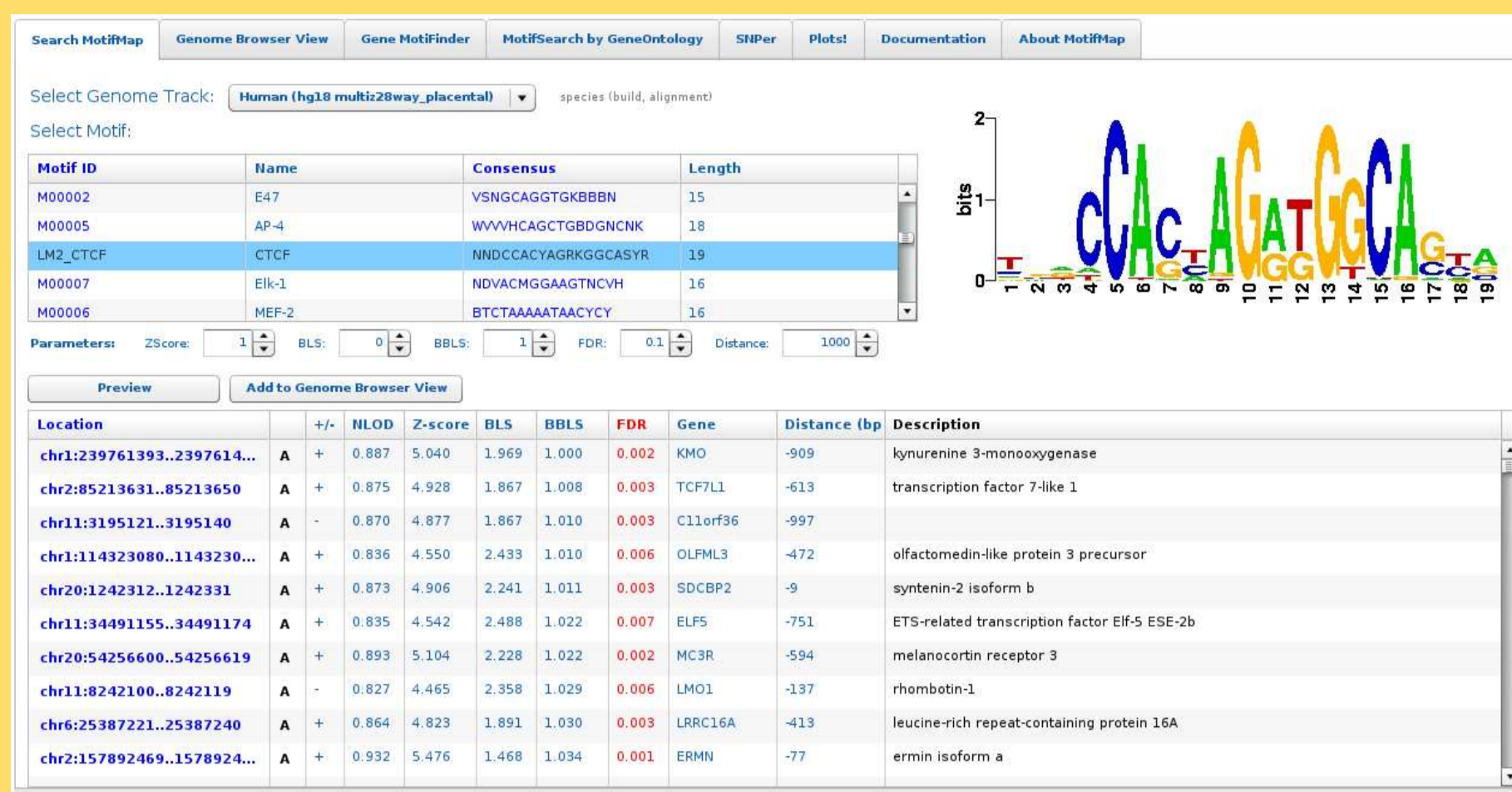


# MotifMap: Integrative genome-wide maps of candidate regulatory motif sites for model species

Kenneth M. Daily, Paul Rigor, Sholeh Forouzan, Yimeng Dou, Vishal Rajesh Patel, Xiaohui Xie, Pierre Baldi  
UCI Dept. of Computer Science, Institute for Genomics and Bioinformatics, Center for Complex Biological Systems, Dept. of Biological Chemistry

<http://motifmap.ics.uci.edu/>



Search transcription factor targeted genes, motifs associated with genes based on Gene Ontology, construct regulatory and protein interactions networks, and more!

## Introduction

A central challenge of biology is to map and understand the role of the 98% noncoding regions of the human genome.

MotifMap [1] provides a comprehensive map of potential regulatory elements in genomes in an unbiased manner.

The Updated MotifMap provides more species, new alignment, integration with other datasets, and more putative binding sites.

MotifMap derived tools also integrate external databases for SNPs, GO annotation, PPI networks and KEGG pathways.

## References and Affiliations

### References

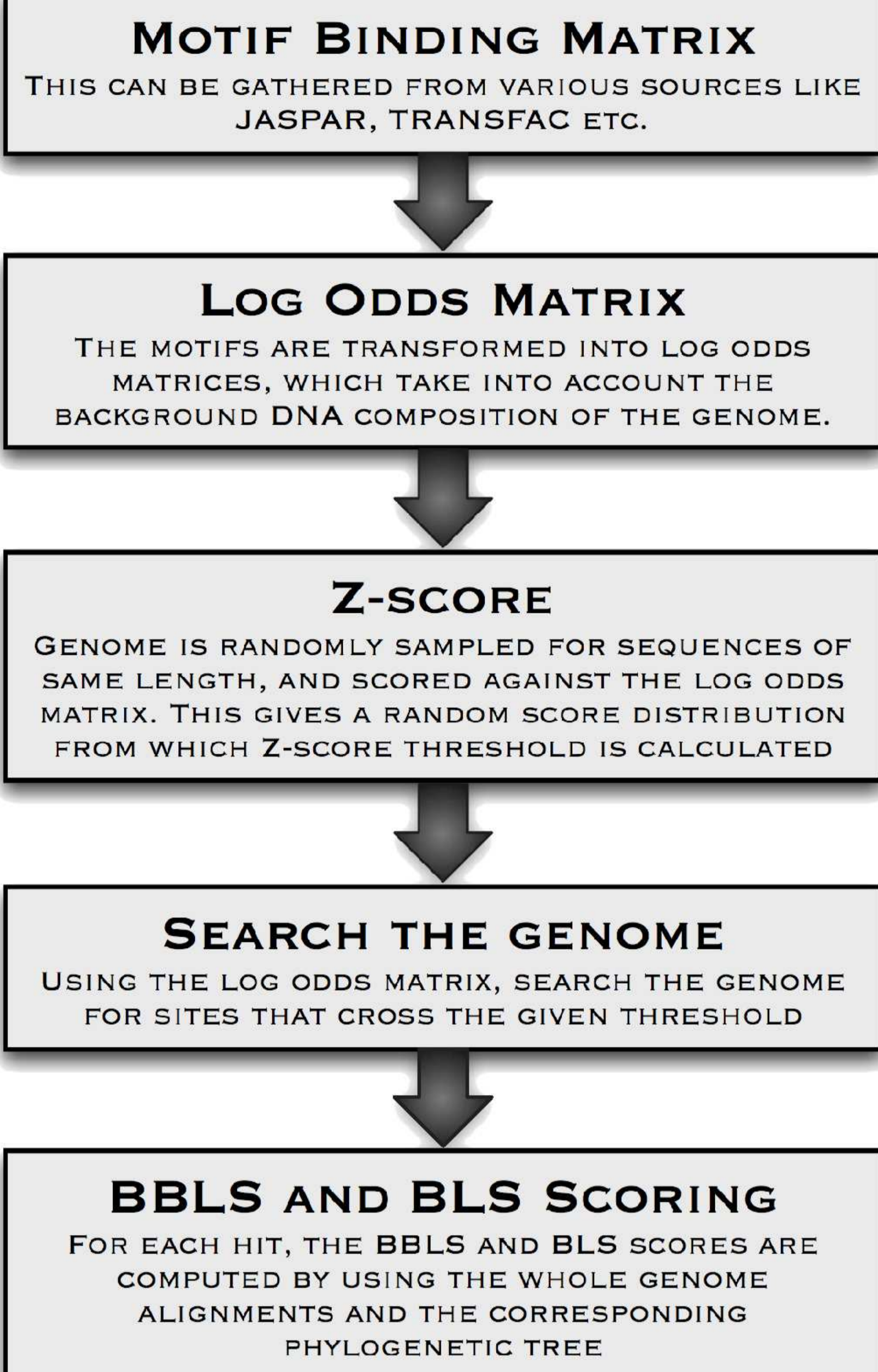
[1] Xie X, Rigor P, Baldi P: MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* 2009, 25(2):167-174, [<http://dx.doi.org/10.1093/bioinformatics/btn605>].

This work was in part supported by National Institutes of Health Biomedical Informatics Training Program Grant 5T15LM007743 and National Science Foundation Grant MRI EIA-0321390 to PB and the Institute for Genomics and Bioinformatics at UCI.

INSTITUTE FOR GENOMICS AND BIOINFORMATICS  
University of California, Irvine

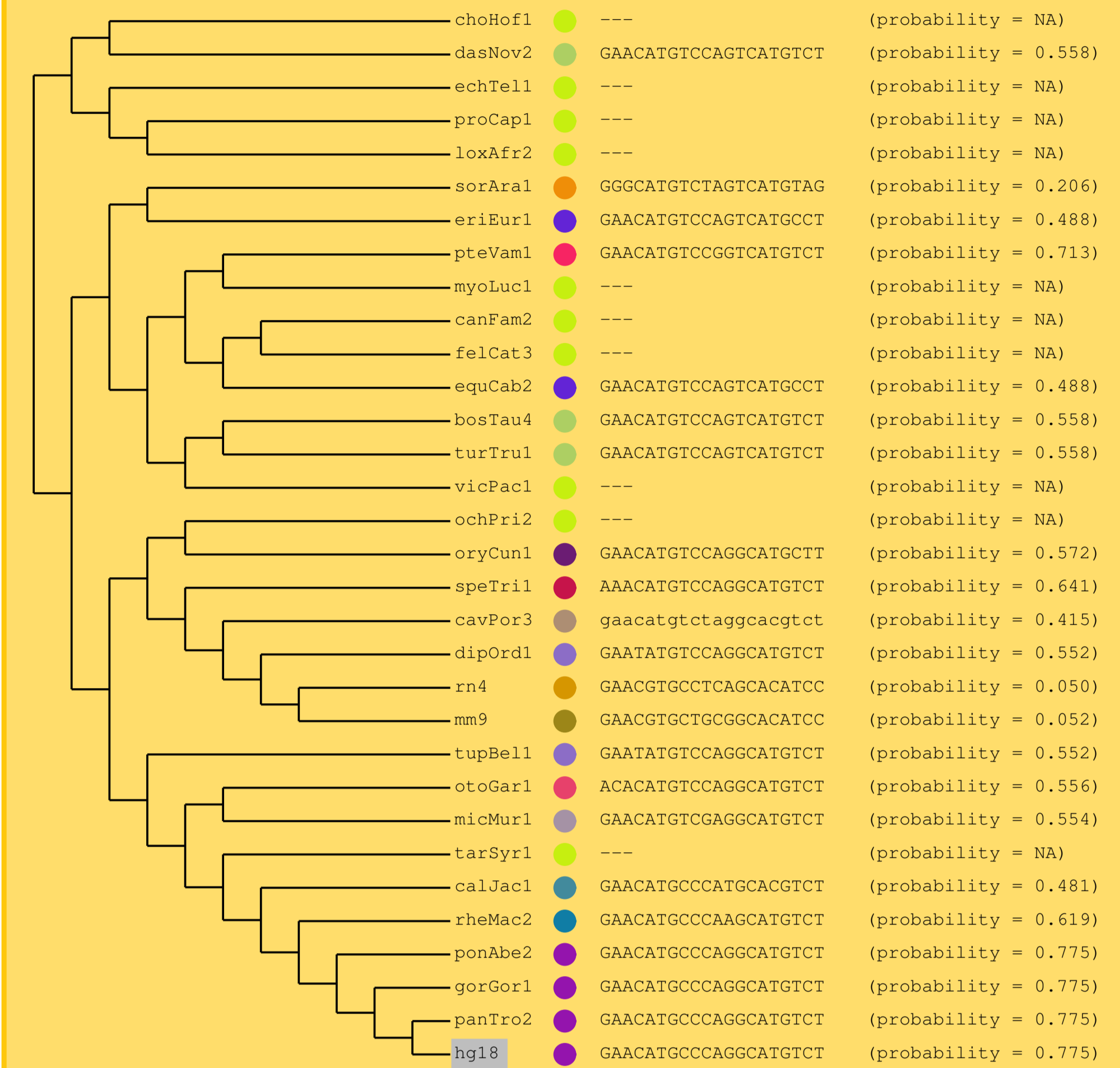
UCIRVINE | UNIVERSITY of CALIFORNIA

## MotifMap Pipeline



## Bayesian Branch Length (BBLs)

$$BBLs(p_V) = \sum_{\sigma_V} P(\sigma_V) BLS(\sigma_V)$$



BBLs accounts for conservation and how well the motif matches in each species.

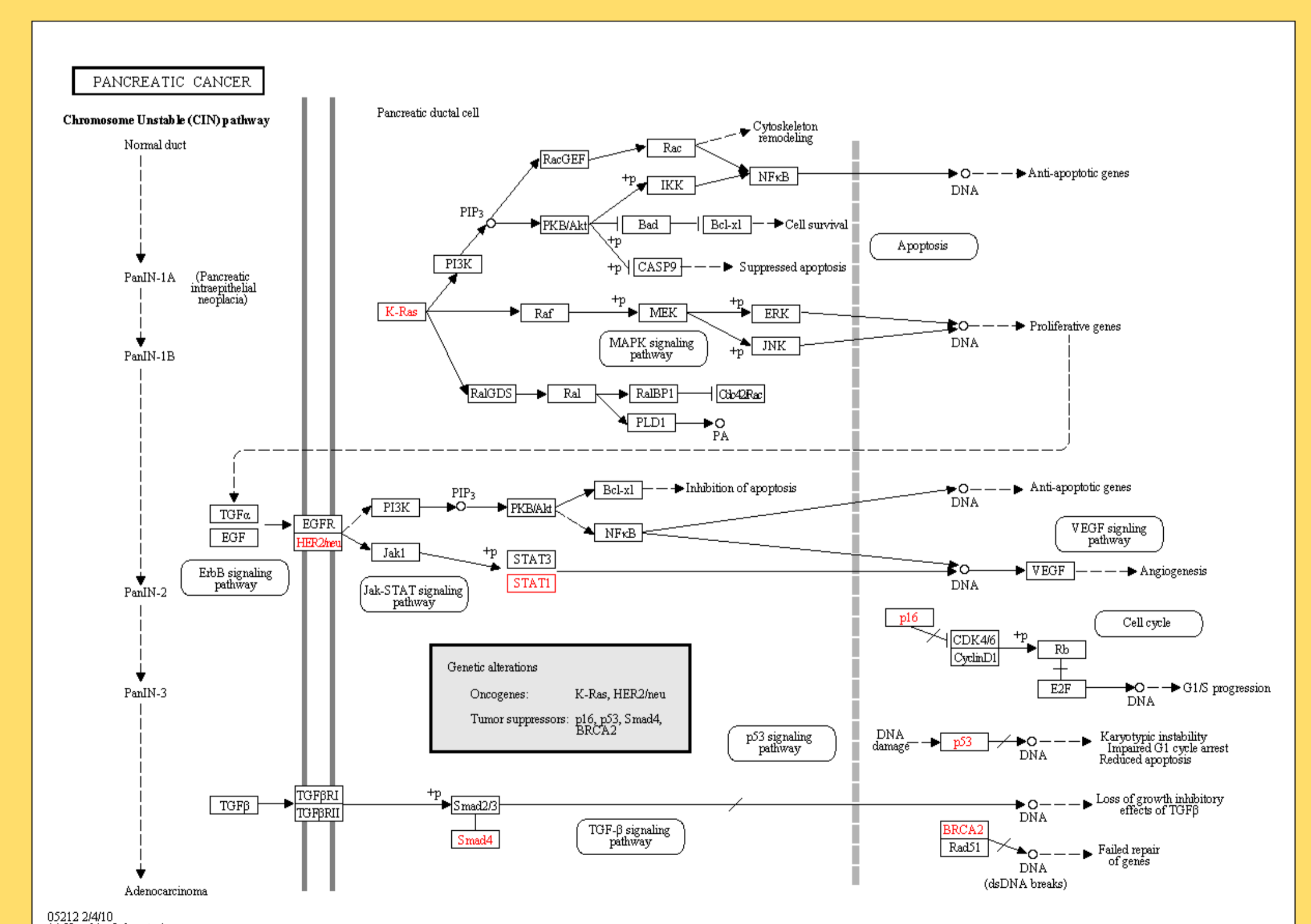
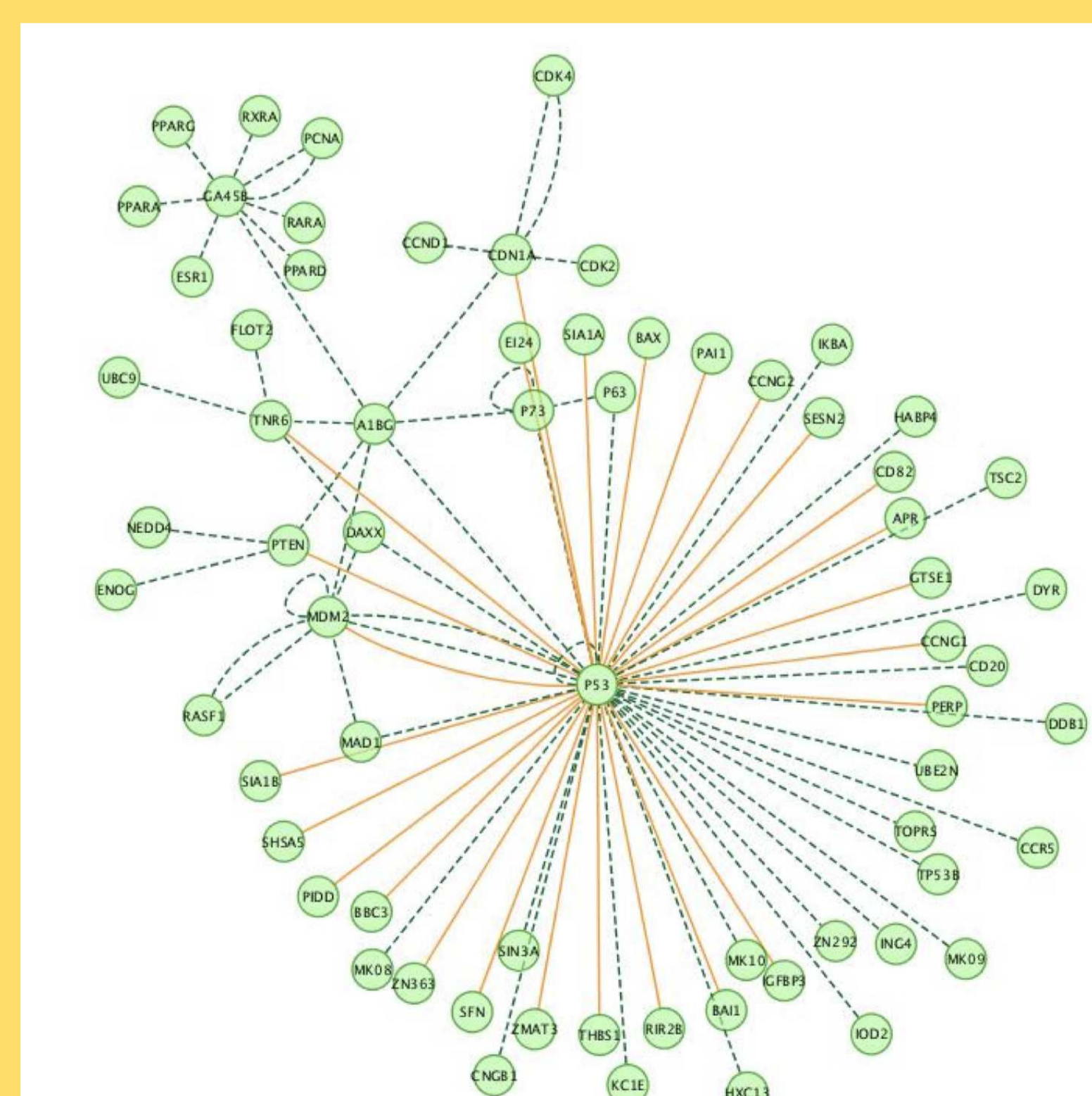
## Currently Available MotifMaps

Species	Total sites (# matrices)	FDR ≤ 0.1 (# matrices)
Human	3,903,125 (598)	874,121 (567)
Mouse	10,212,463 (613)	1,070,609 (571)
Fly	464,316 (116)	122,620 (105)
Yeast	375,135 (321)	108,411 (285)

## Validation Using ChipSeq Data

	NFKB	MYC	P53	STAT1	CTCF	NRSE
AUC						
Original	0.714	0.682	0.857	0.606	0.814	<b>0.941</b>
Updated	<b>0.758</b>	<b>0.711</b>	<b>0.886</b>	<b>0.623</b>	<b>0.829</b>	0.940
Number of sites						
Original	11,636	55,271	28,635	6,134	69,446	13,055
Updated	13,924	100,311	24,880	9,537	53,794	7,488

## PPI Integration



Network showing interacting partners of p53 with transcriptional regulation from MotifMap (orange) and protein-protein interactions (green dashed). KEGG pathway for the role of STAT1 in pancreatic cancer (MotifMap interactions in red).



