processor



Not essential for
IL-10 induction by LPS

Both, IRF1 and STAT3 are
essential for IL-10
induction by interferon-alpha

Molecular basis for two pathways converging on the IL-10 promoter elucidated

Genomatix

Acknowledgements

The IL-10 promoter analysis was an international cooperation

Institut für Immunologie
University of Munich, Germany

Elke M. Benkhart
Maciej Siedlar
Angela Wedel
Mark Lötzerich

Division of Immunology
University of Leicester, UK

Marion Frankenberger
Löms Ziegler-Heitbrock

GSF-National Research Center for
Environment and Health
Gauting, Germany

Annette Schäfer

...and *Genomatix*

...now it is your turn!