The 9aaTAD activation domains in the four Yamanaka Oct4, Sox2, Myc, and Klf4 transcription factors essential during the stem cell development

Martin Piskacek[1*], Kristina Jendruchova[1], Martina Rezacova[1], Marek Havelka[1]
Norbert Gasparik[1], Alena Hofrova[1] and Andrea Knight[2*]

Running title: ***Nine-amino-acid TransActivation Domains***

Department of Pathological Physiology, Faculty of Medicine, Masaryk University Brno, Kamenice 5, 625 00 Brno, Czech Republic

[1] Laboratory of Cancer Biology and Genetics

[2] Gamma Delta T Cell Laboratory

All authors contributed equally to this work

[*] These authors should be considered as joint senior authors and are both corresponding authors:

piskacek@med.muni.cz and knight@med.muni.cz

Keywords: activation domain, 9aaTAD, phase separation, condensate, Oct4, Pou5f, Sox2, Myc and Klf4.

Abstract

Somatic cells can be reprogrammed by the Yamanaka factors Oct4, Sox2, Myc and Klf4 activators into induced pluripotent stem cells. Throughout their genome, the Oct4, Sox2 and Klf4 cooperate with mediators of transcription, where the DNA binding sites serve as scaffolds for the phase-separated transcriptional condensates at distinct genome loci. In this study, we identified the 9aaTAD activation domains as the common interaction interface of the Yamanaka factors for transcription machinery. All four activation domains were identified by our online 9aaTAD prediction service and experimentally confirmed as strong activators of transcription. We considered the mediator interactions granted by 9aaTADs as part of the Yamanaka factors ability to reprogram cell fate.

Introduction

Our longstanding effort has been to determine the Nine-amino-acid TransActivation Domains (9aaTADs) in eukaryotic activators. Previously, we identified the 9aaTAD in a large set of transcription activators that universally recruit multiple mediators of transcription (1–11). The 9aaTAD family is represented by Gal4, p53, MLL, TCF3/E2A and SREBP1. The 9aaTADs were found in the SP/KLF family (Sp1, Sp2, Sp3, Sp4, Klf1, Klf2, Klf3, Klf4, Klf5, Klf6, Klf7, Klf8, Klf12, Klf15, WT1), the SOX family (Sox18 and SoxE), in hormone receptors (RARa, HNF4, PPAR, VDR, NHR49), in yeast transcription factors (Oaf1, Pip2, Pdr1, Pdr3, Rtg3, Gln3, Gcn4, Pho4, Msn2, Msn4, Met4), and in artificial activators of transcription (P201, B42, p53-ECapLL, KBP 2.20, pRJR200, G80BP-A, G80BP-B (12–19).

Although the activation domains have enormous variability, they are universally recognized by transcriptional machinery throughout eukaryotes (20). The 9aaTADs have the competence to activate transcription as small peptides (9 to 14 amino acids long). Currently, all tested human 9aaTADs have been shown to be functional in yeast. Therefore, we considered the 9aaTADs universal function in all eukaryotes as a further ground property of the 9aaTAD family (15, 17, 18, 20, 21). Besides the amino acid pattern, a specific distribution of amino acids in the 9aaTAD is essential. Among the characteristic features is a tandem of hydrophobic clusters surrounded by hydrophilic regions (15, 18). These hydrophobic clusters are often accompanied or are fully substituted by hydrophilic amino acids with aromatic side chains as are tryptophan (W) or tyrosine (Y), e.g. in p53. Accumulation of valines or isoleucines can compromise or fully inactivate the 9aaTAD function (22).

The 9aaTADs are well balanced with hydrophilic amino acids, which may be either positively or negatively charged. Some of them are acidic and negatively charged *e.g.* Gal4 others are positively charged *e.g.* Sp1. The general acidic character of activation domains is old and sadly fixed error, which still persists for Gal4 and also other activators (23). The Gal4 activation domain (DDVY N YLFD) is conserved only in the 9aaTAD region especially in ancestral orthologs as are *Hansenula fabianii* or *Wickerhamomyces ciferrii*, whose activation domain is hardly acidic (NDFY S LIFN)(24). Moreover, we recently demonstrated that the exchange of all negatively charged residues for positively charged ones and *vice versa* in Gal4 and respectively in Sp1 activation domains did not interfere with their function as strong activators of transcription (22).

The 9aaTADs bind to one or more of the mediator's domains on MED15 or CBP/p300 (1, 23). From structural data for the E2A and MLL activation domains in complex with the KIX

domain of CBP mediator, the 9aaTADs form short helices whose lengths vary from 9 to 12 aa (14). Online 9aaTAD prediction (using a residue position matrix search and amino acid clustering) is available on www.piskacek.org. The curated 9aaTADs have been annotated on the protein database UniProt, which now accounts for 145 annotations with 39 human activators (https://www.uniprot.org/uniprot/?query=9aatad&sort=score) (Including the current study, the list of 9aaTAD annotations will be further extended on 165 annotations with 52 human activators).

The Yamanaka transcription factors including the Oct4, Sox2, Myc and Klf4 activators, are all essential for chromatin remodelling and gene activation during the cell reprogramming (25). These activators can reprogram somatic cells into induced pluripotent stem cells (iPSCs) (26). In pluripotent cells, Oct4 / Pou5f from POU family associates with Sox2 to maintain pluripotency or with Sox17 to induce primitive endoderm commitment. The direct interaction between Oct4 and Sox2 is DNA dependent and involves the POU helix A1 from Oct4 and the HMG helix A3 from Sox2 (27). Klf4 cooperates with Oct4 and Sox2 (28, 29) to establish embryonic stem cells (ESCs). Oct4 and Sox2 can form a complex and cooperate during cell reprogramming (30, 31). The activation domain and associated SUMO-interacting motif (SIM, amino acid pattern K/RxE) are required for KLF4 protein stability and are essential for cell pluripotency (32–34).

The SOX family are close relatives of transcription factor SRY (SOX for SRY-related HMG box and SRY for sex-determining region Y) (35–37). The SOX transcription factors are divided into groups A to H and are involved in cell fate determination, development and cancer (38–41).

The most prominent member of the SOX family is Sox2, which supports cellular reprogramming and stem cell pluripotency (40). The Sox2 transactivation is associated with the general TFIID activator complex (42) and with stem cell XPC activator complex (38, 43, 44), with Oct4 (45), with Pax6 (46), with Trim24 (47) and with PARP-1 (48). Sox2 post-translational modifications have been reported to modulate transactivation activity (49–51). Sox2-mediated cell reprogramming could be further enhanced by Sox2 fusion with a strong viral activator VP16 (52). Sox1, Sox3, and Sox15 can replace the function of Sox2 in mouse ES (53, 54).

The gene expression programs, which are controlled by master transcription factors, define the identity of each cell. Recent studies have revealed that master activators form phase-separated condensates with the mediators of transcription at specific genomic loci containing their binding sites (55). The phase separation is a universal cooperative mechanism for

transcription, focusing the transcriptional machinery onto DNA enhancer sites, including Oct4, Sox2 and Klf4 binding sites (56, 57). Moreover, activation domains drive nucleosome eviction after activators binding to specific genomic loci (58) and after Oct4 binding, the chromatin accessibility is facilitated by chromatin remodelling factors (59, 60).

In our study, we have focused on the of transcription factors Oct4, Sox2, Myc and Klf4 to determinate their activation interface for the interaction with mediators of transcription, which is a part of cell reprogramming.

Results

POU family including Oct4

We have performed our online 9aaTAD prediction for Pou5f / Oct4 (9aaTAD prediction service online, www.piskacek.org) (18) and revealed a single perfect hit to amino acid sequence GHLASDFAF (**Figure 1**). By sequence alignment, we have identified the activation domains in other members of the POU family (**Figure 1**).

We generated LexA constructs, which included the prokaryotic DNA binding domain LexA and POU's region coding for the 9aaTAD from selected members of clade 1, 3, and 5. The constructs were tested for their ability to activate transcription (**Figure 2**). Both tested human POU activation domains of Pit1 and Pou5f1 have the capacity to activate transcription. Less activity was observed for Pou5 orthologs in Callorhinus ursinus (cur Pou5, Northern fur seal, gene ID: XP_025704146) and Sarcophilus harrisii (shr Pou5, Tasmanian devil, gene ID: G3VJ82), but none in Lingula unguis (lak Pou, brachiopod, gene ID: XP_013392366) or Callorhinchus milii (cmk Pou5, Ghost shark, gene ID: XP_007894073).

SOX family

We have performed online 9aaTAD prediction for Sox2 and revealed two hits above 80% match (**Figure 1**). The first hit was positioned within the DNA binding domain and was therefore excluded. The second hit was located in the Sox2 region 272-280, which corresponds to reported 9aaTAD activation domains in Sox18 (61) and SoxE (12). By sequence alignment, we have identified activation domains in other human members of the SOX family and their distal orthologs from early metazoans, such as Mnemiopsis leidyi (Ctenophora), Amphimedon queenslandica (Porifera), Strongylocentrotus purpuratus (Echinodermata) and Saccoglossus kowalevskii (Hemichordata), which shared significant similarity with Sox2 and other members

of the family (**Figure 3**). Importantly, members of the groups B and F have supported the common origin of their activation domains. In contrast, the well-studied paralogs SRY, Sox15 and Sox30 (62–66) have largely diverged and do not share a clear 9aaTAD motif or significant activation domain homology with the aforementioned members of SOX family.

We generated LexA constructs, which included the prokaryotic DNA binding domain LexA and SOX's region coding for the 9aaTAD from selected members of clade B, C, E and F. The constructs were tested for their ability to activate transcription (**Figure 2**). All tested SOX activation domains have the capacity to activate transcription.

MYC family

We have analysed the 9aaTAD activation domains in MYC activators. The MYC family has multiple conserved regions including MYC boxes MB0 to MBIV and the DNA binding domain (67) (**Figure 4**). From previous studies, the c-Myc region 105-143 was essential for transforming activity (68, 69). Insertion at codon 105 or removal of region 106-143 diminished c-Myc transcriptional activity (70). Based on this, we predicted the activation domain to be localised between MYC boxes MBI and MBII (**Figure 4**).

To identify the 9aaTAD activation domain, we ran our online 9aaTAD prediction service for trans-activating region 55-120 (localised between MYC boxes MBI and MBII) and revealed a 92% hit (sequence EMVT E LLGG, region 100-120) (**Figure 1**). Similarly, we revealed a 100% hit for the c-Myc mutant, which has an intact activation potential despite of the internal deletion (fusion result for modified trans-activation region with modified sequence EMVT E LLVS, fusion of regions 100-105 to 144-146).

We have analysed several previously reported mutants with surprising phenotypes. Reported insertion of glutamic acid into c-Myc (EMVT EL/ E /LVS, region 100-105 / inserted E / 144-146) (70) diminished the 9aaTAD pattern and its function. The activation domain 9aaTAD was also destroyed in mutants with deletion 3-103 and 106-143. Nevertheless, partial activity was observed in mutants with deleted regions 41-103, 56-103, 93-103 (including also small inserts due to DNA manipulation), which seem to have a partial repair of their 9aaTAD motif (QSEL / E LLGG, ELLD / E LLGG and GSSI / E LLGG).

The longest reported construct with full activity included the c-Myc region 1-262 (69). Despite of the presence of the activation region, a longer construct including region 1-336 has lost abruptly the capacity to activate transcription. Similarly, the mutant with deleted region 7-91 lost the capacity to activate transcription (70). Additionally, mutants with deleted regions 3-53 or 145-304 have no capacity to activate transcription regardless of their activation domains

(regions 55-100). Together, we concluded that in these mutants with intact activation domains, the local structural aberrations hindered function of their activation domains.

The 9aaTAD activation domains are well conserved in the active human Myc paralog N-Myc, but completely absent in human paralog L-Myc, what well corresponds with its poor activation potential (71). Interestingly, we found conservation of the activation domains in all three Myc paralogs in lobe-finned bony fish coelacanth (*Latimeria chalumnae*, fin-to-limb transition, related to lungfishes and tetrapods) but not in reptiles and other higher metazoans, which suggests that loss of L-Myc activation capacity have occurred already in early tetrapod evolution (**Suppl. Figure S1**). Besides the L-Myc clade, we found conservation of the activation domains in the entire Myc family from human to the last unicellular ancestor of animals, the *Monosiga brevicollis* (72).

The regions coding for predicted activation domains were fused with the DNA binding domain of LexA in construct i) mbr Myc for flagellate filozoans *Monosiga brevicollis* (gene ID: A9V5B4), ii) construct aqu Myc for poriferans *Amphimedon queenslandica* (gene ID: XP003390966), iii) constructs lch c-Myc, lch N-Myc, lch L-Myc for lobe-finned fish *Latimeria chalumnae* (gene IDs: XP005992710, XP005993959, XP006009573), and iv) constructs hsa c-Myc and hsa N-Myc for humans. No construct was created for the human L-Myc, which has deleted activation domain and was reported as a poor activator (71). All tested Myc activation domains from human and coelacanth have the capacity to activate transcription (**Figure 2**). Previously, the accumulations of valines and isoleucines in activation domains of SP and KLF transcription factors were linked with their full or partial inhibition, which corresponds well to lower activities of mbr and aqu Myc constructs in our transactivation assay (22).

Discussion

The activation domains have been reported as intrinsically disordered regions, which facilitate fuzzy binding with mediators of transcription. Their binding mechanism should be sequence independent because mutation of nearly every residue within is tolerated (23). In spite of enormous activation domain variability observed within the natural and also artificial activation domains, we generated a 9aaTAD pattern and prediction algorithm based on the amino acid distribution and clustering (18). In this and our other studies, the 9aaTAD online prediction (https://www.med.muni.cz/9aaTAD/index.php) generated reliable predictions, which led to the identification of over hundred activation domains, which have been experimentally confirmed as powerful activators of transcription. Moreover, the 9aaTAD structures seem to be helical after binding to mediators and their interaction with mediators were found not fussy but rather well fixed after binding e.g. MLL, E2A, Myb, p53 (1, 14, 73, 74). We also observed high similarity and partial identity in some of the 9aaTADs e.g. MLL and E2A (SDIM D FVLK - SDLL D FSAM) or human Myc and rat p53 (EMVT E LLGG - QDVA E LLEG) or human / Crassostrea gigas KLF4 and Oaf2 (LLDL D FIL / LLDY D FIL - LFDY D FLF)(14, 22). All that supports that 9aaTAD domains have a natural prevalence for some variants despite their tentatively huge variability.

In our study, we have identified activation domains in all four Yamanaka factors, Oct4, Sox2, Myc, and Klf4. All of them fulfil the criteria for the 9aaTAD domains, which activated transcription as short peptides. The loss of 9aaTAD in human L-Myc (as well as in reptiles and higher animals) explained its poor activator function and pointed to the functional diversification of MYC family during vertebrates' evolution. The members of SOX clades B, F, C, E possess functional 9aaTADs, could partially substitute for each other and their function also largely diversified first in vertebrates. Similarly, the KLF family and their successor SP family (Klf1, Klf2, Klf3, Klf4, Klf5, Klf6, Klf7, Klf8, Klf12, Klf15, WT1, respectively Sp1, Sp2, Sp3, and Sp4), all possess a functional 9aaTAD and have versatile functions in cell fate determination.

We found the 9aaTAD pattern deterioration in Callorhinchus milii (Ghost shark) cmk Pit1 (Pou1 clade) indicating that the Pit1 9aaTAD activation domain is the latter evolution event in vertebrates. Similarly, Pou5 orthologs in Callorhinus ursinus (cur Pou5, Northern fur seal) and Sarcophilus harrisii (Tasmanian devil) have less transcriptional activity and even the loss

of activity was found in Lingula unguis (brachiopod), which has also a deteriorated 9aaTAD motif. Moreover, we found a similar trend in Myc orthologs from flagellate filozoans Monosiga brevicollis and poriferans Amphimedon queenslandica, where the activation domains have substantially compromised function (disadvantageous accumulation of valines and isoleucines in their 9aaTADs, e.g. inactivated 9aaTAD in Sp2).

By cooperative interactions, active enhancers interact with mediators of transcription to enable assembly of the transcriptional machinery. The phase separation contributes to stabilization of transcription machinery at the specific genomic loci even at very low protein concentrations. Throughout the genome, the Oct4, Sox2 and Klf4 binding sites and mediators of transcription distribution have an exponential occurrence correlation, whereby the DNA binding sites serve as scaffolds for the phase-separated transcriptional condensate and cooperate during cell reprogramming. In summary, we consider the mediator interaction interface facilitated by the 9aaTADs in higher vertebrates as a part of Yamanaka factors ability to reprogram cell fate.

Methods

Constructs. The construct pBTM116-HA was generated by insertion of an HA cassette into the EcoRI site of the vector pBTM116. All constructs were generated by PCR and subcloned into the pBTM116 EcoRI and BamHI sites. All constructs had a spacer of three amino acids inserted into the EcoRI site: the peptide GSG. All constructs were sequenced by Eurofins Genomics. Further detailed information about constructs and primer sequences is available on request.

Assessment of enzyme activities. β-galactosidase activity was determined in the yeast strain L40 [58, 59]. The strain L40 has a chromosomally integrated lacZ reporter driven by the lexA operator. In all hybrid assays, we used the 2μ vector pBTM116 for generation of LexA hybrids. The yeast strain L40, Saccharomyces cerevisiae genotype: MATa ade2 his3 leu2 trp1 LYS::lexA-HIS3 URA3::lexA-LacZ, is deposited at the ATCC (#MYA-3332). The average value of the β-galactosidase activities from three independent transformants is presented as a percentage of the reference with the standard deviations (means plus and minus SDs; n = 3). We standardized all results to the previously reported Gal4 construct HaY including the 9aaTAD with the activity set to 100% [14].

Databases used in the study. UniProt, ExPASy, NCBI, Ensembl Metazoa, KEGG http://www.genome.jp, Japanese Lamprey Genome Project http://jlampreygenome.imcb.a-star.edu.sg/blast/, Sun Yat-Sen University Lancelet Genome Project http://genome.bucm.edu.cn/lancelet/blast.php, Compagen Genomics Platform http://www.compagen.org/blast.html, Florida University Neurobase https://neurobase.rc.ufl.edu/, UCSC genomic annotation https://genome.ucsc.edu, NIH https://research.nhgri.nih.gov/mnemiopsis/blast/.

Acknowledgements

This work was supported by the Ministry of Health of the Czech Republic AZV NV19-05-00410.

Author contributions

MP, KJ, MR, and MH performed the experiments. MP conceived the project. MP and AK wrote the manuscript. All authors have contributed critical intellectual content and have approved the final manuscript.

Conflict of interest

The authors declare no potential conflicts of interest.

Figure Legends

Figure 1. Activation domain prediction. The online 9aaTAD prediction service for activation domain was applied on Pou5F / Oct4, Sox2, Myc (prediction for trans-activation region 50-120 localized experimentally) and Klf4. Algorithm for the 9aaTAD amino acid pattern was applied in the search, and region clustering conformity was assessed by percentage.

Figure 2. Transactivation Assay. The regions with the identified 9aaTAD activation domains were tested in a reporter assay with hybrid LexA DNA binding domain for the capacity to activate transcription. The average value of the β-galactosidase activities from two independent transformants is presented as a percentage of the reference with standard deviation (means and plusmn; SD; n = 3). We standardized the results to positive control p53 construct 6p53, which was set to 100%. The 9aaTADs activation domains are colored for faster orientation. Deuterated 9aaTADs or their mismatch are in grey. Abbreviations: Homo sapiens (hsa), Callorhinchus milii (cmk, Ghost shark, gene ID: XP_007894073), Callorhinus ursinus (cur Pou5, Northern fur seal, gene ID: XP_025704146), Sarcophilus harrisii (shr Pou5, Tasmanian devil, gene ID: G3VJ82), Lingula unguis (lak Pou, brachiopod, gene ID: XP_013392366), Mnemiopsis leidyi (mle SoxB, comb jelly, gene ID: A0A059XHC3), Amphimedon queenslandica (aqu SoxB, gene ID: B1A9Y6), Strongylocentrotus purpuratus (spu SoxB, purple sea urchin, gene ID: Q9Y0D7 and spu SoxF, gene ID: W4YEI9), Saccoglossus kowalevskii (sko SoxB, hemichordate, common ancestor of chordata, gene ID: Q7YTD4 and sko SoxF, gene ID: B5THP2), *Monosiga brevicollis* (mbe Myc, filozoans, last unicellular ancestor of animals, gene ID: A9V5B4), *Amphimedon queenslandica* (aqu Myc, poriferans, gene ID: XP003390966), *Latimeria chalumnae* (lch c-Myc, lch N-Myc, and lch L-Myc, lobe-finned fish, related to lungfishes and tetrapods, gene IDs: XP005992710, XP005993959, XP006009573).

Figure 3. Alignment of SOX family. The C-terminal regions of SOX proteins were aligned by sequence similarities to Sox2 and predicted 9aaTAD activation domain (9aaTAD prediction service online, www.Piskacek.org). The Sox orthologs from early diverged eukaryotes Mnemiopsis leidyi (mle SoxB, gene ID: A0A059XHC3), Amphimedon queenslandica (aqu SoxB, gene ID: B1A9Y6), Strongylocentrotus purpuratus (spu SoxB, gene ID: Q9Y0D7 and spu SoxF, gene ID: W4YEI9) and Saccoglossus kowalevskii (sko SoxB, gene ID: Q7YTD4 and sko SoxF, gene ID: B5THP2) have strong conservation of 9aaTAD. Diversification of the

9aaTAD motif, especially in position p5 (in red) could be seen in human SoxF but not in spu SoxF and sko SoxF. The members of SoxE clade contain two activation domains; the 9aaTADs-I have more conserved sequence but the 9aaTADs-II have more conserved position in the protein and are associated with the SOX C-terminal domains, similarly to the 9aaTADs in all other SOX clades. The 9aaTADs activation domains are colored for faster orientation. Dots in sequences represent stop codon.

Figure 4. Alignment of the MYC family. The N-terminal regions of MYC proteins were aligned by sequence similarities and their predicted 9aaTAD activation domains are shown. The 9aaTADs activation domains are colored for faster orientation. A) Contrary to coelacanth (Latimeria chalumnae, fin-to-limb transition, related to lungfishes and tetrapods), the L-Myc from humans, reptiles and higher animals have no 9aaTAD motif and the corresponding region is deleted. B) The MYC clades with conservation MYB box O, I and II are shown. The deletion of activation domain in L clade are in grey box. Conservation of threonine in position 55 is highlighted in red. Pseudo repeat of the 9aaTAD motif with similarity to Sp2 9aaTAD is shown.

References

1. Teufel,D.P., Freund,S.M., Bycroft,M. and Fersht,A.R. (2007) Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 7009–7014.

2. Gamper,A.M. and Roeder,R.G. (2008) Multivalent Binding of p53 to the STAGA Complex Mediates Coactivator Recruitment after UV Damage. *Mol Cell Biol*, **28**, 2517–2527.

3. Feng,H., Jenkins,L.M.M., Durell,S.R., Hayashi,R., Mazur,S.J., Cherry,S., Tropea,J.E., Miller,M., Wlodawer,A., Appella,E., *et al.* (2009) Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. *Structure*, **17**, 202–210.

4. Ferreon,J.C., Lee,C.W., Arai,M., Martinez-Yamout,M.A., Dyson,H.J. and Wright,P.E. (2009) Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 6591–6596.

5. Jenkins,L.M.M., Yamaguchi,H., Hayashi,R., Cherry,S., Tropea,J.E., Miller,M., Wlodawer,A., Appella,E. and Mazur,S.J. (2009) Two distinct motifs within the p53 transactivation domain bind to the Taz2 domain of p300 and are differentially affected by phosphorylation. *Biochemistry*, **48**, 1244–1255.

6. Thakur,J.K., Arthanari,H., Yang,F., Chau,K.H., Wagner,G. and Näär,A.M. (2009) Mediator subunit Gal11p/MED15 is required for fatty acid-dependent gene activation by yeast transcription factor Oaf1p. *J. Biol. Chem.*, **284**, 4422–4428.

7. Choi,Y., Asada,S. and Uesugi,M. (2000) Divergent hTAFII31-binding motifs hidden in activation domains. *J. Biol. Chem.*, **275**, 15912–15916.

8. Uesugi,M. and Verdine,G.L. (1999) The alpha-helical FXXPhiPhi motif in p53: TAF interaction and discrimination by MDM2. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14801–14806.

9. Piskacek,M. (2009) 9aaTADs mimic DNA to interact with a pseudo-DNA Binding Domain KIX of Med15 (Molecular Chameleons). *Nature Precedings*, 10.1038/npre.2009.3939.1.

10. Piskacek,M. (2009) Common Transactivation Motif 9aaTAD recruits multiple general co-activators TAF9, MED15, CBP and p300. *Nature Precedings*, 10.1038/npre.2009.3488.2.

11. Di Lello,P., Jenkins,L.M.M., Jones,T.N., Nguyen,B.D., Hara,T., Yamaguchi,H., Dikeakos,J.D., Appella,E., Legault,P. and Omichinski,J.G. (2006) Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Mol. Cell*, **22**, 731–740.

12. Haseeb,A. and Lefebvre,V. The SOXE transcription factors—SOX8, SOX9 and SOX10—share a bi-partite transactivation mechanism. *Nucleic Acids Res*, 10.1093/nar/gkz523.

13. Piskacek,M. (2009) 9aaTAD Prediction result (2006). *Nature Precedings*, 10.1038/npre.2009.3984.1.

14. Piskacek,M., Vasku,A., Hajek,R. and Knight,A. (2015) Shared structural features of the 9aaTAD family in complex with CBP. *Mol Biosyst*, **11**, 844–851.

15. Piskacek,M., Havelka,M., Rezacova,M. and Knight,A. (2016) The 9aaTAD Transactivation Domains: From Gal4 to p53. *PLoS ONE*, **11**, e0162842.

16. Piskacek,M., Havelka,M., Rezacova,M. and Knight,A. (2017) The 9aaTAD Is Exclusive Activation Domain in Gal4. *PLoS ONE*, **12**, e0169261.

17. Piskacek,M., Havelka,M., Jendruchova,K. and Knight,A. (2018) Nuclear hormone receptors: Ancient 9aaTAD and evolutionally gained NCoA activation pathways. *The Journal of Steroid Biochemistry and Molecular Biology*, 10.1016/j.jsbmb.2018.11.008.

18. Piskacek,S., Gregor,M., Nemethova,M., Grabner,M., Kovarik,P. and Piskacek,M. (2007) Nine-amino-acid transactivation domain: establishment and prediction utilities. *Genomics*, **89**, 756–768.

19. Sandholzer,J., Hoeth,M., Piskacek,M., Mayer,H. and de Martin,R. (2007) A novel 9-amino-acid transactivation domain in the C-terminal part of Sox18. *Biochem. Biophys. Res. Commun.*, **360**, 370–374.

20. Kakidani,H. and Ptashne,M. (1988) GAL4 activates gene expression in mammalian cells. *Cell*, **52**, 161–167.

21. Fields,S. and Jang,S.K. (1990) Presence of a potent transcription activating sequence in the p53 protein. *Science*, **249**, 1046–1049.

22. Piskacek,M., Havelka,M., Jendruchova,K., Knight,A. and Keegan,L.P. (2019) The evolution of the 9aaTAD domain in Sp2 proteins: inactivation with valines and intron reservoirs. *Cell. Mol. Life Sci.*, 10.1007/s00018-019-03251-w.

23. Tuttle,L.M., Pacheco,D., Warfield,L., Hahn,S. and Klevit,R.E. (2019) Mediator subunit Med15 dictates the conserved 'fuzzy' binding mechanism of yeast transcription activators Gal4 and Gcn4 Molecular Biology.

24. Piskacek,M., Havelka,M., Rezacova,M. and Knight,A. (2017) Gal4 activation domain 9aaTAD could be inactivated by adjacent mini-inhibitory domain and reactivated by distal re-activation domain. *bioRxiv*, 10.1101/110882.

25. Chen,K., Long,Q., Xing,G., Wang,T., Wu,Y., Li,L., Qi,J., Zhou,Y., Ma,B., Schöler,H.R., *et al.* (2019) Heterochromatin loosening by the Oct4 linker region facilitates Klf4 binding and iPSC reprogramming. *The EMBO Journal*, **0**, e99165.

26. Takahashi,K. and Yamanaka,S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, **126**, 663–676.

27. Lam,C.S., Mistri,T.K., Foo,Y.H., Sudhaharan,T., Gan,H.T., Rodda,D., Lim,L.H., Chou,C., Robson,P., Wohland,T., *et al.* (2012) DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. *Biochem. J.*, **448**, 21–33.

28. Nakatake,Y., Fukui,N., Iwamatsu,Y., Masui,S., Takahashi,K., Yagi,R., Yagi,K., Miyazaki,J., Matoba,R., Ko,M.S.H., *et al.* (2006) Klf4 Cooperates with Oct3/4 and Sox2 To Activate the Lefty1 Core Promoter in Embryonic Stem Cells. *Mol Cell Biol*, **26**, 7772–7782.

29. Wei,Z., Yang,Y., Zhang,P., Andrianakos,R., Hasegawa,K., Lyu,J., Chen,X., Bai,G., Liu,C., Pera,M., *et al.* (2009) Klf4 interacts directly with Oct4 and Sox2 to promote reprogramming. *Stem Cells*, **27**, 2969–2978.

30. Lodato,M.A., Ng,C.W., Wamstad,J.A., Cheng,A.W., Thai,K.K., Fraenkel,E., Jaenisch,R. and Boyer,L.A. (2013) SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. *PLOS Genetics*, **9**, e1003288.

31. Ng,C.K.L., Li,N.X., Chee,S., Prabhakar,S., Kolatkar,P.R. and Jauch,R. (2012) Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Res*, **40**, 4933–4941.

32. Dhaliwal,N.K., Abatti,L.E. and Mitchell,J.A. (2019) KLF4 protein stability regulated by interaction with pluripotency transcription factors overrides transcriptional control. *Genes Dev.*, **33**, 1069–1082.

33. Du,J.X., McConnell,B.B. and Yang,V.W. (2010) A Small Ubiquitin-related Modifier-interacting Motif Functions as the Transcriptional Activation Domain of Krüppel-like Factor 4. *J. Biol. Chem.*, **285**, 28298–28308.

34. Presnell,J.S., Schnitzler,C.E. and Browne,W.E. (2015) KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes. *Genome Biol Evol*, **7**, 2289–2309.

35. Bowles,J., Schepers,G. and Koopman,P. (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.*, **227**, 239–255.

36. Koopman,P., Schepers,G., Brenner,S. and Venkatesh,B. (2004) Origin and diversity of the SOX transcription factor gene family: genome-wide analysis in Fugu rubripes. *Gene*, **328**, 177–186.

37. Wegner,M. (2011) SOX after SOX: SOXession regulates neurogenesis. *Genes Dev.*, **25**, 2423–2428.

38. Cattoglio,C., Zhang,E.T., Grubisic,I., Chiba,K., Fong,Y.W. and Tjian,R. (2015) Functional and mechanistic studies of XPC DNA-repair complex as transcriptional coactivator in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E2317-2326.

39. Kormish,J.D., Sinner,D. and Zorn,A.M. (2010) Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. *Dev. Dyn.*, **239**, 56–68.

40. Takahashi,K. and Yamanaka,S. (2016) A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.*, **17**, 183–193.

41. Tosic,N., Petrovic,I., Grujicic,N.K., Davidovic,S., Virijevic,M., Vukovic,N.S., Pavlovic,S. and Stevanovic,M. (2018) Prognostic significance of SOX2, SOX3, SOX11, SOX14 and SOX18 gene expression in adult de novo acute myeloid leukemia. *Leuk. Res.*, **67**, 32–38.

42. Pijnappel,W.W.M.P., Esch,D., Baltissen,M.P.A., Wu,G., Mischerikow,N., Bergsma,A.J., van der Wal,E., Han,D.W., Bruch,H. vom, Moritz,S., *et al.* (2013) A central role for TFIID in the pluripotent transcription circuitry. *Nature*, **495**, 516–519.

43. Lafrance-Vanasse,J., Arseneault,G., Cappadocia,L., Legault,P. and Omichinski,J.G. (2013) Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic Acids Res.*, **41**, 2736–2745.

44. Zhang,E.T., He,Y., Grob,P., Fong,Y.W., Nogales,E. and Tjian,R. (2015) Architecture of the human XPC DNA repair and stem cell coactivator complex. *PNAS*, **112**, 14817–14822.

45. Chang,Y.K., Srivastava,Y., Hu,C., Joyce,A., Yang,X., Zuo,Z., Havranek,J.J., Stormo,G.D. and Jauch,R. (2017) Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.*, **45**, 832–845.

46. Kamachi,Y., Uchikawa,M., Tanouchi,A., Sekido,R. and Kondoh,H. (2001) Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev.*, **15**, 1272–1286.

47. Rafiee,M.-R., Girardot,C., Sigismondo,G. and Krijgsveld,J. (2016) Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol. Cell*, **64**, 624–635.

48. Terrados,G., Finkernagel,F., Stielow,B., Sadic,D., Neubert,J., Herdt,O., Krause,M., Scharfe,M., Jarek,M. and Suske,G. (2012) Genome-wide localization and expression profiling establish Sp2 as a sequence-specific transcription factor regulating vitally important genes. *Nucleic Acids Res.*, **40**, 7844–7857.

49. Kamachi,Y. and Kondoh,H. (2013) Sox proteins: regulators of cell fate specification and differentiation. *Development*, **140**, 4129–4144.

50. Tsuruzoe,S., Ishihara,K., Uchimura,Y., Watanabe,S., Sekita,Y., Aoto,T., Saitoh,H., Yuasa,Y., Niwa,H., Kawasuji,M., *et al.* (2006) Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem. Biophys. Res. Commun.*, **351**, 920–926.

51. Van Hoof,D., Muñoz,J., Braam,S.R., Pinkse,M.W.H., Linding,R., Heck,A.J.R., Mummery,C.L. and Krijgsveld,J. (2009) Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*, **5**, 214–226.

52. Narayan,S., Bryant,G., Shah,S., Berrozpe,G. and Ptashne,M. (2017) OCT4 and SOX2 Work as Transcriptional Activators in Reprogramming Human Fibroblasts. *Cell Rep*, **20**, 1585–1596.

53. Adikusuma,F., Pederick,D., McAninch,D., Hughes,J. and Thomas,P. (2017) Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. *Genetics*, **206**, 1495–1503.

54. Niwa,H., Nakamura,A., Urata,M., Shirae-Kurabayashi,M., Kuraku,S., Russell,S. and Ohtsuka,S. (2016) The evolutionally-conserved function of group B1 Sox family members confers the unique role of Sox2 in mouse ES cells. *BMC Evol. Biol.*, **16**, 173.

55. Strom,A.R. and Brangwynne,C.P. (2019) The liquid nucleome – phase transitions in the nucleus at a glance. *J Cell Sci*, **132**.

56. Shrinivas,K., Sabari,B.R., Coffey,E.L., Klein,I.A., Boija,A., Zamudio,A.V., Schuijers,J., Hannett,N.M., Sharp,P.A., Young,R.A., *et al.* (2019) Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell*, **75**, 549-561.e7.

57. Zamudio,A.V., Dall'Agnese,A., Henninger,J.E., Manteiga,J.C., Afeyan,L.K., Hannett,N.M., Coffey,E.L., Li,C.H., Oksuz,O., Sabari,B.R., *et al.* (2019) Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol. Cell*, 10.1016/j.molcel.2019.08.016.

58. Gutiérrez,J.L., Chandy,M., Carrozza,M.J. and Workman,J.L. (2007) Activation domains drive nucleosome eviction by SWI/SNF. *The EMBO Journal*, **26**, 730–740.

59. King,H.W. and Klose,R.J. (2017) The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife*, **6**, e22631.

60. Suzuki,H.I., Young,R.A. and Sharp,P.A. (2017) Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell*, **168**, 1000-1014.e15.

61. Sandholzer,J., Hoeth,M., Piskacek,M., Mayer,H. and de Martin,R. (2007) A novel 9-amino-acid transactivation domain in the C-terminal part of Sox18. *Biochem. Biophys. Res. Commun.*, **360**, 370–374.

62. Berta,P., Hawkins,J.R., Sinclair,A.H., Taylor,A., Griffiths,B.L., Goodfellow,P.N. and Fellous,M. (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448–450.

63. Feng,C.-W.A., Spiller,C., Merriner,D.J., O'Bryan,M.K., Bowles,J. and Koopman,P. (2017) SOX30 is required for male fertility in mice. *Sci Rep*, **7**, 17619.

64. Ito,S., Yamane,M., Ohtsuka,S. and Niwa,H. (2014) The C-terminal region of Xpc is dispensable for the transcriptional activity of Oct3/4 in mouse embryonic stem cells. *FEBS Lett.*, **588**, 1128–1135.

65. McDowall,S., Argentaro,A., Ranganathan,S., Weller,P., Mertin,S., Mansour,S., Tolmie,J. and Harley,V. (1999) Functional and structural studies of wild type SOX9 and mutations causing campomelic dysplasia. *J. Biol. Chem.*, **274**, 24023–24030.

66. Roumaud,P., Haché,J. and Martin,L.J. (2018) Expression profiles of Sox transcription factors within the postnatal rodent testes. *Mol Cell Biochem*, 10.1007/s11010-018-3302-3.

67. Cowling,V.H. and Cole,M.D. (2006) Mechanism of transcriptional activation by the Myc oncoproteins. *Semin. Cancer Biol.*, **16**, 242–252.

68. Barrett,J.F., Lee,L.A. and Dang,C.V. (2005) Stimulation of Myc transactivation by the TATA binding protein in promoter-reporter assays. *BMC Biochem.*, **6**, 7.

69. Kato,G.J., Barrett,J., Villa-Garcia,M. and Dang,C.V. (1990) An amino-terminal c-myc domain required for neoplastic transformation activates transcription. *Mol. Cell. Biol.*, **10**, 5914–5920.

70. Stone,J., de Lange,T., Ramsay,G., Jakobovits,E., Bishop,J.M., Varmus,H. and Lee,W. (1987) Definition of regions in human c-myc that are involved in transformation and nuclear localization. *Mol Cell Biol*, **7**, 1697–1709.

71. Barrett,J., Birrer,M.J., Kato,G.J., Dosaka-Akita,H. and Dang,C.V. (1992) Activation domains of L-Myc and c-Myc determine their transforming potencies in rat embryo cells. *Mol. Cell. Biol.*, **12**, 3130–3137.

72. King,N., Westbrook,M.J., Young,S.L., Kuo,A., Abedin,M., Chapman,J., Fairclough,S., Hellsten,U., Isogai,Y., Letunic,I., *et al.* (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, **451**, 783–788.

73. Goto,N.K., Zor,T., Martinez-Yamout,M., Dyson,H.J. and Wright,P.E. (2002) Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J. Biol. Chem.*, **277**, 43168–43174.

74. Krois,A.S., Ferreon,J.C., Martinez-Yamout,M.A., Dyson,H.J. and Wright,P.E. (2016) Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1853-1862.

Online prediction for activation domains

The **9aaTAD prediction** result for the **Pou5F / Oct4** activator
less stringent pattern resulted in a single perfect match

online 9aaTAD prediction service
www.med.muni.cz/9aaTAD

Sequence	Start	End	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	RC11	RC12	= Rules for amino acid clustering
GHLASDFAF	3	11	+	+	+	+	+	+	+	+	+	+	+	+	perfect match

The **9aaTAD prediction** result for the **Sox2** activator
less stringent pattern, cut off 80% match

Sequence	Start	End	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	RC11	RC12	= Rules for amino acid clustering
MNAFMVWSR	45	53	+	+	-	+	+	-	+	+	+	+	+	+	83% match
DMISMYLPG	272	280	+	+	+	+	-	-	+	+	+	+	+	+	83% match

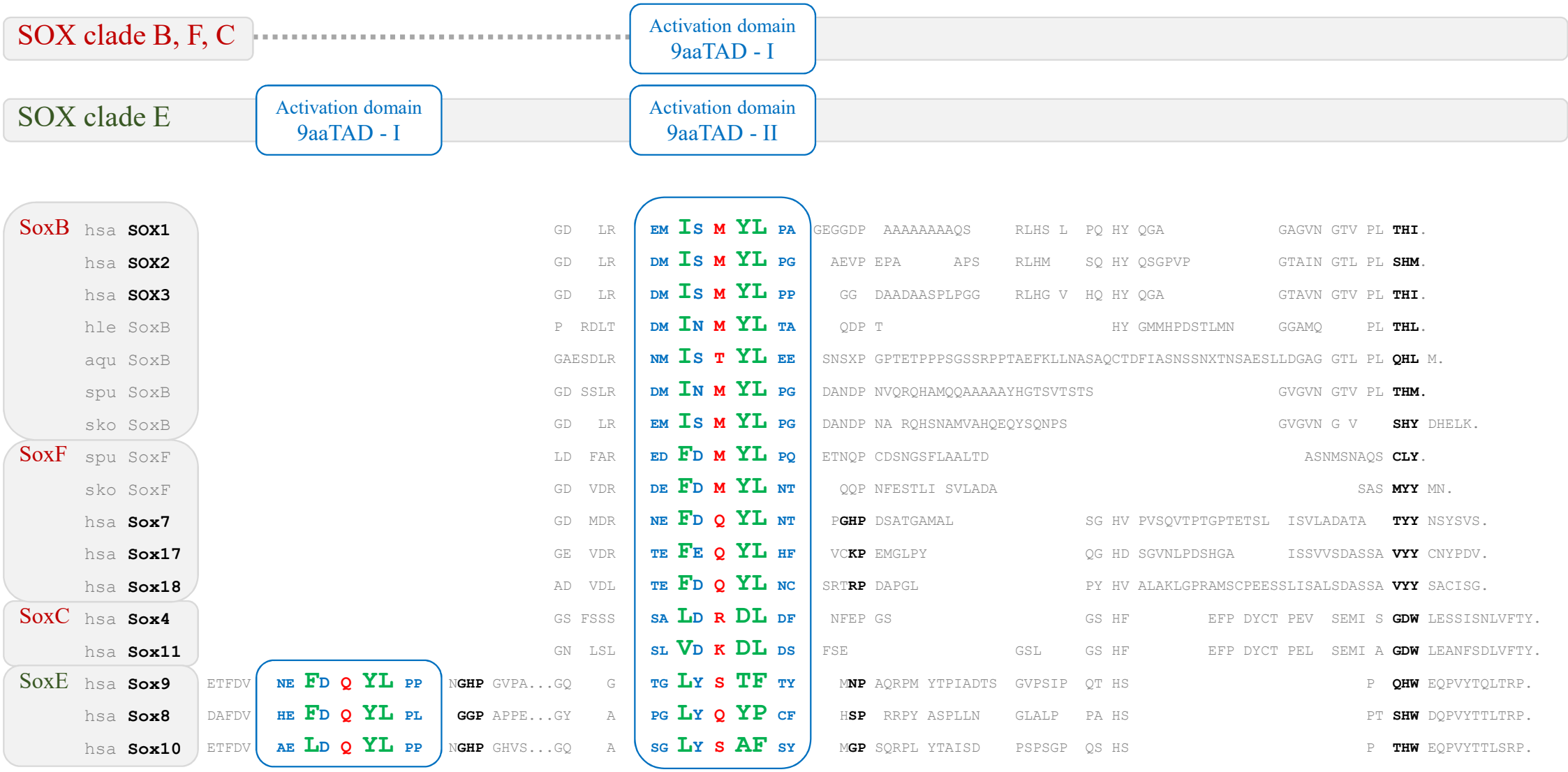
The **9aaTAD prediction** result for the **c-Myc** trans-activating region 1-120
less stringent pattern, cut off 80% match

Sequence	Start	End	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	RC11	RC12	= Rules for amino acid clustering
MPLNVSFTN	1	9	+	-	+	+	+	+	+	+	+	+	+	-	83% match
EMVTELLGG	100	108	+	+	+	+	+	-	+	+	+	+	+	+	92% match

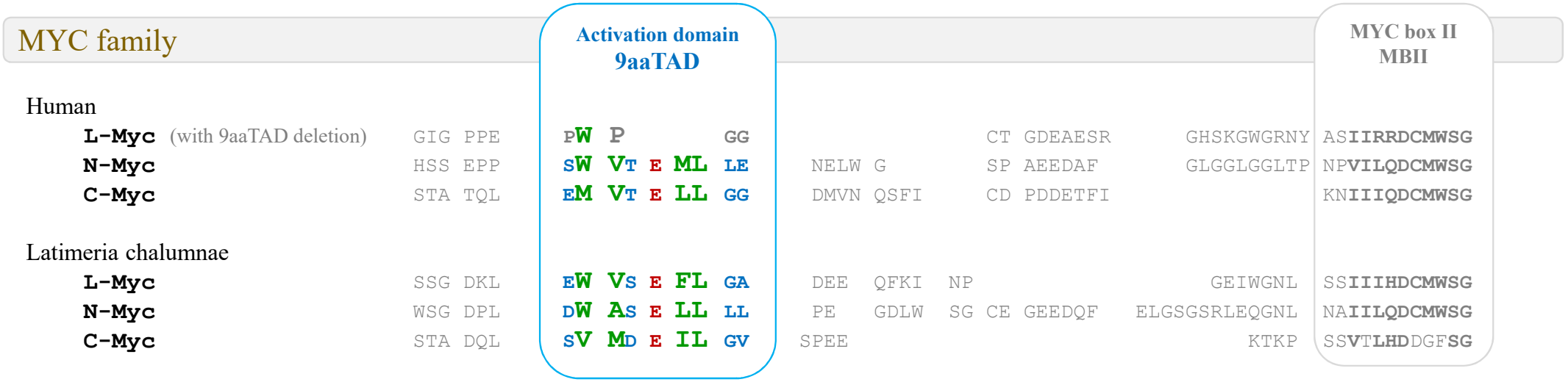
The **9aaTAD prediction** result for the **KLF4** activator
most stringent pattern resulted in a single perfect match

Sequence	Start	End	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	RC11	RC12	= Rules for amino acid clustering
DLLDLDFIL	101	109	+	+	+	+	+	+	+	+	+	+	+	+	perfect match

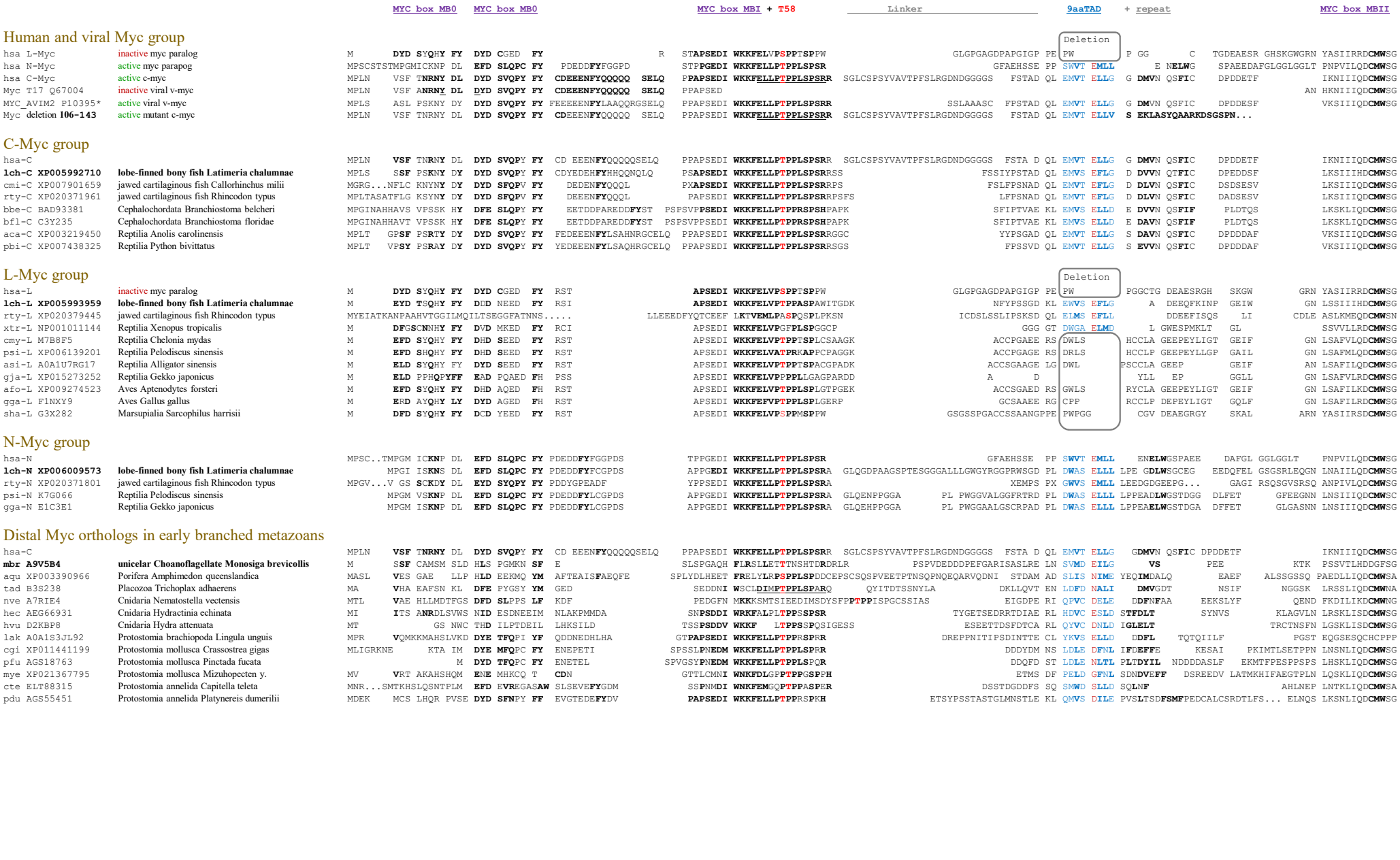
Transactivation assay		peptides used in assay		ID	Position	Reporter activation
		9aaTAD				
Pou1 clade	hsa **Pit1**	MSCQ AF TS A DT FI PLN		**po1**	1-16	30 ±04
	cmk Pit1	MHL vS VE E LR AT VK		**po2**	1-14	1 ±01
Pou5 clade	hsa **Pou5f1/Oct4**	MAG HL AS D FA FS		**po6**	1-12	21 ±04
	cur Pou5	MAG HL SS D LA FS		**po8**	1-12	12 ±02
	shr Pou5	MAG HL AP E YF SP		**po7**	1-12	9 ±01
Pou predecessor	lak Pou	D HL SD M KY MS AHN		**po11**	37-49	0 ±01
SoxB clade	hsa **Sox1-3**	DLR DM IS M YL PP		**hS3**	396-407	217 ±10
SoxC clade	hsa **Sox4**	sA LD R DL DF NFE		**hS4**	426-437	114 ±11
SoxE/F clade **(TADI)**	hsa **Sox17**	VDR tE FE Q YL HF		**hS17**	358-371	111 ±02
SoxE clade **(TADII)**	hsa **Sox9**	tG LY S TF TY MNP		**hS9**	461-471	46 ±04
	hsa **Sox8**	pG LY Q YP CF HSP		**hS8**	401-411	83 ±06
c-Myc clade	hsa **c-Myc**	QL EM VT E LL GG		**hsa-C**	98-108	62 ±09
	lch c-Myc	QL EM VS E FL GD		**lch-C**	79-89	60 ±08
N-Myc clade	hsa **N-Myc**	sW VT E ML LE		**hsa-N**	76-84	128 ±14
	lch N-Myc	L DW AS E LL LL P		**lch-N**	94-104	125 ±20
L-Myc clade	lch L-Myc	KL EW VS E FL GA		**lch-L**	59-69	174 ±44
Myc predecessors	aqu Myc	D sL IS N IM EY E		**aqu**	99-109	12 ± 1
	mbe Myc	LN sV MD E IL GV S		**mbr**	74-85	10 ± 1
Control constructs	p53 - 9aaTAD	LSP DD IE Q WF TE DP		**6p53**		set to 100 ±11
	empty vector	TM DD		**Hdd**		0 ±01


SOX clade B, F, C
Activation domain
9aTAD - I
SOX clade E
Activation domain
9aTAD - I
Activation domain
9aTAD - II
SoxB
hsa SOX1 GD LR EM IS M YL PA GEGGDP AAAAAAAAQS RLHS L PQ HY QGA GAGVN GTV PL THI.
hsa SOX2 GD LR DM IS M YL PG AEVP EPA APS RLHM SQ HY QSGPVP GTAIN GTL PL SHM.
hsa SOX3 GD LR DM IS M YL PP GG DAADAASPLPGG RLHG V HQ HY QGA GTAVN GTV PL THI.
hle SoxB P RDLT DM IN M YL TA QDP T HY GMMHPDSTLMN GGAMQ PL THL.
aqu SoxB GAESDLR NM IS T YL EE SNSXP GPTETPPPSGSSRPPTAEFKLLNASAQCTDFIASNSSNXTNSAESLLDGAG GTL PL QHL M.
spu SoxB GD SSLR DM IN M YL PG DANDP NVQRQHAMQQAAAAAYHGTSVTSTS GVGVN GTV PL THM.
sko SoxB GD LR EM IS M YL PG DANDP NA RQHSNAMVAHQEQYSQNPS GVGVN G V SHY DHELK.
SoxF
spu SoxF LD FAR ED FD M YL PQ ETNQP CDSNGSFLAALTD ASNMSNAQS CLY.
sko SoxF GD VDR DE FD M YL NT QQP NFESTLI SVLADA SAS MYY MN.
hsa Sox7 GD MDR NE FD Q YL NT PGHP DSATGAMAL SG HV PVSQVTPTGPTETSL ISVLADATA TYY NSYSVS.
hsa Sox17 GE VDR TE FE Q YL HF VCKP EMGLPY QG HD SGVNLPDSHGA ISSVVSDASSA VYY CNYPDV.
hsa Sox18 AD VDL TE FD Q YL NC SRTRP DAPGL PY HV ALAKLGPRAMSCPEESSLISALSDASSA VYY SACISG.
SoxC
hsa Sox4 GS FSSS SA LD R DL DF NFEP GS GS HF EFP DYCT PEV SEMI S GDW LESSISNLVFTY.
hsa Sox11 GN LSL SL VD K DL DS FSE GSL GS HF EFP DYCT PEL SEMI A GDW LEANFSDLVFTY.
SoxE
hsa Sox9 ETFDV NE FD Q YL PP NGHP GVPA...GQ G TG LY S TF TY MNP AQRPM YTPIADTS GVPSIP QT HS P QHW EQPVYTQLTRP.
hsa Sox8 DAFDV HE FD Q YL PL GGP APPE...GY A PG LY Q YP CF HSP RRPY ASPLLN GLALP PA HS PT SHW DQPVYTTLTRP.
hsa Sox10 ETFDV AE LD Q YL PP NGHP GHVS...GQ A SG LY S AF SY MGP SQRPL YTAISD PSPSGP QS HS P THW EQPVYTTLSRP.

4A


MYC family
Activation domain
9aaTAD
MYC box II
MBII
Human
L-Myc (with 9aaTAD deletion) GIG PPE pW P GG CT GDEAESR GHSKGWGRNY ASIIRRDCMWSG
N-Myc HSS EPP sW VT E ML LE NELW G SP AEEDAF GLGGLGGLTP NPVILQDCMWSG
C-Myc STA TQL eM VT E LL GG DMVN QSFI CD PDDETFI KNIIIQDCMWSG
Latimeria chalumnae
L-Myc SSG DKL eW VS E FL GA DEE QFKI NP GEIWGNL SSIIIHDCMWSG
N-Myc WSG DPL dW AS E LL LL PE GDLW SG CE GEEDQF ELGSGSRLEQGNL NAIILQDCMWSG
C-Myc STA DQL sV MD E IL GV SPEE KTKP SSVTLHDDGFSG

4B


MYC box MB0
MYC box MB0
MYC box MBI + T58
Linker
9aaTAD
+ repeat
MYC box MBII
Deletion
Human and viral Myc group
hsa L-Myc inactive myc paralog
hsa N-Myc active myc parapog
hsa C-Myc active c-myc
Myc T17 Q67004 inactive viral v-myc
MYC_AVIM2 P10395* active viral v-myc
Myc deletion 106-143 active mutant c-myc
C-Myc group
hsa-C
lch-C XP005992710 lobe-finned bony fish Latimeria chalumnae
cmi-C XP007901659 jawed cartilaginous fish Callorhinchus milii
rty-C XP020371961 jawed cartilaginous fish Rhincodon typus
bbe-C BAD93381 Cephalochordata Branchiostoma belcheri
bfl-C C3Y235 Cephalochordata Branchiostoma floridae
aca-C XP003219450 Reptilia Anolis carolinensis
pbi-C XP007438325 Reptilia Python bivittatus
L-Myc group
hsa-L inactive myc paralog
lch-L XP005993959 lobe-finned bony fish Latimeria chalumnae
rty-L XP020379445 jawed cartilaginous fish Rhincodon typus
xtr-L NP001011144 Reptilia Xenopus tropicalis
cmy-L M7B8F5 Reptilia Chelonia mydas
psi-L XP006139201 Reptilia Pelodiscus sinensis
asi-L A0A1U7RG17 Reptilia Alligator sinensis
gja-L XP015273252 Reptilia Gekko japonicus
afo-L XP009274523 Aves Aptenodytes forsteri
gga-L F1NXY9 Aves Gallus gallus
sha-L G3X282 Marsupialia Sarcophilus harrisii
N-Myc group
hsa-N
lch-N XP006009573 lobe-finned bony fish Latimeria chalumnae
rty-N XP020371801 jawed cartilaginous fish Rhincodon typus
psi-N K7G066 Reptilia Pelodiscus sinensis
gga-N E1C3E1 Reptilia Gekko japonicus
Distal Myc orthologs in early branched metazoans
hsa-C
mbr A9V5B4 unicelar Choanoflagellate Monosiga brevicollis
aqu XP003390966 Porifera Amphimedon queenslandica
tad B3S238 Placozoa Trichoplax adhaerens
nve A7RIE4 Cnidaria Nematostella vectensis
hec AEG66931 Cnidaria Hydractinia echinata
hvu D2KBP8 Cnidaria Hydra attenuata
lak A0A1S3JL92 Protostomia brachiopoda Lingula unguis
cgi XP011441199 Protostomia mollusca Crassostrea gigas
pfu AGS18763 Protostomia mollusca Pinctada fucata
mye XP021367795 Protostomia mollusca Mizuhopecten y.
cte ELT88315 Protostomia annelida Capitella teleta
pdu AGS55451 Protostomia annelida Platynereis dumerilii

Graphical abstract