Exploring gene promoters for experimentally-proven and putative transcription factor binding sites with TRANSFAC® Professional and the included Match™ tool

TRANSFAC® Professional is a unique knowledge-base containing published data on eukaryotic transcription factors, their experimentally-proven binding sites, and regulated genes. Based on its extensive compilation of binding sites, consensus binding motifs are derived in the form of positional weight matrices which can be used with the included Match™ tool to search DNA sequences for predicted transcription factor binding sites. Comparative promoter analysis of high-throughput data based on TRANSFAC® positional weight matrices is provided in the ExPlain™ Analysis System.

This application note demonstrates the advantages of TRANSFAC® Professional versus the TRANSFAC® Public release through investigation of the VEGF gene and its binding factors.

VEGF gene regulation

In this example we've chosen to investigate regulatory mechanisms for VEGF (vascular endothelial growth factor), a well characterized gene involved in a variety of processes including angiogenesis, heart development, cell migration and more. As discussed in a recent review article (Rapisarda A and Melillo G, Adv Cancer Res. 2012, 114:237-267), VEGF's role as a major driver of tumor angiogenesis has made it a popular target of therapeutic intervention. Yet anti-VEGF therapies introduced with hopes for combatting several types of malignancies via inhibiting VEGF activity and thereby inhibiting tumor angiogenesis have proven to be ineffective for many patients, suggesting that there is more to learn about regulation of VEGF, its role in tumor angiogenesis, and other influencing factors.

We begin by querying TRANSFAC® Professional in order to identify which transcription factors have been shown in the literature to be able to interact with binding sites in the VEGF gene promoter. A Gene/Protein search for "VEGFA" followed by a secondary search for factors known to bind the gene quickly identifies 36 factor-gene interactions (for

BINDING FACTOR	VEGFA(h)	VEGFA(m)	VEGFA(r)
AML2 (m.s.)	+		
AP-2alpha (h)	+		
Arnt (h,r)	+ (+)		+
c-Fos (h)	+		
c-Myc (h)	+		
CSDA-isoform3 (h)		+	
DbpA (m)		+ (+)	
Egr-1 (h)	+		
ER-alpha (h), ER-alpha-L (h)	+		+ (+)
ER-beta (h), ER-beta-isoform1 (h)	+		+ (+)
Foxm1 (h)	+		
HAF (h)	+		
Hey2 (h)	+		
HIF-1alpha (h,r), HIF-1alpha-isoform1 (h)	+ (+)		
HIF2A (h)	+		
JunD (h)	+		
Osx (m.s.)		+	
Smad3 (h)	+		
Sp1 (h, m, r, m.s.)	+	+	
Sp3 (h, m)	+	+	
Sp4 (h)	+		
STAT3 (h)	+		
TEF-3 (m.s.)	+		
WT1 (m.s.)	+		
YB-1B (h)		+ (+)	

Table 1: Transcription factors with binding site(s) in the human, mouse or rat VEGFA genes. + indicates the binding site is present in TRANSFAC® Professional, (+) indicates the binding site is present in TRANSFAC® Public.

the human, mouse, and rat gene) with 25 different transcription factors (combining orthologs from human, mouse, and rat) (Table 1). A comparable Factor search for "VEGF" as a regulated gene in TRANSFAC® Public identifies only five transcription factor binding sites in the human, mouse and rat genes, missing important regulators such as c-Fos, STAT3, Smad3, ER and more.

Continuing on to the Locus Report for human VEGFA in TRANSFAC® Professional and the corresponding Gene Report/Entry for VEGF (VEGFA) in TRANSFAC® Public in search of more detailed information about the exact nature of the described factor-DNA interactions, we find that TRANSFAC® Professional provides data not just about individual site-based interactions but also provides data about composite elements (coordinate binding sites which act synergistically or antagonistically), ChIP-chip and ChIP-seq fragments within the vicinity of the gene, additional functionally characterized regions within the gene, as well as information about miRNAs which target VEGF mRNA as an independent mechanism of regulation (Table 2). In contrast, TRANSFAC® Public only provides information about the site-based interactions.

We would now like to have a closer look at the properties of the transcription factors found to regulate VEGFA (Table 1). In the online version of TRANSFAC® Professional this can be easily done by mapping the factors in the search result above (Table 1) to assigned ontology terms describing functional characteristics of the genes which encode the transcription factors, followed by Set Analysis for identification of statistically over-represented terms. (Note: TRANSFAC® Professional contains ontology assignments for factors only. Ontology assignments for the complete set of human, mouse, rat and other genes requires a separate subscription to PROTEOME™) .

The 36 binding factors (factor-site interactions for human, mouse, and rat gene) map to 31 encoding genes, 23 of which are shown to be expressed in various tumors and 10 of which are assigned to the gene ontology term "BP: response to hypoxia". Having identified

Human VEGFA	TRANSFAC® Professional	TRANSFAC® Public
Binding factors (monomers)	21	2
Binding sites (Factor-site interactions)	43 site entries (66 factor-site interactions)	2 site entries (1 factor-site interaction)
In vivo TF-binding fragments (ChIP-seq or ChIP-chip)	473 (within 150kb of gene), 16 (within 11,000bp promoter region)	-
Modified histone binding sites (see associated Promoter Report)	39	-
Composite elements	2	-
miRNAs (miRNA-mRNA interactions)	9	-
References	118	3

Table 2: Table 2. Overview of data provided in the human VEGFA Locus Report or its associated Promoter Report, including features mapped to specific nucleotide coordinates, in TRANSFAC® Professional.

a set of transcription factors that are known to be able to bind VEGFA and have also been shown to play a role in response to hypoxia, we're interested in what other genes they may also regulate. With the tight integration between the matrix library and the Match tool in TRANSFAC® Professional, we can easily retrieve the postional weight matrices (consensus binding motifs) for our transcription factors of interest using the "search within results" option and then choosing to forward the matrix list to the Profile generation tool within Match for creation of a Match profile.

We find that all of the selected transcription factors are associated with at least one matrix (Table 3), in comparison to TRANSFAC® Public which

Binding factors of VEGFA which show response to hypoxia	TRANSFAC® Professional matrix	TRANSFAC® Public matrix
ARNT	+	+*
EGR1	+	+
EPAS1	+	
ESR1	+	+
FOXM1	+	
HIF1A	+	
JUND	+	+
SMAD3	+	
Sp1	+	+
Sp3	+	

Table 3: Overview of positional weight matrix coverage by transcription factor. *Note that TRANSFAC® Public includes matrices for the AhR:Arnt complex, but not for the HIF1alpha:Arnt complex.

lacks matrices for five of the factors including HIF1A and EPAS1 which are known to be particularly important for hypoxia-dependent gene regulation.

Our created profile representing all 10 selected transcription factors can now be used by Match for a matrix based binding site search in promoters of other genes which are shown to be co-expressed with VEGF under hypoxic conditions. As TRANSFAC® Professional contains promoter sequences for all human, mouse, and rat genes it's easy to export the desired FASTA-formatted sequences (ranging from 10,000 bp upstream of the TSS to 1,000 bp downstream of the TSS) for use with Match. In this case, our interest is in the promoters of genes such as ACE2, CCL2, CXCR4, NOS3 and others which, like VEGF, have been shown to play a role in hypoxia and are also expressed in capillaries, the site of angiogenesis.

Conclusion

In this application note we have demonstrated significant differences in the content between TRANSFAC® Professional and TRANSFAC® Public, as well as significant differences with regard to the usability of that content due to greater connectivity with Match and other supporting tools in TRANSFAC® Professional.

BIOBASE
BIOLOGICAL DATABASES

To learn more about TRANSFAC Professional, please visit our website at **www.biobase-international.com/product/transcription-factor-binding-sites** or email **info@biobase-international.com**.