

Supplementary information

Integrative epigenetic and molecular analysis reveals a novel promoter for a new isoform of the transcription factor TEAD4

Shima Rashidiani ¹, Gizaw Mamo ¹, Benjámín Farkas ¹, András Szabadi^{1,2}, Bálint Farkas ^{3,4}, Veronika Uszkai ³, András Császár ³, Barbara Brandt ⁵, Kálmán Kovács ^{3,4}, Marianna Pap ^{5,#} and Tibor A. Rauch ^{1,#,*}

¹ Institute of Biochemistry and Medical Chemistry, Medical School, University of Pécs; 7624 Pécs, Hungary

² Department of Dentistry, Oral and Maxillofacial Surgery, Medical School, University of Pécs, 7623 Pécs, Hungary

³ Department of Obstetrics and Gynecology, Medical School, University of Pécs, 7624 Pécs, Hungary;

⁴ National Laboratory of Human Reproduction, University of Pécs, 7624 Pécs, Hungary

⁵ Department of Medical Biology and Central Electron Microscope Laboratory, Medical School, University of Pécs; 7624 Pécs, Hungary

* Correspondence: tibor.rauch@aok.pte.hu

Shared authors

Supplementary information

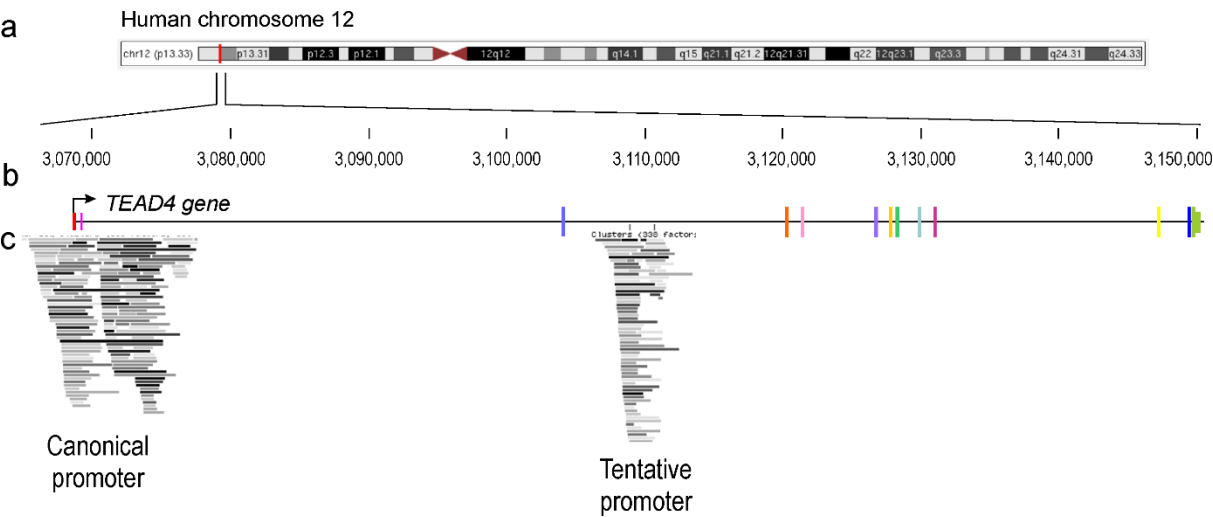


Figure S1. The transcription factor ChIP-Seq data in the TEAD4 locus. (a) The TEAD4 gene is located on the left arm of human chromosome 12, as indicated by the vertical red line. The exon/intron structure of the TEAD4 gene is shown in (b) with colored vertical lines representing exons and thin horizontal lines representing intronic regions. (c) The location of ChIP-Seq data for specific transcription factors is represented by gray horizontal lines.

Supplementary information

>TEAD4-ΔN

CCAGCCCTCCCCTTGAGCACCTCTTACCCTCCTGGGCCGGCTCCCGAGCCCGGGGTGCTTGCCTTCTCCTGCT
CACGGCCGCTTTCATTTCTGCCCTGTCACTGTGTGACCCTCCCCCTGGCACAGCCAACGCTGGCCACCCTGAC
CTCCTTTGGCCAGGCTCACAGTCGGCCTAGCCTAGCGTGCAGTCGGAACGAGCTGATTGCCCGCTACATCAA
GCTCCGGACAGGGAAGACCCGCACCAGGAAGCAGGTCTCCAGCCACATCCAGGTGCTGGCTCGTCGCAAAGCT
CGCGAGATCCAGGCCAAGCTAAAGGACCAGGCAGCTAAGGACAAGGCCCTGCAGAGCATGGCTGCCATGTCGT
CTGCACAGATCATCTCCGCCACGGCCTTCCACAGTAGCATGGCCCTCGCCCGGGGCCCCGGCCGCCAGCAGT
CTCAGGGTTTGGCAAGGAGCTTTGCCAGGCCAAGCCGGAACGTCCCATGATGTGAAGCCTTTCTCTCAGCAA
ACCTATGCTGTCCAGCCTCCGCTGCCTCTGCCAGGGTTTGAGTCTCCTGCAGGGCCCCGCCCATCGCCCTCTG
CGCCCCCGGCACCCCCATGGCAGGGCCGCAGCGTGGCCAGCTCCAAGCTCTGGATGTTGGAGTTCTCTGCCTT
CCTGGAGCAGCAGCAGGACCCGGACACGTACAACAAGCACCTGTTTCGTGCACATTGGCCAGTCCAGCCCAAGC
TACAGCGACCCCTACCTCGAAGCCGTGGACATCCGCCAAATCTATGACAAATTCCCGGAGAAAAAGGGTGGAC
TCAAGGATCTCTTCGAACGGGGACCCTCCAATGCCTTTTTTCTTGTGAAGTTCTGGGCAGACCTCAACACCAA
CATCGAGGATGAAGGCAGCTCCTTCTATGGGGTCTCCAGCCAGTATGAGAGCCCCGAGAACATGATCATCACC
TGCTCCACGAAGGTCTGCTCTTTTCGGCAAGCAGGTGGTGGAGAAAGTTGAGACAGAGTATGCTCGCTATGAGA
ATGGACACTACTCTTACCGCATCCACCGGTCCCCGCTCTGTGAGTACATGATCAACTTCATCCACAAGCTCAA
GCACCTCCCTGAGAAGTACATGATGAACAGCGTGTGGAGAACTTCACCATCCTGCAGGTGGTCACCAACAGA
GACACACAGGAGACCTTGCTGTGCATTGCCTATGTCTTTGAGGTGTCAGCCAGTGAGCACGGGGCTCAGCACC
ACATCTACAGGCTGGTGAAAGAATGAGAGACTCGGGGAGCAGGGAGGGGGGAAG

Figure S2. The primary nucleotide sequence of the mRNA encoding the truncated TEAD4 isoform. The sequence contains alternating exons, indicated by yellow and gray shading. The novel (non-coding) exon sequence is underlined in red, while the open reading frame (ORF) is marked in bold.

Supplementary information

The Sequence Manipulation Suite: Multiple Align Show

```

TEAD4_FL      GTCCTCCCTCGCACACTCGAGGCCAGGGGGCGGGAGGGCGGCGCTCCGGTCCGCGCGCGCTCCGCGACAAAGATCGCCGGGCGCCGCAACA 90
TEAD4-N      CCAGGCCCTCC-----CCTTGAG-----CACCTCTTACC-----CTCCTGCGCGGCTCCCGAGCCCGGGTG 87

TEAD4_FL      GTTGGCGGTCGGGGCGGACTCCTTGGAACTGGCTTAGCGGCAACCATCCACCTTCCCGCACCTGGGACCGGTCCAAAGAGCGCTCCTCG 180
TEAD4-N      G-----TGG-----CCTT-----TTCCTG----- 71

TEAD4_FL      AAGCGGAGCCTTGGAGGGCAGGCGCGGCACCATTAACCTCCAAACAGTGGAGCTCTCCACCTCCCTGAGGGGAGCACCGCCTCTGGGG 270
TEAD4-N      -----CTCAGGCGCGCTT-----T----- 85

TEAD4_FL      CAGTCAGGCACTGGACAAAGCCCATCGACCAATGACCGACAGGGCGTGTGAGCCCGGATATTGAGCAGAGTTTCAGGAGGCCCTCGC-CA 359
TEAD4-N      -----CATTTCTGCGCTGCACTGTGTGACCTTCCCGCTG-GCACAGC---CAA---CGCTGGCCAG 140

TEAD4_FL      TCTACCCGCCCTGTGGCAGGCGCAAAATCATCCTCTCGGACGAGGGCAAGATGTATGGTCGGAACGAGCTGATTGCCCGCTACATCAAGC 449
TEAD4-N      CCTGACCTCCTTTGGCCAGGCTCACAGTGGCTAG-----CCTAG-CGTGCAAGTCGGAACGAGCTGATTGCCCGCTACATCAAGC 221

TEAD4_FL      TCCGGACAGGGAAGACCCGACACAGGAAGCAGGTCTCCAGCCACATCCAGGTGCTGGCTCGTGGCAAAGCTCGCGAGATCCAGGCCAAGC 539
TEAD4-N      TCCGGACAGGGAAGACCCGACACAGGAAGCAGGTCTCCAGCCACATCCAGGTGCTGGCTCGTGGCAAAGCTCGCGAGATCCAGGCCAAGC 311

TEAD4_FL      TAAAGGACAGGCGAGCTAAGGACAAAGGCCCTGCAGAGCATGGCTGCCATGTGCTCTGCACAGATCATCTCCGCCACGGCCTTCCACAGTA 629
TEAD4-N      TAAAGGACAGGCGAGCTAAGGACAAAGGCCCTGCAGAGCATGGCTGCCATGTGCTCTGCACAGATCATCTCCGCCACGGCCTTCCACAGTA 401

TEAD4_FL      GCATGGCCCTCGCCCGGGGGCCCGCGCCGCCAGCAGTCTCAGGGTTTTGGCAAGGAGCTTTGCCAGGCCAAGCCGGAACGTCCCATGATG 719
TEAD4-N      CATGGCCCTCGCCCGGGGGCCCGCGCCGCCAGCAGTCTCAGGGTTTTGGCAAGGAGCTTTGCCAGGCCAAGCCGGAACGTCCCATGATG 491

TEAD4_FL      TGAAGCCTTTTCTCTCAGCAAAACCTATGCTGTCCAGCCTCCCGCTGCCCTTGCAGGGTTTGAAGTCTCCTGCAAGGGCCCGCCCATCGCCCT 809
TEAD4-N      TGAAGCCTTTTCTCTCAGCAAAACCTATGCTGTCCAGCCTCCCGCTGCCCTTGCAGGGTTTGAAGTCTCCTGCAAGGGCCCGCCCATCGCCCT 581

TEAD4_FL      CTGCGCCCCCGGCACCCCCATGGCAGGGCCGCGAGCTGGCCAGCTCCAAAGCTCTGGATGTTGGAGTTCTCTGCCCTTCTGGAGCAGCAGC 899
TEAD4-N      CTGCGCCCCCGGCACCCCCATGGCAGGGCCGCGAGCTGGCCAGCTCCAAAGCTCTGGATGTTGGAGTTCTCTGCCCTTCTGGAGCAGCAGC 671

TEAD4_FL      AGGACCCGGACAGCTACAACAAGCACTGTTGCTGCACATTGGCCAGTCCAGCCCAAGCTACAGCGACCCCTACCTCGAAGCCGTGGACA 989
TEAD4-N      AGGACCCGGACAGCTACAACAAGCACTGTTGCTGCACATTGGCCAGTCCAGCCCAAGCTACAGCGACCCCTACCTCGAAGCCGTGGACA 761

TEAD4_FL      TCCGCCAAATCTATGACAAATTCCCGGAGAAAAAGGGTGGACTCAAGGATCTCTTCGAACGGGGACCCCTCCAATGCCCTTTTTTCTGTGTA 1079
TEAD4-N      TCCGCCAAATCTATGACAAATTCCCGGAGAAAAAGGGTGGACTCAAGGATCTCTTCGAACGGGGACCCCTCCAATGCCCTTTTTTCTGTGTA 851

TEAD4_FL      AGTTCTGGGCGAGACCTCAACACCAACATCGAGGATGAAGGCAGTCCCTTCTATGGGGTCTCCAGCCAGTATGAGAGCCCGAGAACATGA 1169
TEAD4-N      AGTTCTGGGCGAGACCTCAACACCAACATCGAGGATGAAGGCAGTCCCTTCTATGGGGTCTCCAGCCAGTATGAGAGCCCGAGAACATGA 941

TEAD4_FL      TCATCACCTGCTCCACGAAGGTCTGCTCTTTCCGCAAGCAGGTGGTGGAGAAAGTTGAGACAGAGTATGCTCGCTATGAGAATGGACACT 1259
TEAD4-N      TCATCACCTGCTCCACGAAGGTCTGCTCTTTCCGCAAGCAGGTGGTGGAGAAAGTTGAGACAGAGTATGCTCGCTATGAGAATGGACACT 1031

TEAD4_FL      ACTCTTACCGCATCCACCGGTCCCGCTCTGTGAGTACATGATCAACTTCATCCACAAGCTCAAGCACTCCCTGAGAAGTACATGATGA 1349
TEAD4-N      ACTCTTACCGCATCCACCGGTCCCGCTCTGTGAGTACATGATCAACTTCATCCACAAGCTCAAGCACTCCCTGAGAAGTACATGATGA 1121

TEAD4_FL      ACAGCGTGTGGAGAACTTCACCATCCTGCAGGTGGTCAACCAACAGAGACACACAGGAGACCTTGTGTGCAATTGCCCTATGTCTTTGAGG 1439
TEAD4-N      ACAGCGTGTGGAGAACTTCACCATCCTGCAGGTGGTCAACCAACAGAGACACACAGGAGACCTTGTGTGCAATTGCCCTATGTCTTTGAGG 1211

TEAD4_FL      TGTAGCCAGTGAGCACGGGGCTCAGCACCAACATCTACAGGCTGGTGAAGAAATGAGAGACTCGGGGAGCAGGGAGGGGGGAAGAGACGT 1529
TEAD4-N      TGTAGCCAGTGAGCACGGGGCTCAGCACCAACATCTACAGGCTGGTGAAG----- 1260

TEAD4_FL      GTGTGCAGGAAACGGGGACGTGGGGAGGGGACCTGCAGGGGCAGCCCTCTGAAGTGCCAAAGAGAGCTGAGAGGAGCAGTTGTGACTTAC 1619
TEAD4-N      -----A----- 1261

TEAD4_FL      CAGGAACAACTGTGCTTGAACC TGAAGGTGCCCAACCCCAATTAACCCAACTGCTGTGATTTTCAGA 1690
TEAD4-N      GAA-----TGA-----SAGACTGGCGAGCGAGGAGGGGGGAG 1295

```

Figure S3.
Pairwise

alignment of mRNAs encoding TEAD4 isoforms. The nucleotide sequences of the full-length coding transcript of TEAD4 and the newly identified TEAD4-ΔN isoform encoding transcript are aligned using . The nucleotides highlighted in red are identical residues.

Supplementary information

Analysis	Name	Forward/Reverse	Notes
5'RACE	Anchor_F1	GACCACGCGTATCGATGTCGACTTTTT TTTTTTTTTTTV	
	Anchor_F2	GTATCGATGTCGACTTTTTTTTTT	
	GenSpec_RT	ACTCAAACCCTGGCAG	Reverse transcription
	GenSpec_R1	CATGGGACGTTCCGGC	
	GenSpec_R2	AAGCTCCTTGCCAAAAC	
Gene expr.	TEAD4-C_F	TGAGCAGAGTTTCCAGGAGG	End-point PCR
	TEAD4-C_R	CCTTGCCAAAACCCTGAGAC	
	TEAD4-del_F	TCCCCTTGAGCACCTCTTAC	End-point PCR
	TEAD4-del_R	TCTCAGGGTTTTGGCAAGGA	
Promoter cloning	TEAD- PromC_oF	TCATGCACTTCCTATCCCA	Nested PCR
	TEAD- PromC_oR	TGAGGGCTGGAATGAGATGC	
	TEAD- PromC_InF	aactcgagCCACCCAACACAAGGCTAAG	
	TEAD- PromC_InR	aaaagcttGGCTGGAATGAGATGCGC	
	TEAD- PromAlt_oF	TCTCTTTGGCAGGAGGGAAG	Nested PCR
	TEAD- PromAlt_oR	CAGGAGGGTAAGAGGTGCTC	
	TEAD- PromAlt_inF	aactcgagCAGTCTGCTTCCCTGAGTCA	
	TEAD- PromAlt_inR	aaaagcttGTTAATCTGCCGTGCTGCTT	
	Prom-Seq_F	gtacgggagggtacttgagcg	Sanger Seq
	Prom-Seq_R	ccacctgatatgtcatctgtaaaa	
Fusing cDNAs to FP	TEAD4-EGFP_F	aaaaagcttGCCATGGAGGGCACGGC	TTG mutated to ATG to increase translation
	TEAD4-EGFP_R	aaggatccTTCTTTCACCAGCCTGTA	
	TEAD4-RFP_F	aaaagcttATGGCTGCCATGTCGTCTG	
	TEAD4-RFP_R	aaggatccTTCTTTCACCAGCCTGTA	
	EGFP-Seq_F	CCTCGCCCTCGCCGGA	Sanger Seq
	RFP-Seq_F	GCCCTCGTAGGGCTTGC	
DNA methylation anal.	BS_PromC_F	ATGGTATGTTATATGGATTTTAAAT	
	BS_PromC_R	CAAAAAAACTACTCCCTTCCTAC	
	BS_PromAlt_F	TAGGTTTTGTGGGATTTAGATATT	
	BS_PromAlt_R	CATCCACACACTAAAAATAAACTC	
	T7	GCCAAGCTCTAATACGACTCACTATA GG	Sanger Seq
	Sp6	GCCATTTAGGTGACACTATAGAATAC	

Table S1. Oligonucleotides used in end-point PCR, cloning and DNA methylation (i.e., Bisulfite Sequencing) analysis

Supplementary information

Analysis	Name	Forward/Reverse	Notes
E M S A	TEAD4-ΔN-Bio_F	CCCTCTGGGAGGCGGAATGGGAGGCCGAGCTG	5'-biotyn
	TEAD4-ΔN_F	CCCTCTGGGAGGCGGAATGGGAGGCCGAGCTG	
	TEAD4-ΔN_R	CAGCTCGGCCTCCCATTCGCTCCAGAGGG	
	TEAD4-ΔN-Mut_Bio_F	CCCTCTGGGAGGCGGACGGGGAGGCCGAGCTG	5'-biotyn
	TEAD4-ΔN-Mut_F	CCCTCTGGGAGGCGGACGGGGAGGCCGAGCTG	
	TEAD4-ΔN-Mut_R	CAGCTCGGCCTCCCCGTCCGCTCCAGAGGG	
	TEAD4-cons-F-Bio	CGGCGATGTGACCTGGAATGTGGCGTCCGTAT	5'-biotyn
	TEAD4-cons-F	CGGCGATGTGACCTGGAATGTGGCGTCCGTAT	
	TEAD4-cons-R	ATACGGACGCCACATTCCAGGTCACATCGCCG	

Table S2. Oligonucleotides for DNA-protein interaction studies (EMSAs). The corresponding complementary reverse oligos were annealed to the 5'-biotinylated oligos. In competition assays, un-biotinylated forward were annealed and used in essays with varying molar concentrations.