Comparative Genomics
Gene Regulatory Networks (GRNs)

Anil Jegga
Biomedical Informatics

Contact Information:
Anil Jegga
Biomedical Informatics
Room # 232, S Building 10th Floor, CCHMC
Homepage: http://anil.cchmc.org
Tel: 513-636-0261
E-mail: anil.jegga@cchmc.org

Session 1: Overview of GRNs (Feb 23)

a. Computational Approaches
b. Cis-Element Identification
c. Comparative Genomics
d. Regulatory region variations
e. p53 case study

Session 2: Database Session (Feb 24)

a. Genome Browsers
b. Promoter Analysis, TFBS Search
c. Co-regulated gene analysis

Some Basic Questions….

- **What is Gene Expression?**

  It is the ability of a gene to produce a biologically active protein.

- **What is a Promoter?**

  Combination of short sequence elements to which RNA polymerase binds in order to Initiate transcription.

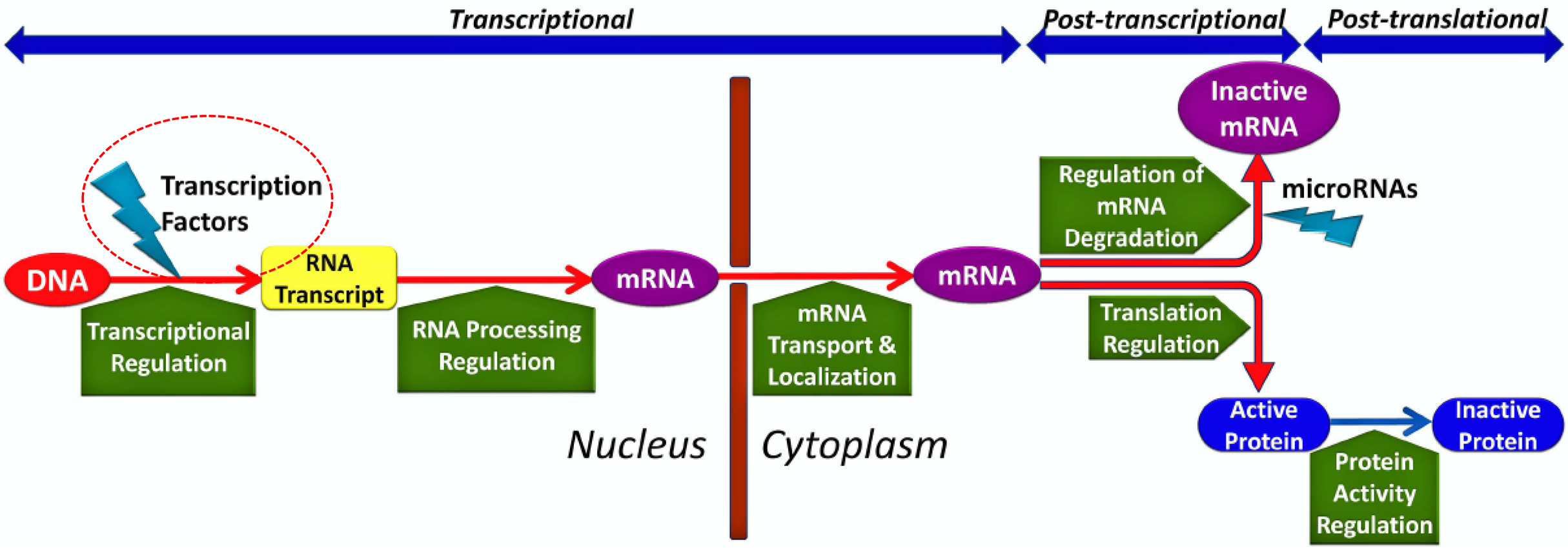
- **What is an Enhancer?**

  Set of short sequence elements which Stimulate transcription. Function independent of position or orientation.

- **What is a Silencer/Repressor?**

  Set of short sequence elements which Suppress transcription.

Different levels of gene regulation



Points to Remember….

- *Cis*-regulators are on the same chromosome as the gene.
- Trans-regulators are transcription factors that act on many different genes, on different chromosomes.
- Transcription factors are proteins with DNA-binding domains, which bind to specific DNA sequences.
- Transcription factors usually do not work alone.

Points to Remember….

- Elementary Units of Transcriptional Regulatory Regions are Transcription Factor Binding Sites (TFBSs) or *cis*-elements.
- Control Regions
  - Promoters
  - Enhancers
  - Silencers
  - Insulators
- Regulatory output of control region depends on specific combination of its elements + order and orientation.
- Genes are controlled by several control regions located upstream or downstream.

In summary….

- TFBSs or *cis*-elements are the fundamental units of gene regulation.
- TFBSs are small nucleotide fragments (usually ≤ 30 bp).
- Regulatory proteins, namely the Transcription Factors (TFs), bind in a sequence-specific manner to TFBSs to activate or suppress gene transcription (gene expression).
- Regulatory regions can occur anywhere including the coding regions and they can be several thousand base pairs away from the target gene.
- TFBSs are a critical component in gene regulation and that identification of TFBSs is a central problem for understanding gene regulation in molecular biology and genetics.

Computational Approaches

TFBS Discovery – Computational Search….

Two Problems:

1. Searching for additional members of a known DNA binding motif (site search problem) *or* finding novel target genes of a TF or TFs and
2. Discovering novel DNA binding motifs in collections of functionally related sequences (sequence motif discovery problem)

TFBS Discovery – Computational Search….

- Several methods available to search for binding sites – ranging from pattern search (e.g. scanning the genome for all occurrences of a specific short sequence or a known TFBS) to using PWMs.
- Several algorithms available for sequence motif discovery. Hypothesis: a set of sequences share a binding motif for functional reasons.
- Recent advances in sequencing have led to the introduction of comparative genomics approaches to DNA binding motif discovery – Phylogenetic Footprinting

TFBS Discovery – Pattern Search - 1

Seed or Probe Sequence

TFBS-1:
CGCCATATAAGGAGCAGGAA

TFBS-2:
TGGAGTGGCCC

Target Sequence: Specific promoters or genome-wide scan

```
>ATG4C range=chr1:63021391-63022390
Cctcccaaaatgctgggattacaggcgtcagccgccgcgcc
tggccaaaaatttccacgttataaaaagcatcttcggaccg
tccttactgctttaggtgacttttcttttccttctttttt
ttttttttccccacaaagcctcgctctaaagtgatctgtg
ttgtgagaatataggagaaatggagtttggtttccc
>MAP1LC3C range=chr1:240229009-240230008
gtatggggaacccccacccacccacacacacacacacacat
ctagtgtcaaaagtgatctgtgttgtgagaatataggagaa
atggagtttggtttcccctatgtatgaccagagacccagtt
tcttaactctaaaagcaagatcttggg
```

Parameters: Specify mismatches, context…

Too many false positives!

TFBS Discovery – Pattern Search - 2

```
>TFBS-1
CGCCATATAAGGAGCAGGAA
>TFBS-2
CGCCTTATATGGAGTGGCCC
>TFBS-3
GACCAAATAAGGCAAGGTGG
>TFBS-4
TACCAAATAAGGGCAGGCTG
>TFBS-5
AGCCATATGTGGACAGATGG
>TFBS-6
CGCCTTCTTTGGGCAGCGCG
>TFBS-7
ACCCAAATATGGCGACGGCC
>TFBS-8
GTCCTTATATGGACTCATCT
>TFBS-9
ATCCTTTTATGGCCCTGTCC
>TFBS-10
ACCCAAATATGGAAATATTG
>TFBS-11
GCCCATATTTGGCGATCTTC
>TFBS-12
GCCCATATTTGGCGATCTTC
>TFBS-13
ATCCCTATTTGGCCATCCCT
>TFBS-14
CTCCCTATTTGGCCATCCCC
>TFBS-15
TTCCTTACATGGTCTGGGGG
>TFBS-16
TTCCATACATGGGCTAAGGG
>TFBS-17
GTCCATATTAGGACATCTGC
>TFBS-18
GTCCATATTAGGACATCTGC
>TFBS-19
GTCCATATATGGGCAGCGAC
>TFBS-20
TCCCATATATGGCCATGTAC
```

IUPAC

Code	Description
A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil
R	Purine (A or G)
Y	Pyrimidine (C, T, or U)
M	C or A
K	T, U, or G
W	T, U, or A
S	C or G
B	C, T, U, or G (not A)
D	A, T, U, or G (not C)
H	A, T, U, or C (not G)
V	A, C, or G (not T, not U)
N	Any base (A, C, G, T, or U)

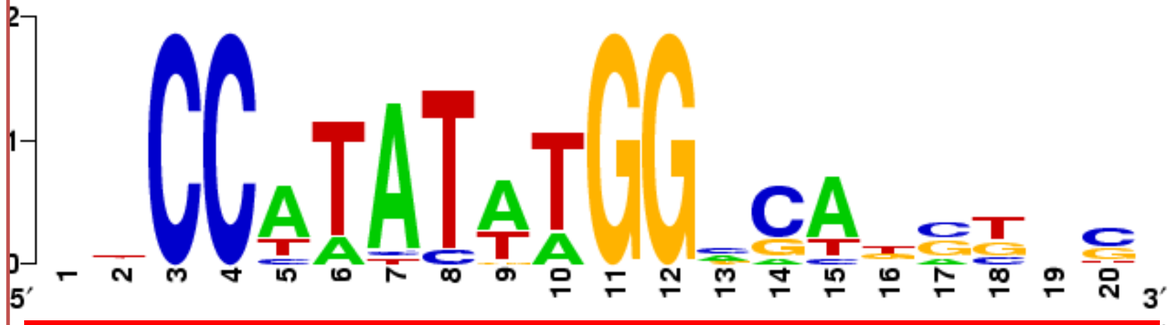
TFBS Discovery – Pattern Search - 2

```
>TFBS-1
CGCCATATAAGGAGCAGGAA
>TFBS-2
CGCCTTATATGGAGTGGCCC
>TFBS-3
GACCAAATAAGGCAAGGTGG
>TFBS-4
TACCAAATAAGGGCAGGCTG
>TFBS-5
AGCCATATGTGGACAGATGG
>TFBS-6
CGCCTTCTTTGGGCAGCGCG
>TFBS-7
ACCCAAATATGGCGACGGCC
>TFBS-8
GTCCTTATATGGACTCATCT
>TFBS-9
ATCCTTTTATGGCCCTGTCC
>TFBS-10
ACCCAAATATGGAAATATTG
>TFBS-11
GCCCATATTTGGCGATCTTC
>TFBS-12
GCCCATATTTGGCGATCTTC
>TFBS-13
ATCCCTATTTGGCCATCCCT
>TFBS-14
CTCCCTATTTGGCCATCCCC
>TFBS-15
TTCCTTACATGGTCTGGGGG
>TFBS-16
TTCCATACATGGGCTAAGGG
>TFBS-17
GTCCATATTAGGACATCTGC
>TFBS-18
GTCCATATTAGGACATCTGC
>TFBS-19
GTCCATATATGGGCAGCGAC
>TFBS-20
TCCCATATATGGCCATGTAC
```

Alignment

```
CGCCATATAAGGAGCAGGAA
CGCCTTATATGGAGTGGCCC
TTCCTTACATGGTCTGGGGG
TTCCATACATGGGCTAAGGG
GTCCTTATATGGACTCATCT
TACCAAATAAGGGCAGGCTG
GACCAAATAAGGCAAGGTGG
CGCCTTCTTTGGGCAGCGCG
AGCCATATGTGGACAGATGG
ACCCAAATATGGAAATATTG
ATCCCTATTTGGCCATCCCT
CTCCCTATTTGGCCATCCCC
GTCCATATATGGGCAGCGAC
ACCCAAATATGGCGACGGCC
ATCCTTTTATGGCCCTGTCC
GTCCATATTAGGACATCTGC
GTCCATATTAGGACATCTGC
GCCCATATTTGGCGATCTTC
GCCCATATTTGGCGATCTTC
TCCCATATATGGCCATGTAC
```

Use an IUPAC consensus instead of a single specific pattern



nnCCATatwtGGncaksknsn

- **False positives**
- **Need to frame specific additional rules (e.g. at which positions mismatches are allowed?)**

TFBS Discovery – Position Weight Matrix (PWM)

- TFBS are often modeled by position weight matrices (or position specific scoring matrices), which is a probabilistic model that characterizes the DNA binding preferences of a TF.
- Since a PWM is built from a collection of aligned DNA binding sites that are likely to be bound by a common TF, they offer a sensitive way to represent the specificity of transcription factor/DNA interfaces.

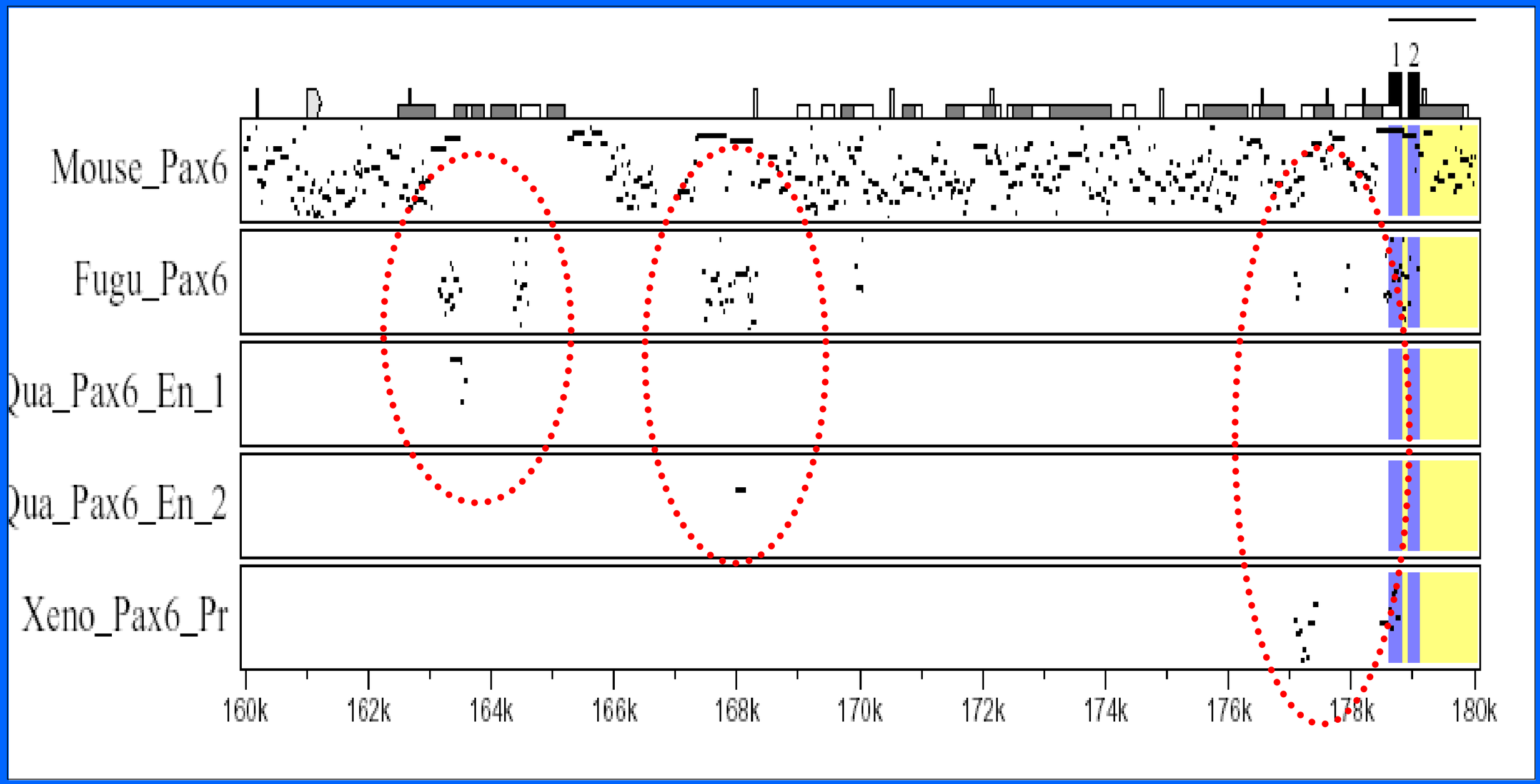
Pos.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	5	2	0	0	13	4	**18**	0	12	5	0	0	7	2	14	2	4	0	3	1
C	4	5	**20**	**20**	2	0	1	2	0	0	0	0	8	13	2	2	8	4	7	10
G	7	4	0	0	0	0	0	0	1	0	**20**	**20**	4	5	0	7	8	6	6	7
T	4	9	0	0	5	**16**	1	**18**	7	15	0	0	1	0	4	9	0	10	4	2
IUPAC	N	N	C	C	A	T	A	T	W	T	G	G	N	C	A	K	S	K	N	S

False positives: On average, for every 1000 bp, 80-100 TFBS!

What is functional and what is not?

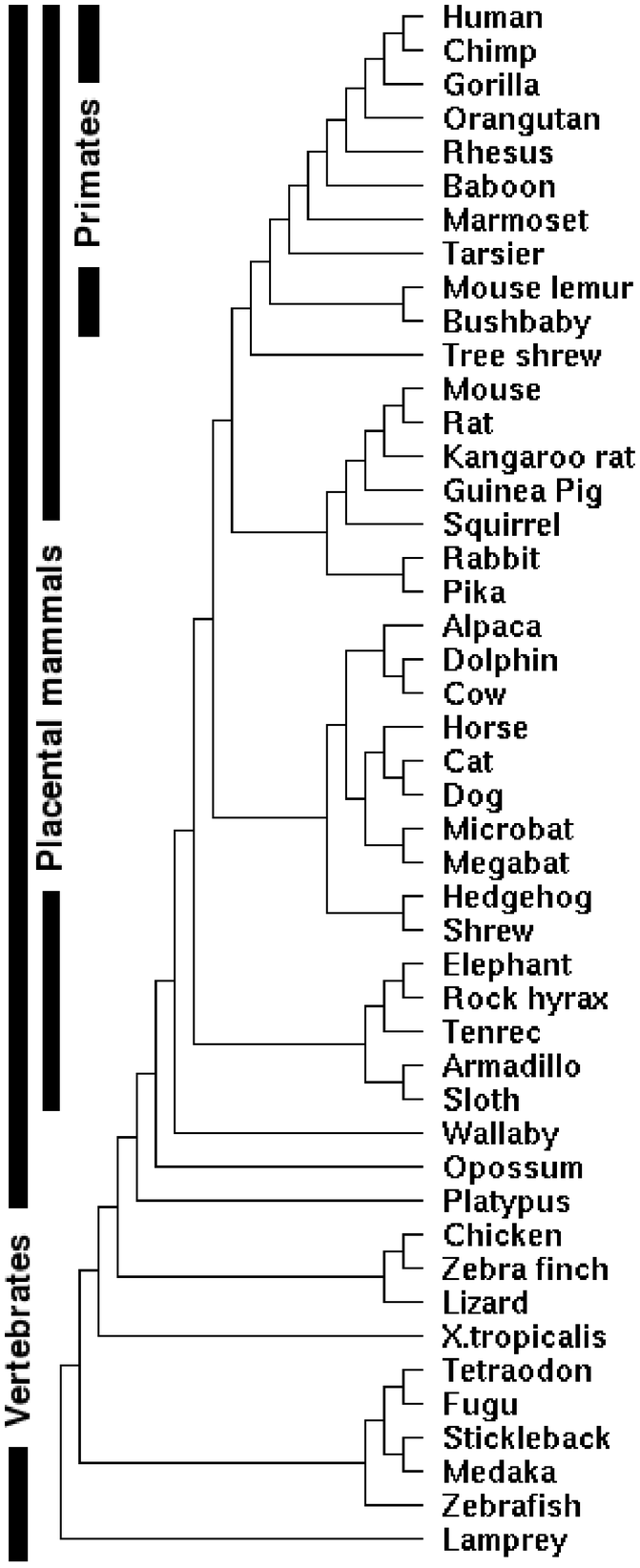
GRNs – Comparative Genomics: Motivation

Comparative Analysis of Homologous Sequences: After >900 Myrs of divergence essential functional elements remain conserved! E.g. PAX6: Human Vs Puffer fish (*Fugu rubripes)*



http://bio.cse.psu.edu/Multipipmaker

46 Species



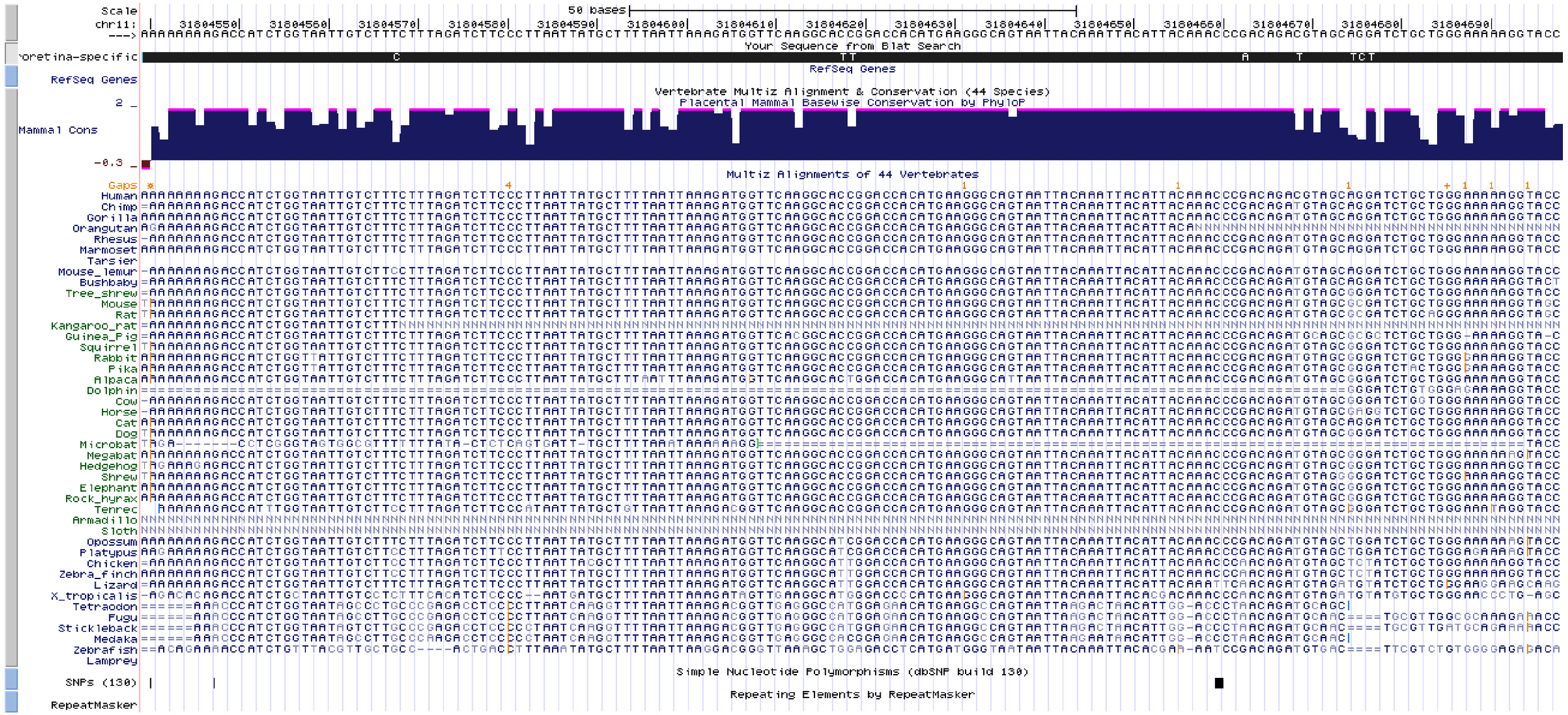

```
>Quail_Pax6_neuroretina-specific gene
enhancer region
GGTACCTTTTTCCCAGCAGATAGAGCTACATCTGTTGGGT
TTGTAATGTAATTTGTAATTACTGCCCTTCATGTGGTCCA
ATGCCTTGAACCATCTTTAATTAAAAGCATAATTAAGGGA
AGATCTAAAGGAAGACAATTACCAGATGGTCTTTTTTTTTT
TTTTTTTTTAGAAGCGGTTCGTTGCTCGGAGGGCGCAGCC
CGGTCCGCTTCGGACTCGGCTTAAGGGCCGGAGGGGTCGG
AGAGGGAGGGGGGGGGGGGTCCGAGCCAGGGCTCGGG
```

Align to human genome

159/277 bp align with human genome at 95.0% sequence similarity!

Human: chr11:31,847,964-31,848,122

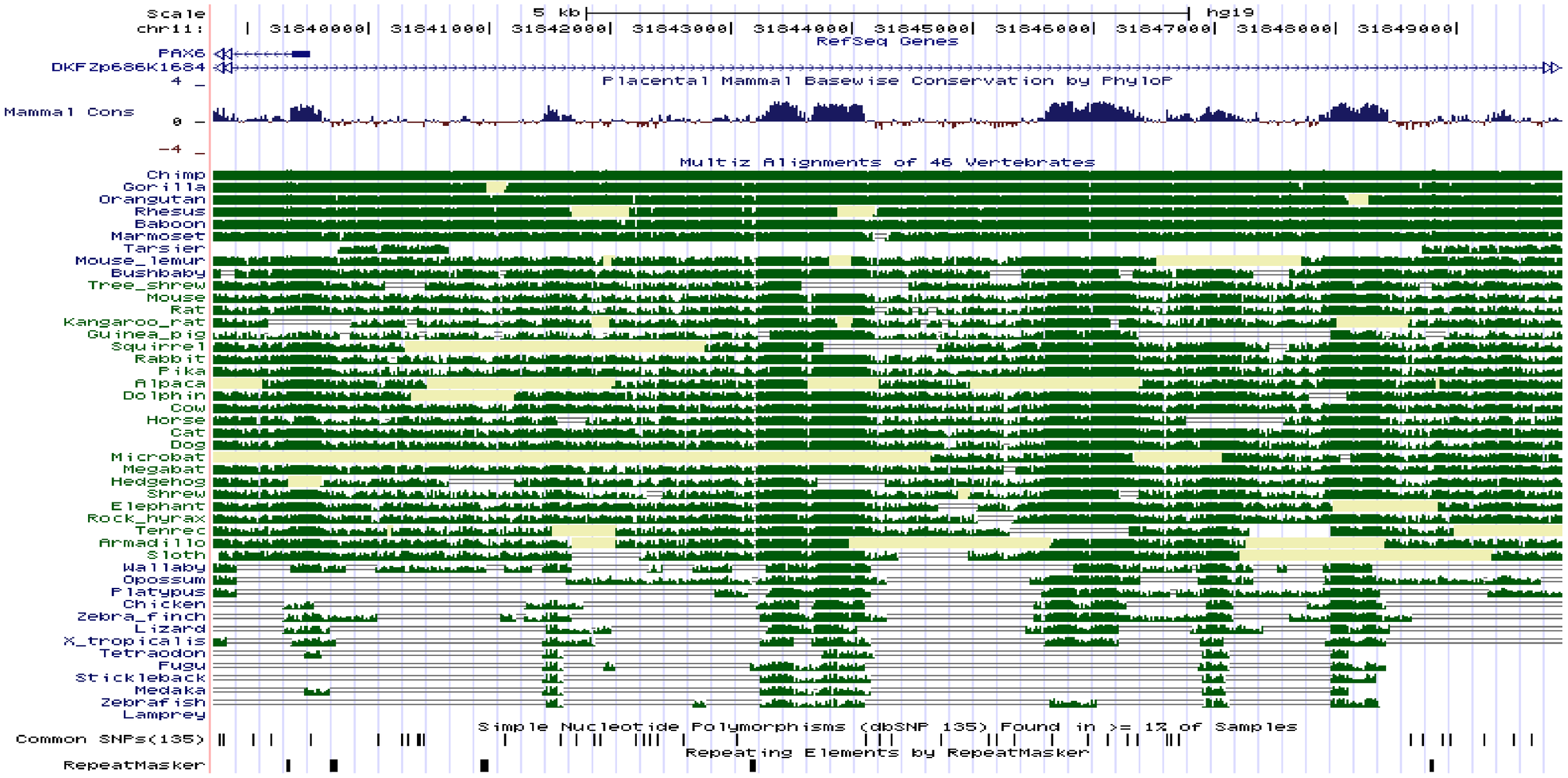
159 bp: Highly conserved in almost all of the species



37/46 Species

```
Alignment block 4 of 6 in window, 31804609 - 31804674, 66 bps
         Human  ttcaaggcaccggaccacatgaa-gggcagtaattacaaattacatta-caaacccgacagacgtagc
         Chimp  ..........................................................t.....
       Gorilla  ..........................................................t.....
     Orangutan  ..................................................----------------
        Rhesus  ..........................................................t.....
      Marmoset  ..........................................................t.....
   Mouse lemur  ..........................................................t.....
     Bushbaby  ..........................................................t.....
   Tree shrew  ..........................................................t.....
        Mouse  ..........................................................t.....
          Rat  ..........................................................t.....
   Guinea Pig  ....c.......................................................t.c...
     Squirrel  ..........................................................t.....
       Rabbit  ..........................................................t.....
         Pika  ..........................................................t.....
       Alpaca  .........t.................t................................t.....
          Cow  ..........................................................t.....
        Horse  ..........................................................t.....
          Cat  ..........................................................t.....
          Dog  ..........................................................t.....
      Megabat  ..........................................................t.....
     Hedgehog  ..........................................................t.....
        Shrew  ..........................................................t...g
     Elephant  ..........................................................t.....
   Rock hyrax  ..........................................................t.....
       Tenrec  ..........................................................t.....
      Opossum  .........t................................................t.....
     Platypus  .........t................................................t.....
      Chicken  .........tt.....................................a.....t.....
  Zebra finch  .........tt.....................................a.....t.....
       Lizard  .........tt.................................tt.a.....t....a
X. tropicalis  ..g......tg.....c......g.....................cg.........a.....t....a
     Tetraodon  ..g.g...cat...ga.........c.........ag.c.a.....g.g-.c..ta.....t.c...
         Fugu  ..g.g...cat...ga.........c.........ag.c.a.....g.g-.c..ta.....t.c.a.
   Stickleback  ..g.g...cat...ga.........c.........ag.c.a.....g.g-.c..ta.....t.c.a.
        Medaka  ..g.g...ca....ga.........c.........ag...a.....g.g-.c..ta.....t.c.a.
    Zebrafish  g.t..a..tgga....t.....t....t.a..............cg.ca-..t........t..ga.
```

Human: chr11:31,838,708-31,849,878
Pax6 upstream region with several blocks of HCNRs



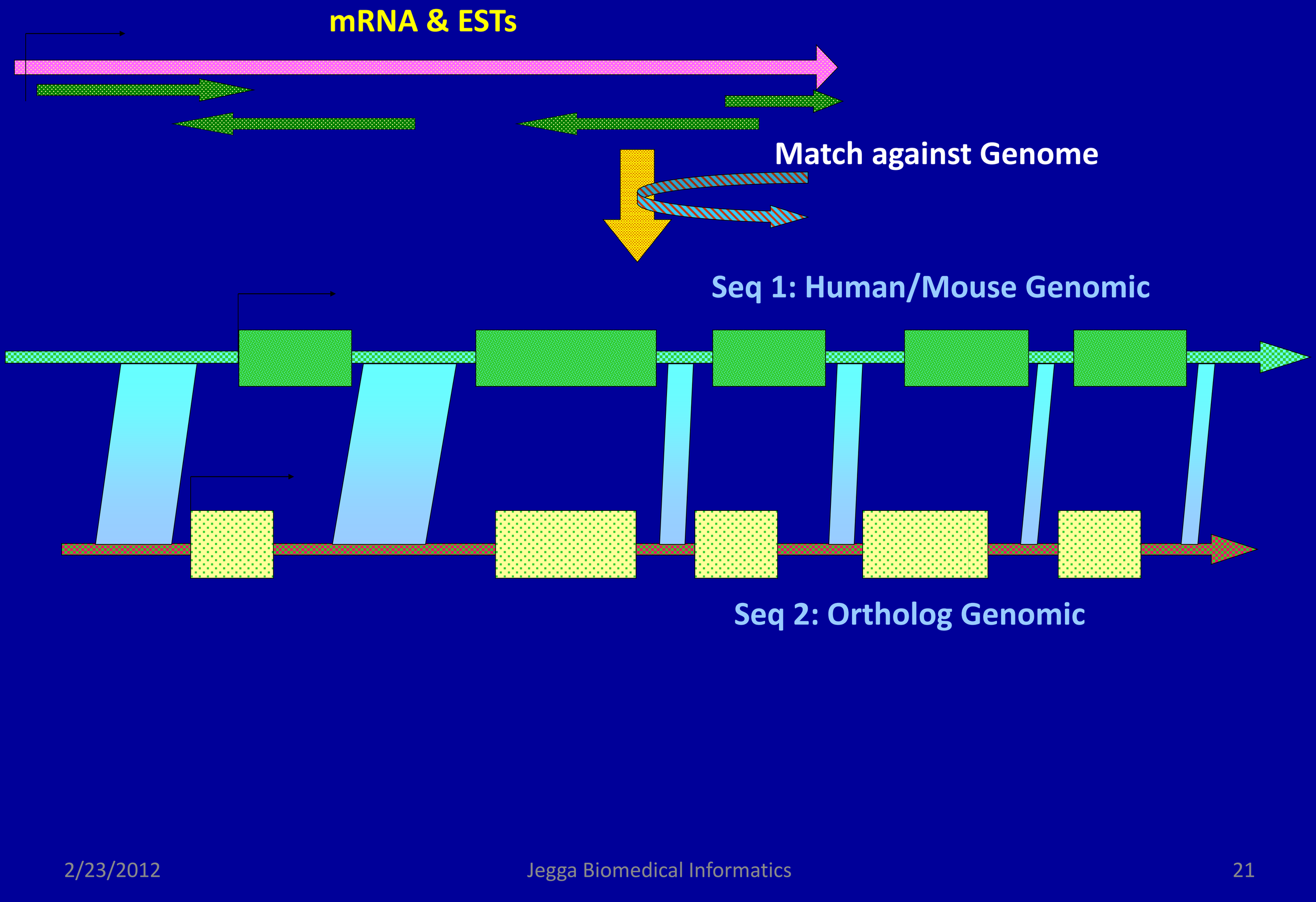
Phylogenetic Footprinting: is a technique used to identify TFBSs within a non-coding region through comparative genomics of distantly-related (human Vs mouse or chicken) species. When this technique is used with a closely-related species (human Vs chimp), this is called **Phylogenetic Shadowing**.

Ultra Conserved Region (UCRs

- There are **481** segments longer than **200** base pairs that are absolutely conserved (**100%** identity with no insertions or deletions) between orthologous regions of the human, rat, and mouse genomes. Almost all of these UCRs are also conserved in the chicken and dog genomes, with an average of 95-99% identity, respectively. Several of these are significantly conserved in fish too.
- These ultra conserved elements of the human genome are most often located either overlapping exons in genes involved in RNA processing or in introns or nearby genes involved in the regulation of transcription and development.
- Interestingly, these regions have little variation (20-fold fewer SNPs).

Bejerano et al., 2004. Science 304(5675): 1321-5

Cycle of Events in Identifying Potential Regulatory Regions



>Seq 1 Genomic

AGAGAAAATTGCTAGAGC
TCAGGAGTTTGAGACCAG
CCTGGGCAATAGAGTAAG
ACTTTGTCTCTATCAAAAA
TTTAAAAATTAACTGGGCT
TGGCGGTGTGCACCTGT
GGTCCAGCTACTCAGGAG
GCTGAGGTGGGAGGATT
GCTTGAGCCCAAGAGGTT
GAGGCTGCAGTAAGCCGT

TF Binding Sites

V$ETSF/ETS1_B	8333 - 8347
V$STAT/STAT1_01	8335 - 8355
V$ETSF/PU1_B	8335 - 8350
V$ETSF/GABP_B	8336 - 8347
V$ETSF/NRF2_01	8338 - 8347
V$CLOX/CDPCR3_01	8363 - 8377
V$EVI1/EVI1_01	8373 - 8388

BlastZ

Local Alignment Number 5

Similarity Score: 3074
Match Percentage: **51 %**
Number of Matches: 96
Number of Mismatches: 39
Total Length of Gaps: 52
Begins at (8281,8874) and
Ends at (8416,9059)

Seq 1	<--> Seq 2	Sim%	No. of Nt
8281-8300	<--> 8874-8893	**70%**	(20 nt)
8301-8310	<--> 8902-8911	**90%**	(10 nt)
8311-8324	<--> 8923-8936	**57%**	(14 nt)
8325-8376	<--> 8947-8998	**62%**	(52 nt)
8378-8386	<--> 8999-9007	**67%**	(9 nt)
8387-8416	<--> 9030-9059	**90%**	(30 nt)

Trafac

Seq 1	<--> Seq 2	Sim% Nt	Hits
8301-8310	<--> 8902-8911	90% (10 nt)	**3**
8311-8324	<--> 8923-8936	57% (14 nt)	**2**
8325-8376	<--> 8947-8998	62% (52 nt)	**3**
8378-8386	<--> 8999-9007	67% (9 nt)	**0**
8387-8416	<--> 9030-9059	90% (30 nt)	**4**

>Seq 2 Genomic

GACTGAGGGCTTGTGAAA
CAGCAAGAACCTGTCTCA
AAAAACAGTGGGCAGGGA
GGGGATTAATGAATAGGCA
GCTACGTTCTGGGACTGG
AGGGACTCGAGGTGGCTA
GAAAGCAAGAGGTACTGG
GAGACAAGGCTGCAGACA
TTTCTTTTTTTTTTTTTTTT
TTTGAGACAGAGTC

TF Binding Sites

V$ETSF/ETS1_B	8880-8894
V$STAT/STAT1_01	8881-8901
V$ETSF/PU1_B	8882-8897
V$ETSF/NRF2_01	8892-8902
V$CLOX/CDPCR3_01	8908-8922
V$GATA/GATA_C	8916-8928
V$FKHD/FREAC2_01	8923-8938

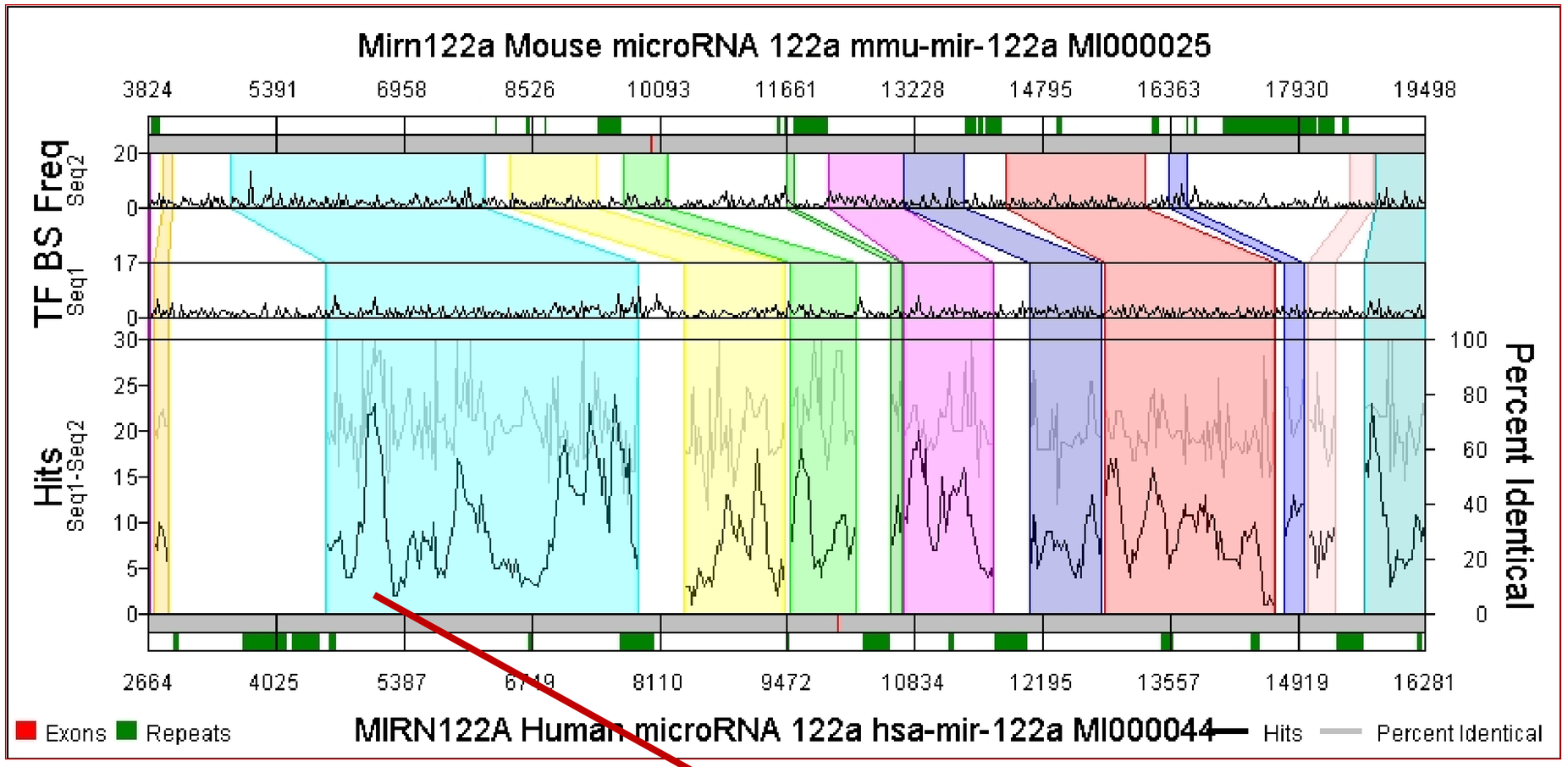
Sequence Conservation – Reduction in search space

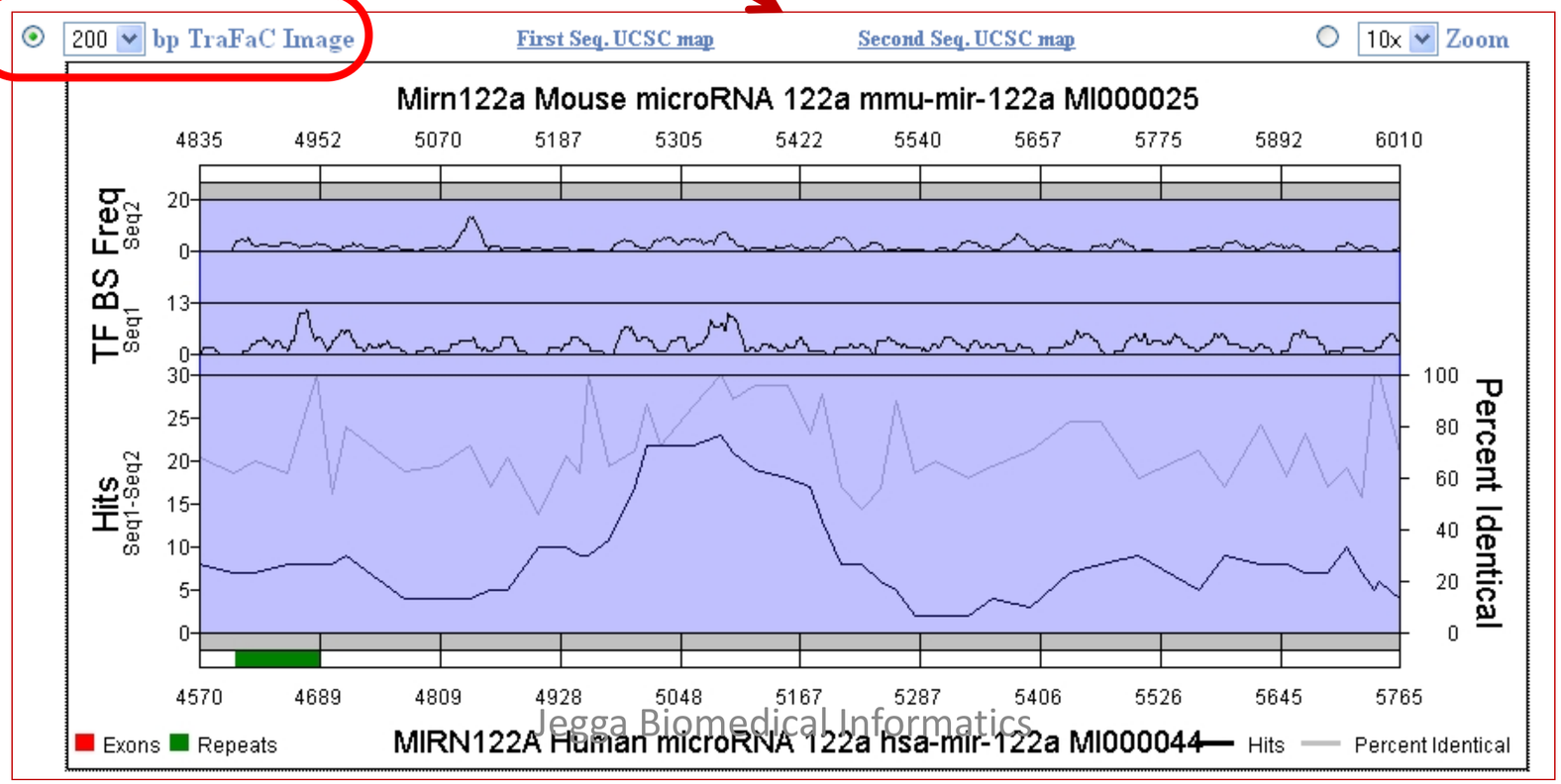
Gene	Genomic Sequences		Total Sequence length (bp) (a)	Sum of Conserved Regions (bp) (b)	% Sequence to be searched for shared TF binding sites. (b/a)
	Human (bp)	Mouse (bp)			
ADA	36741	29807	66548	16668	25
APEX	22527	21963	44490	5982	13
XRCC1	37785	37349	75134	11260	15
ERCC2	54336	32595	86931	14257	16
CD4	39512	43508	83020	18828	23
PAX6	378625	400000	778625	187432	24
ATM	162429	116461	268890	58828	21
MYO7A	106974	75825	182799	50247	27

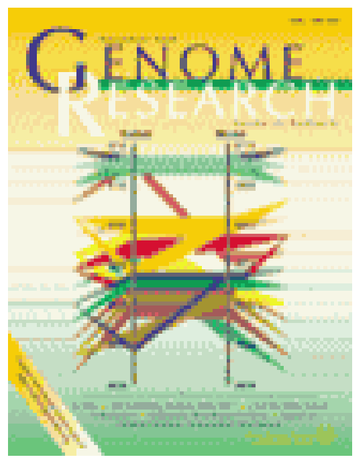
Sequence Conservation – Filtering the TFBSs

Gene	Number of TF binding sites in genomic sequences		Total number of TF binding sites of the two sequences (a)	TF binding sites in common within blastz aligned regions (b)	% Reduction based on # shared cis-elements within blastz-aligned sequence blocks (100 - b/a%)
	Human	Mouse			
ADA	2225	1740	3965	798	80
APEX	1850	1721	3571	458	87
XRCC1	2276	2209	4485	469	90
ERCC2	1806	1051	2857	299	90
CD4	2609	2609	5218	910	83
PAX6	29398	27321	56719	12444	78
ATM	12653	8753	21406	4093	81
MYO7A	7492	5146	12638	3116	75

GenomeTrafac: http://genometrafac.cchmc.org

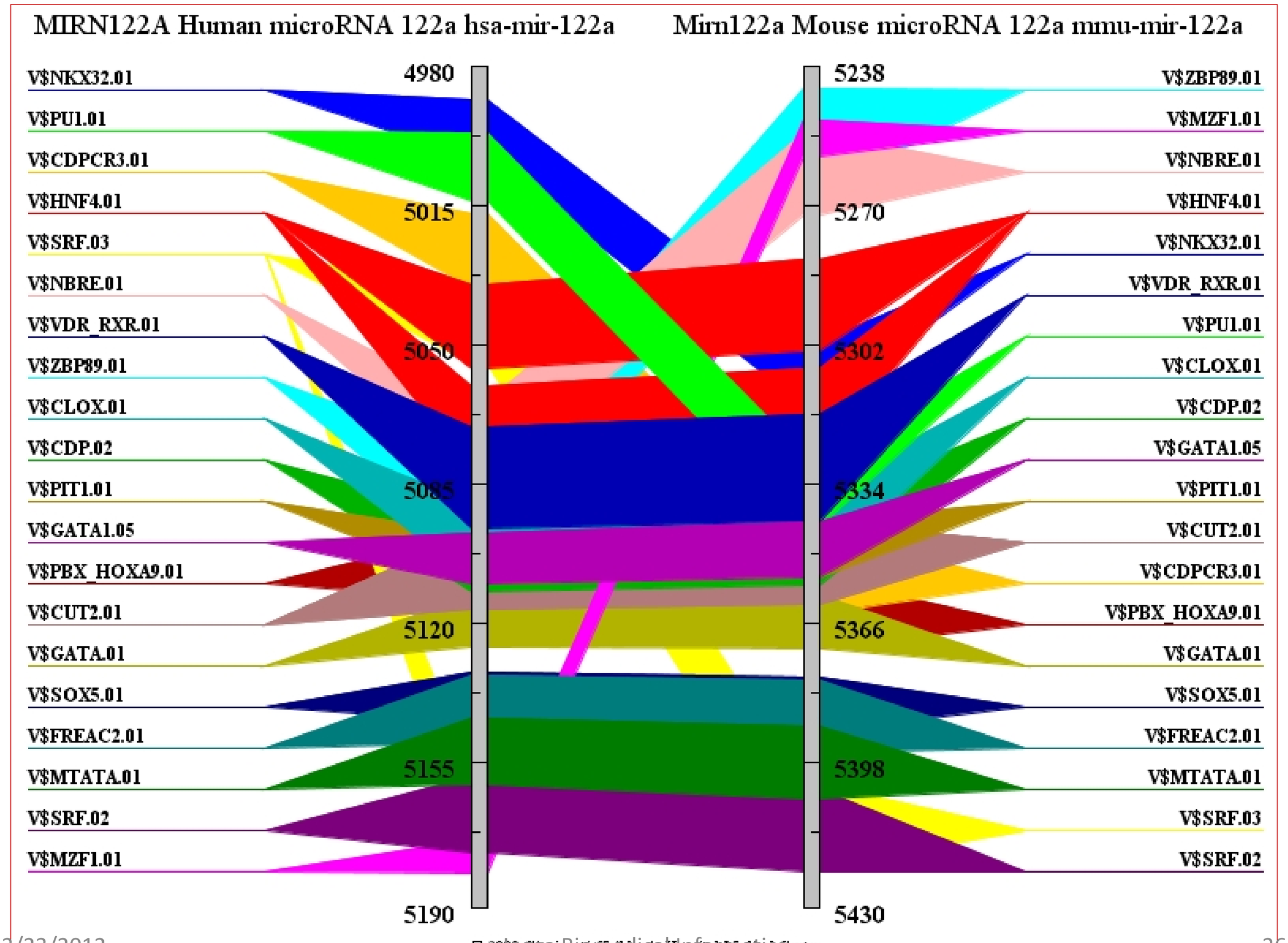






Jegga et al., 2002
Genome Research
12 (9): 1408-1417

GenomeTrafac: http://genometrafac.cchmc.org



GenomeTrafac: http://genometrafac.cchmc.org

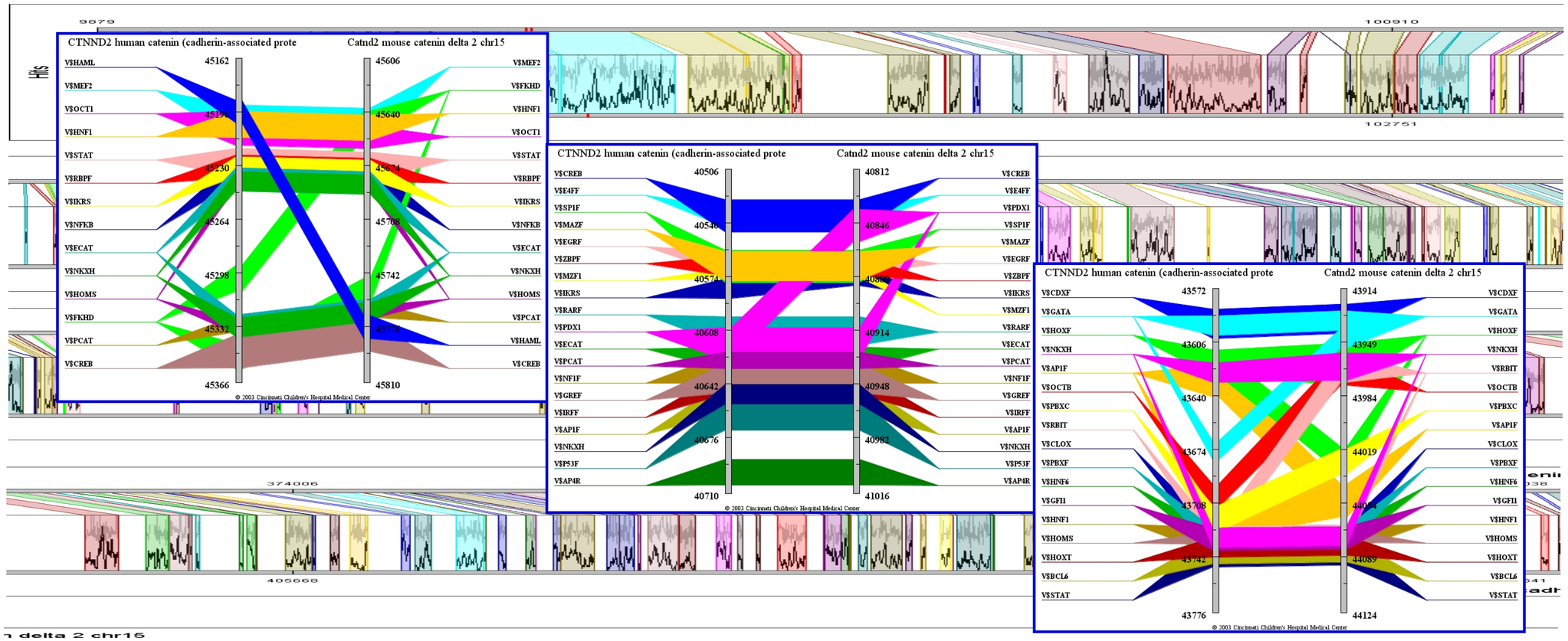
Shared *Cis*-elements

(Genomatix Matrix Family Library Version 5.0 (January 2005))
(For details and annotations of TFBS-PWMs, please register at Genomatix)

Family/Matrix	Description	hgMIRN122A				mgMirn122a			
		Begin	End	Sequence		Begin	End	Sequence	
V$NKXH/V$NKX32.01	Homeodomain protein NKX3.2 (BAPX1, NKX3B, Bagpipe homolog)	4993	5007	CCCCCACTCAGCAGA	-	5301	5315	CTGACTTAGTGGACT	+
V$ETSF/V$PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	5001	5017	CAGCAGAGGAATGGACT	+	5326	5342	CCTCTCTTCCCCCACAA	-
V$CLOX/V$CDPCR3.01	Cut-like homeodomain protein	5020	5038	CCAATCTTGCTGAGTGTGT	-	5343	5361	TCGATAATTTAATGTGACT	-
V$HNF4/V$HNF4.01	Hepatic nuclear factor 4	5037	5057	GTTTGACCAAAGGTGGTGCTG	+	5283	5303	GTTTGACCAAAGGTGACTCTG	+
V$SRFF/V$SRF.03	Serum responsive factor	5038	5056	TTTGACCAAAGGTGGTGCT	-	5399	5417	GGATCCCATAAAGGGAGAG	-
V$HNF4/V$HNF4.01	Hepatic nuclear factor 4	5061	5081	TAGTGGCCTAAGGTCGTGCCC	+	5307	5327	TAGTGGACTAAGGTCATGCCC	+
V$RORA/V$NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)	5065	5083	GGCCTAAGGTCGTGCCCTC	+	5255	5273	GGGAGCTGGACCTTCGGTT	-
V$RXRF/V$VDR_RXR.01	VDR/RXR Vitamin D receptor RXR heterodimer site	5071	5095	AGGTCGTGCCCTCCCTCCCCCACTG	-	5317	5341	AGGTCATGCCCTCTCTTCCCCCACA	-
V$ZBPF/V$ZBP89.01	Zinc finger transcription factor ZBP-89	5077	5099	TGCCCTCCCTCCCCCACTGAATC	+	5245	5267	GGGGCATGGGGGGAGCTGGACCT	-
V$CLOX/V$CLOX.01	Clox	5089	5107	CCCACTGAATCTGATAAAA	+	5334	5352	CCCCCACAATCGATAATTT	+

Regulomorph (Regulatory Polymorphisms)

CTNND2 catenin (cadherin-associated protein), delta 2) 1225 AA, 4746 bp mRNA



Genomic Sequence: >1 million bp

1st Intron: >192 kb!!

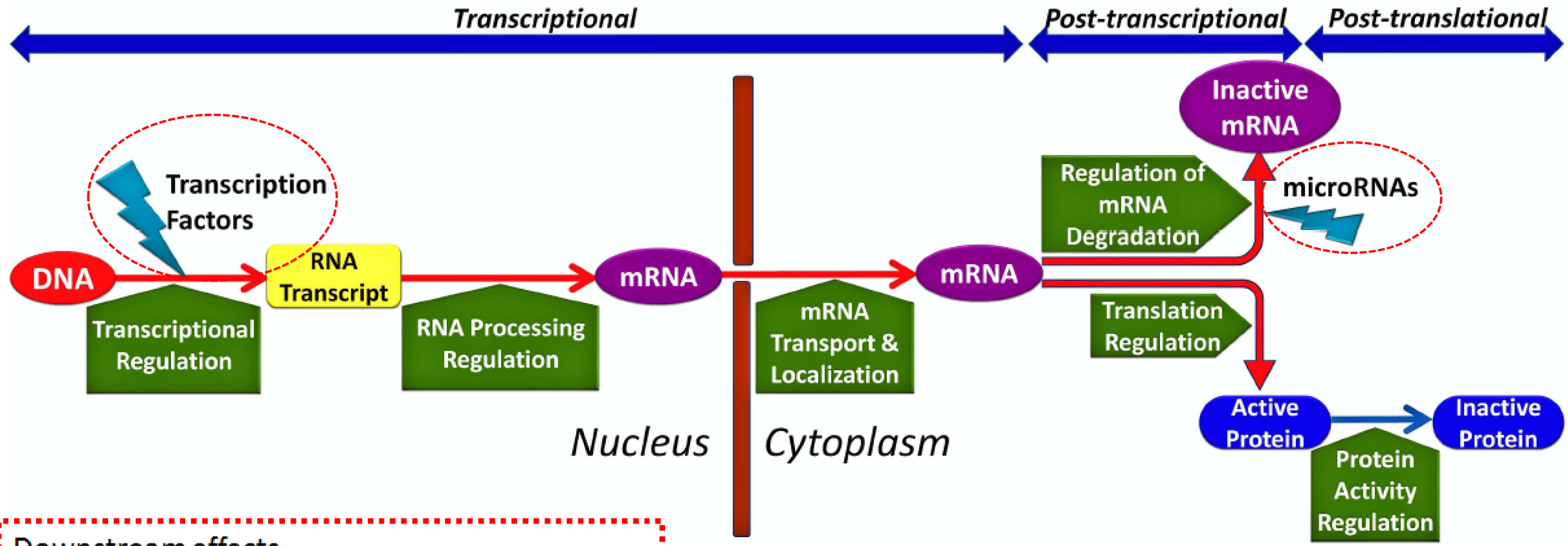
GENESYMBOL	INTRON	FLANKING	5'UTR	3'UTR	NON-SYNON	SYNON
CTNND2	1318	7	0	2	2	6

p53 Networks

Some well known and not so well known facts about p53

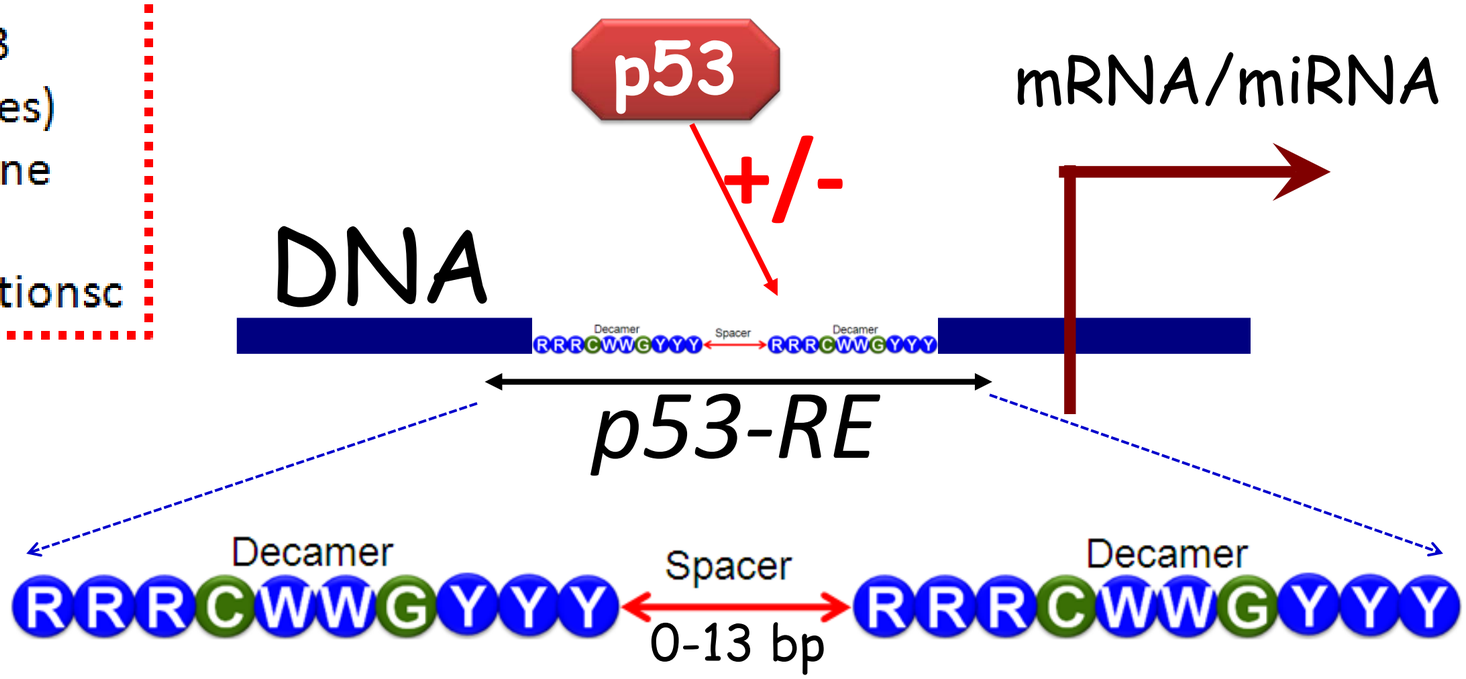
- One of the most widely studied protein
- Tumor suppressor in humans and other mammals
- Loss or mutation: Strongly associated with an increased susceptibility to cancer (>50% of human cancers)
- p53-null mice develop normally – No major functions in normal physiology(?)
- Other functions of p53 are being uncovered
  - Regulating longevity and ageing
  - Glycolytic pathways - endurance and overall fitness
  - Apoptotic responses during ischaemic & other types of stress
  - Angiogenesis
  - Reproduction
  - Development

Different levels of gene regulation



Downstream effects

- Changes in the downstream target genes or microRNAs of p53 that prevent p53 binding (polymorphic p53 binding sites)
- Changes in the downstream target gene 3'UTRs that prevent or make them susceptible to p53-induced miRNA actionsc



Functional evolution of the p53 regulatory network through its target response elements

Anil G. Jegga*†, Alberto Inga‡, Daniel Menendez§, Bruce J. Aronow*†, and Michael A. Resnick§¶

*Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229-3039; †Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267; ‡Molecular Mutagenesis Unit, National Institute for Cancer Research, 16132 Genoa, Italy; and §Laboratory Molecular Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709

Edited by Eviatar Nevo, University of Haifa, Haifa, Israel, and approved December 5, 2007 (received for review May 18, 2007)

Transcriptional network evolution is central to the development of tive functional consequences of diverged REs on promoter

Goal: To address the evolution of the p53 master regulatory network

Question: To what extent functional variation tracks with sequence variation?

How: Analysis of 47 previously validated human and mouse p53 target REs - Combination of sequence inspection, RE rules, and direct determination of p53-mediated transactivation capacity in yeast and mouse cells.

Results: Functional differences were often not predicted from consensus sequence evaluations. Of the established human p53 REs analyzed, 91% had sequence conservation in at least one nonprimate species compared with 67.5% for functional conservation. Surprisingly, there was almost no conservation of functional REs for genes involved in DNA metabolism or repair between humans and rodents, suggesting important differences in p53 stress responses and cancer development.

Sequence ID	Total length of allgn sequences	p53RE length	Mismatches to Human	Mismatches to Consensus	Heatmap Rank

RE Sequence Alignment

- A — Not part of p53 RE sequence
- p53 RE CWWG core
- Spacer sequence
- N — No information
- Original mismatch present in human RE
- Mismatch vs. human RE and consensus
- Mismatch vs. human RE but match vs. consensus
- Mismatch vs. human spacer

Relative predicted transactivation (Heatmap Rank)

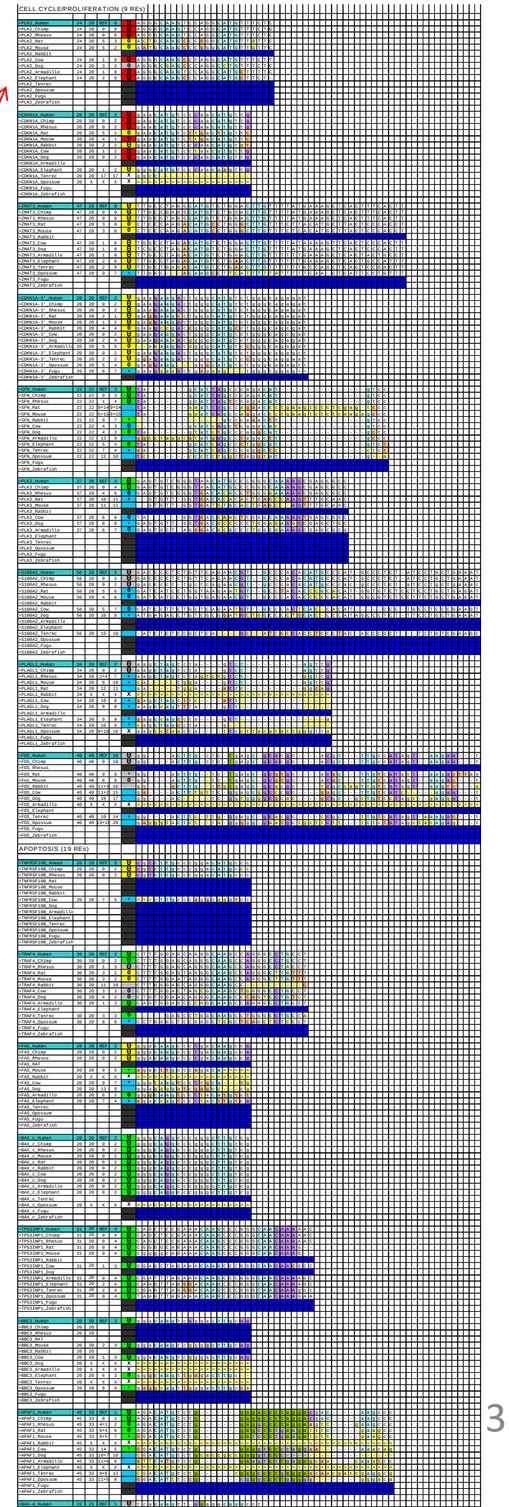
- High
- Moderate
- Slight
- Poor
- Nonresponsive to p53

p53 RE sequence conservation from reference RE (Heatmap Rank)

- 0 — High (0-2MM)
- 0 — Moderate (3-6MM)
- 0 — Poor (7-10MM)
- Very poor (11-15MM)
- X — No conservation(>15MM)
- No sequence information

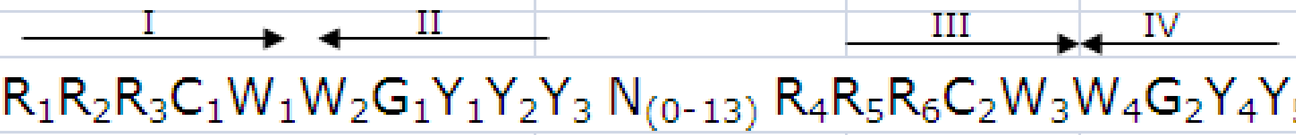
CELL CYCLE/PROLIFERATION (9 REs)

>PLK2_Human	24	20	REF	0	0	A	G	G	G	C	A	A	G	T	C	C	A	G	G	C	A	T	G	T	T	T	C	T	C
>PLK2_Chimp	24	20	0	0	0	A	G	G	G	C	A	A	G	T	C	C	A	G	G	C	A	T	G	T	T	T	C	T	G
>PLK2_Rhesus	24	20	0	0	0	A	G	G	G	C	A	A	G	T	C	C	A	G	G	C	A	T	G	T	T	T	C	T	C
>PLK2_Rat	24	20	5	3	0	A	C	T	G	C	A	A	G	C	C	C	G	G	G	C	A	T	G	T	T	G	T	T	C
>PLK2_Mouse	24	20	5	2	0	A	A	T	G	C	A	A	G	C	C	C	G	G	G	C	A	T	G	T	T	G	T	T	C
>PLK2_Rabbit						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
>PLK2_Cow	24	20	1	0	0	A	G	G	G	C	A	A	G	C	C	C	A	G	G	C	A	T	G	T	T	T	C	T	T
>PLK2_Dog	24	20	3	2	0	A	G	G	G	C	G	A	G	C	C	C	A	G	G	C	C	T	G	T	T	T	C	T	C
>PLK2_Armadillo	24	20	1	0	0	A	G	G	G	C	A	A	G	T	C	C	A	G	G	C	A	T	G	C	T	T	T	T	C
>PLK2_Elephant	24	20	2	0	0	A	G	G	G	C	A	A	G	C	C	C	A	G	G	C	A	T	G	C	T	T	C	T	-
>PLK2_Tenrec						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
>PLK2_Opossum						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
>PLK2_Fugu						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
>PLK2_Zebrafish						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
>CDKN1A_Human	20	20	REF	2	0	g	a	a	c	a	t	g	t	c	c	c	a	a	c	a	t	g	t	t	g				
>CDKN1A_Chimp	20	20	0	2	0	g	a	a	c	a	t	g	t	c	c	c	a	a	c	a	t	g	t	t	g				
>CDKN1A_Rhesus	20	20	0	2	0	g	a	a	c	a	t	g	t	c	c	c	a	a	c	a	t	g	t	t	g				
>CDKN1A_Rat	20	20	5	1	0	g	a	a	c	a	t	g	t	c	t	t	g	a	c	t	g	t	t	c						
>CDKN1A_Mouse	20	20	4	1	0	g	a	a	c	a	t	g	t	c	t	t	g	a	c	a	t	g	t	t	c				
>CDKN1A_Rabbit	20	20	2	2	0	g	a	a	c	a	t	g	t	c	c	c	a	a	c	a	t	g	t	g	t				
>CDKN1A_Cow	20	20	1	2	0	g	a	a	c	a	t	g	t	c	c	t	a	a	c	a	t	g	t	t	g				
>CDKN1A_Dog	20	20	0	2	0	g	a	a	c	a	t	g	t	c	c	c	a	a	c	a	t	g	t	t	g				
>CDKN1A_Armadillo						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N				
>CDKN1A_Elephant	20	20	1	2	0	g	g	a	c	a	t	g	t	c	c	c	a	a	c	a	a	g	t	t	g				
>CDKN1A_Tenrec	20	20	17	17	X	g	g	c	c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
>CDKN1A_Opossum	20	X	X	X	X	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=				
>CDKN1A_Fugu						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N				
>CDKN1A_Zebrafish						N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N				

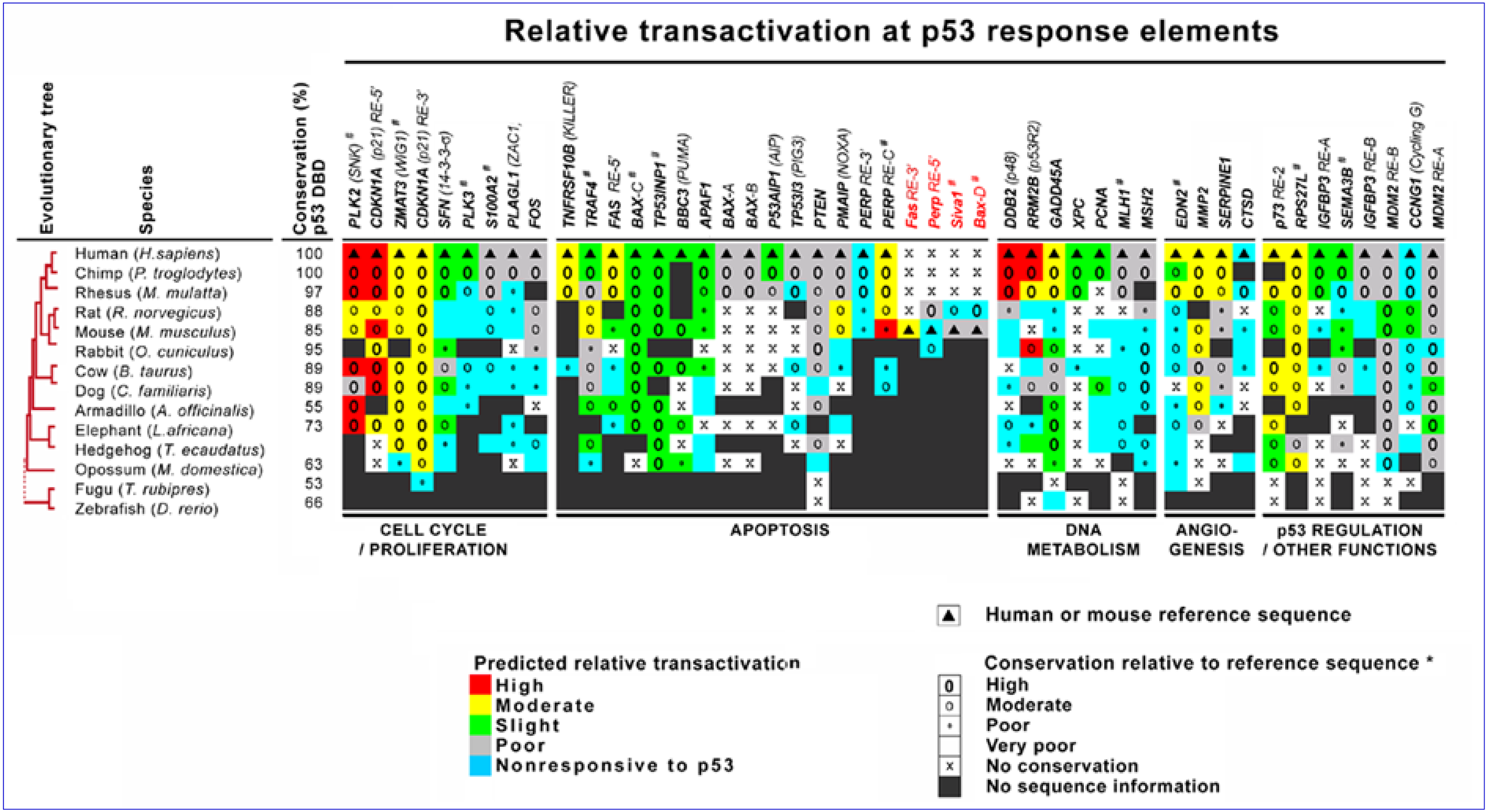


RE sequence features and functional score*

consensus p53 RE



	Mismatch position	Spacer (N) length	WW sequences	SCORE
RE matches consensus	NA	0	AT	High
			AA, TT, TA	Moderate
		1, 2	AT	Slight
			AA, TT, TA	Poor
		>2	any	Poor
A: RE with 1 mismatch in conserved C or G	C_1 or G_2	0	AT	Moderate
			AA, TT, TA	Slight
		≥1	AT	Poor
			AA, TT, TA	Non-responsive
	G_1 or C_2	0	AT	Slight
			AA, TT, TA	Poor
		≥1	any	Non-responsive
B: RE with 1 mismatch in WW	W_1 or W_4	0	AT	Moderate
			AA, TT, TA	Slight
		1, 2	any	Poor
		>2	AT	Poor
			AA, TT, TA	Non-responsive
	W_2 or W_3	0	AT	Moderate
			AA, TT, TA	Slight
		1, 2	AT	Poor
			AA, TT, TA	Non-responsive
		>2	AT	Poor
			AA, TT, TA	Non-responsive
C: RE with 1 mismatch in R or Y	R_1 or Y_1	0	AT	High
			AA, TT, TA	Moderate
		1, 2	AT	Slight
			AA, TT, TA	Poor
		>2	any	Non-responsive
	R_2 or Y_2	0	AT	Slight
			AA, TT, TA	Poor
		≥1	any	Non-responsive
	R_3 or Y_3	0	AT	Slight
			AA, TT, TA	Poor
		≥1	any	Non-responsive



Jegga et al., 2008

Variation in TFBSs can alter transcriptional control

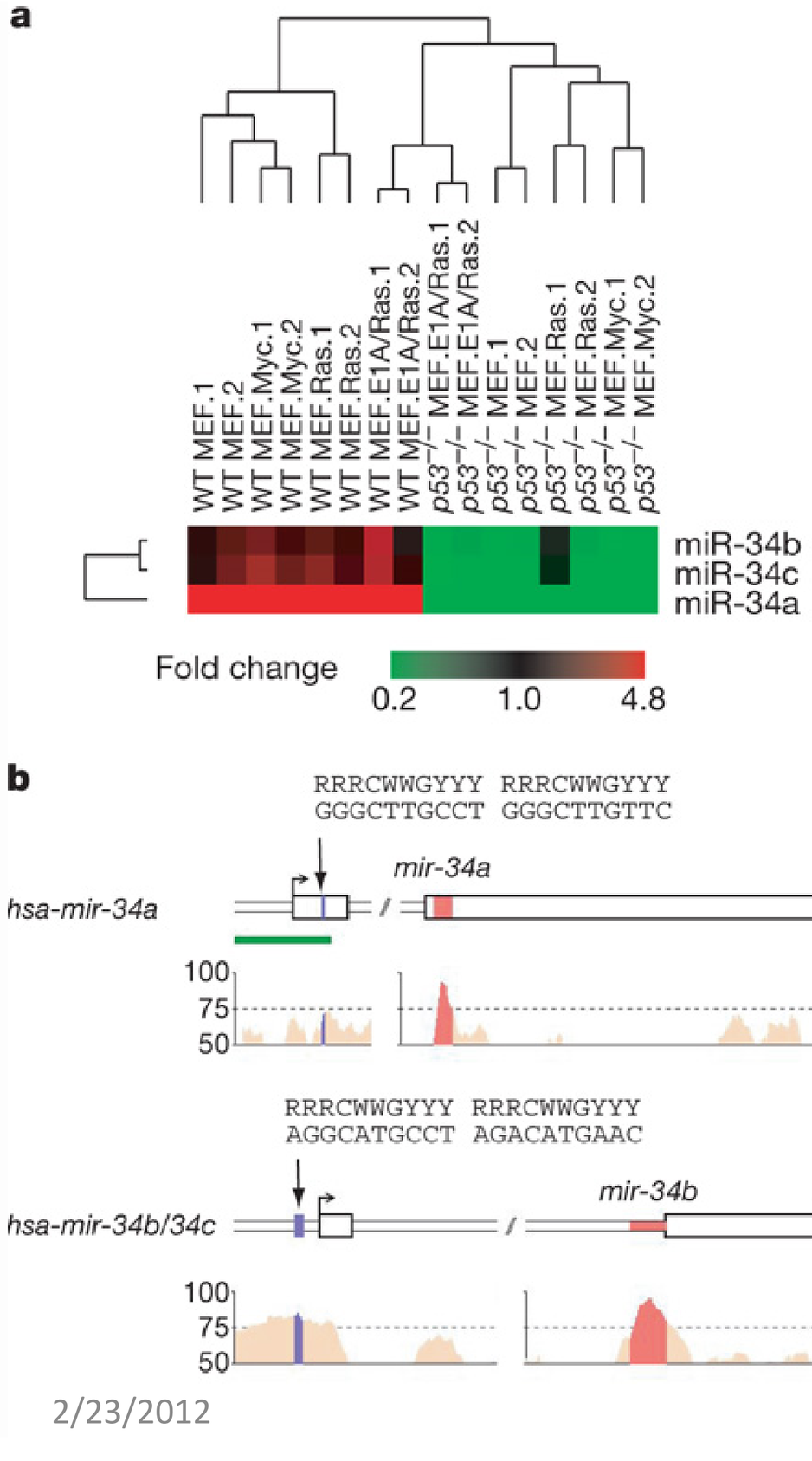
1. Human *FLT1* promoter: Genetic variation (C → T)
2. The infrequent FLT1-T variant results in the inclusion of this gene into the network coordinated by the p53

FLT1: fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)

Menendez et al., 2006 PNAS 103(5): 1406-1411

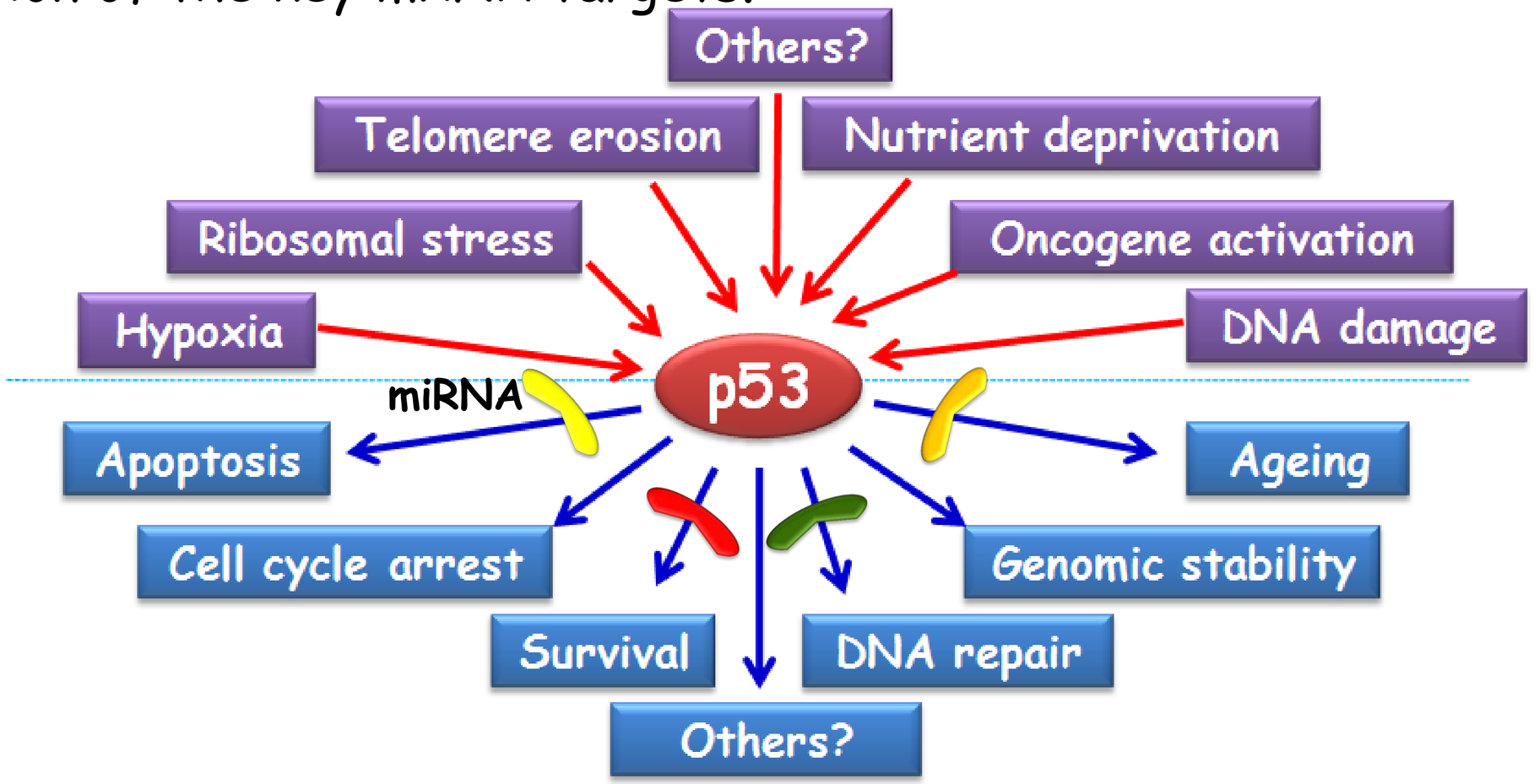
Microregulation of the master regulatory network

p53 transactivates miR-34a



1. Bommer, G. T. et al. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. Curr. Biol. 17, 1298-1307 (2007).
2. Chang, T. C. et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol. Cell 26, 745-752 (2007).
3. He, L. et al. A microRNA component of the p53 tumour suppressor network. Nature 447, 1130-1134 (2007).
4. Raver-Shapira, N. et al. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. Mol. Cell 26, 731-743 (2007).
5. Tarasov, V. et al. Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. Cell Cycle 6, 1586-1593 (2007).

Hypothesis: Multiple miRNAs are involved in p53 tumor suppressor network to provide the p53 with greater flexibility in rapidly responding to different growth condition changes perhaps by having unique miRNAs mediate the regulation of the key mRNA targets.



Research article

Highly accessed | Open Access

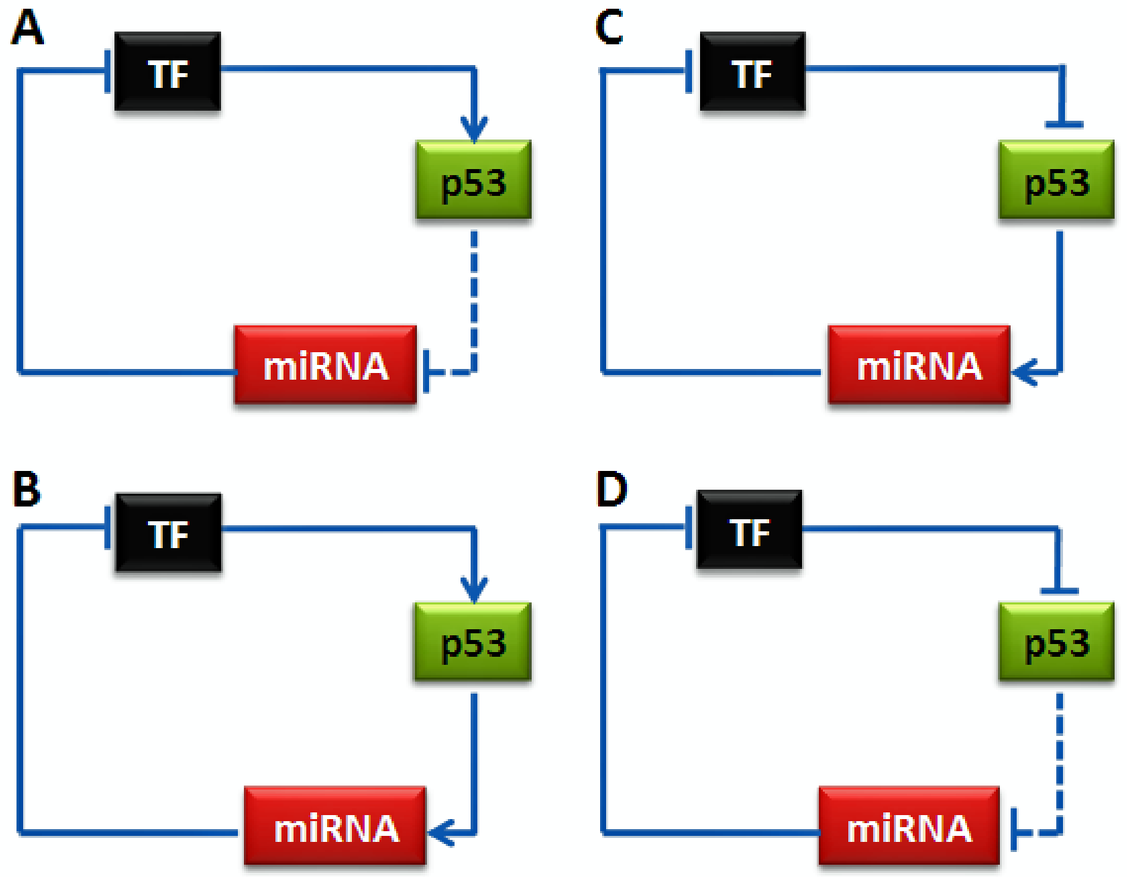
Dissecting microregulation of a master regulatory network

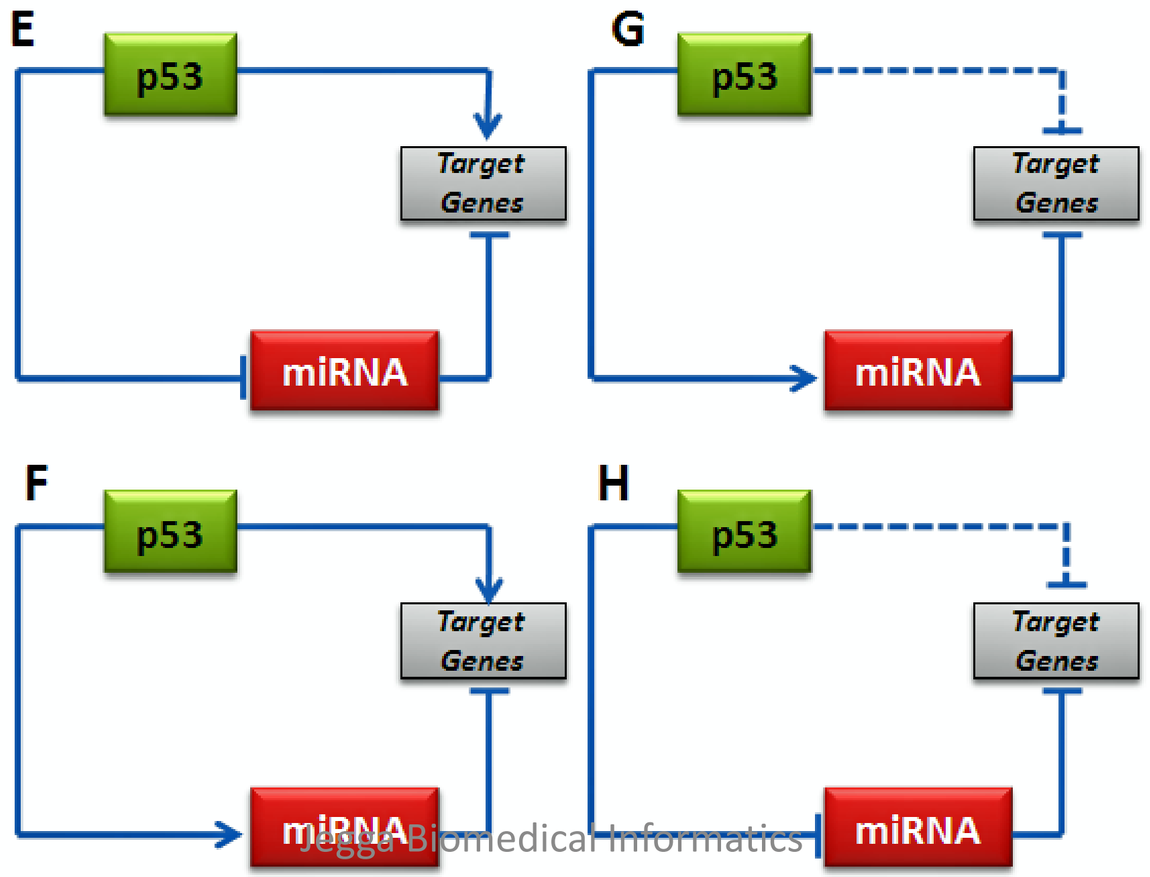
Amit U Sinha* [1], **Vivek Kaimal*** [2,4], **Jing Chen**[2,4] and **Anil G Jegga**[3,4]

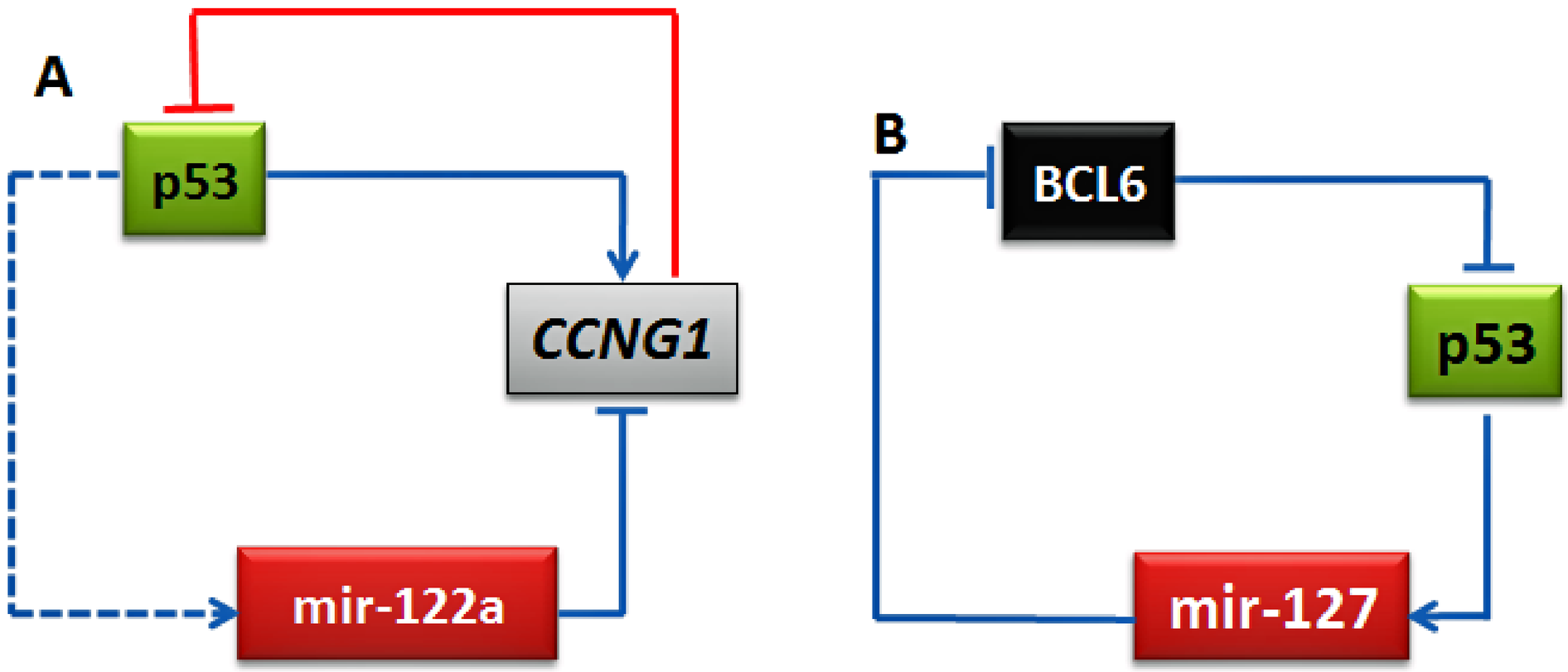
1 Department of Computer Science, University of Cincinnati, Ohio, USA
2 Department of Biomedical Engineering, University of Cincinnati, Ohio, USA
3 Department of Pediatrics, University of Cincinnati College of Medicine, Ohio, USA
4 Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Ohio, USA

author email corresponding author email * Contributed equally

BMC Genomics 2008, **9**:88 doi:10.1186/1471-2164-9-88


A
TF
p53
miRNA
C
TF
p53
miRNA
B
TF
p53
miRNA
D
TF
p53
miRNA


E
p53
Target Genes
miRNA
G
p53
Target Genes
miRNA
F
p53
Target Genes
miRNA
H
p53
Target Genes
miRNA

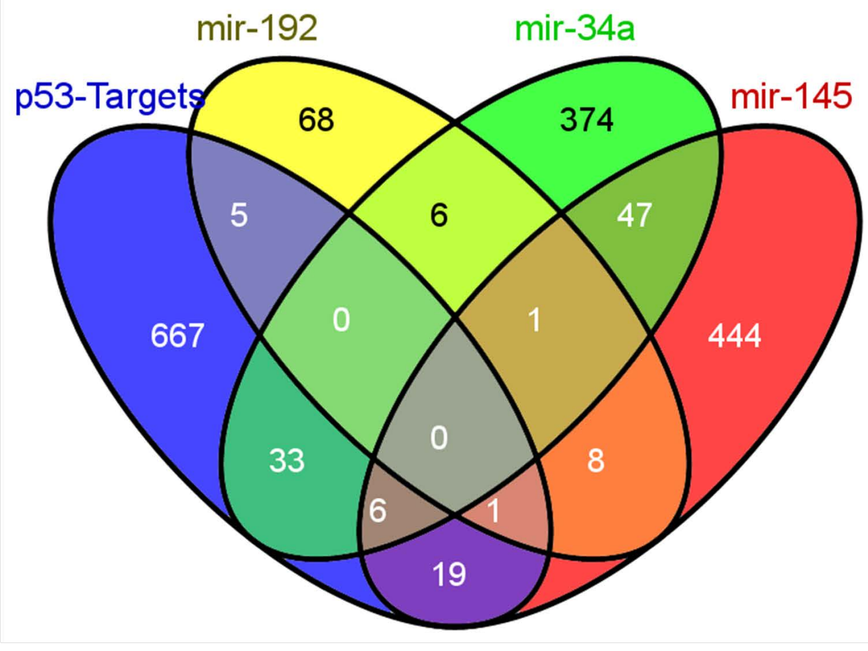

A
p53
CCNG1
mir-122a
B
BCL6
p53
mir-127
C
p53
RB1
mir-106a
D
SIRT1
p53
mir-34a

- p53
- miR-34a
- miR-192
- miR-145
- p53 target genes
- p53-miR targets that are not p53 targets
- miR-34a enriched p53 target genes

A network representation of

a. p53-target genes
b. 3 known p53-target microRNAs
c. Targets of 3 p53-miRs
d. p53 interactome

The size of the nodes is proportional to the degree (i.e. no. of edges).
Below is the Venn representation



Tomorrow's Session

Feb 24, 2012

- Genome Browsers
  - –Downloading promoters
- Identifying putative TFBS in a sequence
- Identifying putative regulatory regions
- Analyzing coexpressed genes for shared regulatory mechanisms