

Published in final edited form as:

*Biochem Biophys Res Commun.* 2009 September 4; 386(4): 554–558. doi:10.1016/j.bbrc.2009.06.045.

## New TFII-I family target genes involved in embryonic development

**Aleksandr V. Makeyev** and **Dashzeveg Bayarsaihan**\*

Department of Reconstructive Sciences, University of Connecticut Health Center, 262 Farmington avenue, Farmington, CT 06030, USA

### Abstract

Two members of the TFII-I family transcription factor genes, *GTF2I* and *GTF2IRD1*, are the prime candidates responsible for the craniofacial and cognitive abnormalities of Williams syndrome patients. We have previously generated mouse lines with targeted disruption of *Gtf2ird1* and *Gtf2i*. Microarray analysis revealed significant changes in the expression profile of mutant embryos. Here we described three unknown genes that were dramatically down-regulated in mutants. The *2410018M08Rik/Scand3* gene encodes a protein of unknown function with CHCH and hATC domains. *Scand3* is downregulated during mouse embryonic stem cell (ES) differentiation. *4933436H12Rik* is a testis-specific gene, which encodes a protein with no known domains. It is expressed in mouse ES cells. *1110008P08Rik/Kbtbd7* encodes an adapter protein with BTB/POZ, BACK and Kelch motifs, previously shown to recruit substrates to the enzymatic complexes of the histone modifying or E3 ubiquitin ligase activities. Based on the expression pattern *Kbtbd7* may have a specific role in brain development and functioning. All three genes possess well-conserved TFII-I-binding consensus sites within proximal promoters. Therefore our analysis suggests that these genes can be direct targets of TFII-I proteins and their impaired expression, as a result of the *GTF2I* and *GTF2IRD1* haploinsufficiency, could contribute to the etiology of Williams syndrome.

### Keywords

Williams syndrome; TFII-I; transcription factors; *Gtf2ird1*; *Gtf2i*; ES differentiation; microarray

## INTRODUCTION

The *GTF2I* and *GTF2IRD1* paralogs encode two principal members of the TFII-I family of transcription factors. The recent research links the haploinsufficiency of these two genes to the facial dysmorphism and cognitive defects of Williams syndrome (WS; OMIM 194050), the autosomal dominant disorder resulting from the hemizygous deletion of 1.5-1.8 Mb segment on chromosome 7q11.23 [1].

The homozygous loss of either the *Gtf2i* or *Gtf2ird1* function in mice results in similar multiple phenotypic manifestations, including embryonic lethality, brain hemorrhage, vasculogenic, craniofacial, and neural tube defects [2]. In line with the complexity of the

© 2009 Elsevier Inc. All rights reserved

\*Correspondence to: Dashzeveg Bayarsaihan, Email: dashzeveg@uchc.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

observed phenotypes, the following microarray analysis has revealed significant changes in the expression of a wide variety of genes in both *Gtf2i* and *Gtf2ird1* mutants (GEO record GSE4437). A total of 217 and 38 up-regulated and 2,356 and 498 down-regulated genes with more than 1.5-fold difference in expression were identified in the *Gtf2i* and *Gtf2ird1* embryonic arrays, respectively [2]. It is remarkable that a substantial portion of these genes (30 up-regulated and 80 down-regulated) showed similar changes in both arrays. This is consistent with the phenotypic similarity of *Gtf2i* and *Gtf2ird1* mutants.

Among genes expression of which was significantly disturbed by both *Gtf2i* and *Gtf2ird1* knockouts, we have found a few candidates with a well-established relevance to the craniofacial development. Therefore, we believe that the transcriptional regulation of these genes by the TFII-I proteins could provide a genotype-phenotype link in the WS (manuscript in preparation). However, in our previous microarray analysis we did not include EST's that were poorly characterized or known only as full-length transcripts obtained and annotated during RIKEN Mouse Gene Encyclopedia Project. A strong response to the *Gtf2i* or *Gtf2ird1* inactivation suggests that these genes could be new direct targets of the TFII-I family transcriptional regulation. Moreover, we found well-conserved TFII-I-binding sites at the putative promoter regions of *2410018M08Rik/Scand3*, *4933436H12Rik*, and *1110008P08Rik/Kbtbd7* genes. Based on the GEO database expression analysis we suggest that all three genes could be important for embryonic development.

## MATERIALS AND METHODS

The genomic sequences were obtained from the Ensemble project database ([www.ensembl.org](http://www.ensembl.org)). The exon-intron junctions of the *4933436H12Rik* gene were found by a comparison of the RIKEN full-length cDNA (Accession AK017082) with contig of RP24-331B9 (Accession AC116515) and RP23-378L12 (Accession AC161514) clones from mouse chromosome 7.

The putative TFII-I and other transcriptional factor binding sites were searched as pattern consensus on both strands within -5,000 to +200 genomic regions using MacVector 7.2 (Oxford Molecular Group). The statistical analysis was performed using InStat 3.0 (GraphPad Software).

The time-course of gene expression in mouse R1 embryonic stem cells differentiating into embryoid bodies *in vitro* was analyzed using datasets from Gene Expression Omnibus (GEO) repository ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/), GDS2667 record) and molecular probes 1429671\_at (*2410018M08Rik/Scand3*) and 1432036\_at (*4933436H12Rik*). The molecular probe 1452700\_s\_at was used for analysis of *Kbtbd7* gene expression in mouse brain regions (GEO GDS2917 record).

## RESULTS AND DISCUSSION

### 2410018M08Rik/Scand3

The *2410018M08Rik* transcript (GenBank Accession AK010551) was down-regulated more than 20-fold in both *Gtf2i* and *Gtf2ird1* mutants. This gene has also been assigned the name *SCAN domain containing 3* (*Scand3*) and received the Entrez and MGI ID numbers 71970 and 1919220, respectively. Mouse *Scand3* consists of 4 exons (Fig. 1A) and was mapped to cytoband 5G1.3. *Scand3* is located at the border between two syntenic regions to human 12q24 and 7p11 and has no obvious human ortholog. The human *SCAND3* gene (GeneID 114821) has 44.5% and 58% similarity to mouse *2410018M08Rik* sequence at the nucleotide and amino acid levels, respectively, and possesses different exon-intron and domain structures. Therefore, it is unlikely that human *SCAND3* has orthologous relationship to

mouse *2410018M08Rik* gene. In addition to the functional *2410018M08Rik/Scand3* gene, we have identified processed pseudogenes at 1p36.21, 8q11.23, 9q21.31, and 19q13.11 (data not shown).

Mouse *2410018M08Rik/Scand3* encodes 756 aa (84.3 kDa) protein (Fig. 1A) which despite of its name (assigned for reason of similarity to the human *SCAND3* gene) does not have the SCAN domain but harbors two other conserved domains: the C-terminal coiled-coil-helix-coiled-coil-helix motif (CHCH, pfam06747) and the N-terminal hAT family dimerisation motif (hATC, pfam05699) [3,4].

Within the 2 kb region upstream from the *Scand3* putative transcriptional start site, we have found two TFII-I binding consensus BRGATTRBR sequences (Fig. 1B), which were deduced from the SELEX experiments [5] and from the comparison of natural TFII-I binding sites. These putative binding sites are complemented by multiple E-boxes that can also serve as potential TFII-I binding sites [6]. In addition, this genomic region contains cAMP-responsive elements and well preserved consensus binding sites for SMAD, STAT, p300, and NFkB transcription factors (Fig. 1B).

The EST sequences of *Scand3* were found in many embryonic and adult mouse tissues including pituitary gland, thymus, testis, mammary gland, eye, brain, liver, and ovary (UniGene Mm.422716), however in terms of the transcript abundance, the large scale transcriptome analysis using massive parallel signature sequencing technology [7] showed the more limited expression (hippocampus, kidney cortex, lung, female white fat and ovary) (GDS868 record).

During mouse embryonic development, *Scand3* is already expressed in the inner cell mass of blastocysts (GDS814 record). Another interesting observation is that the *Scand3* expression decreases dramatically during R1 embryonic stem (ES) cell differentiation into embryoid bodies *in vitro* [8] which similar to expression pattern of pluripotency marks such as *Nanog* and *Pou5f1* (Fig. 1C, GDS2667 record). Similar results were obtained for the mouse V6.5 ES cells (GDS2672 record) and mouse ES J1 cells (GDS2669 and GDS2906 records). These findings indicate that *Scand3* could be involved in the self-renewal of mouse embryonic stem cells.

#### 4933436H12Rik

The *4933436H12Rik* transcript (GenBank Accession AK017082) was also down-regulated 18- and 35-fold in *Gtf2i* and *Gtf2ird1* mutants, respectively. We have determined its exon-intron structure by direct comparison of cDNA (Entrez Gene ID 71244 and MGI:1918494) and mouse genomic sequences (Table 1 and Fig. 2A). We have found that *4933436H12Rik* consists of three exons that can be found in the contig of RP24-331B9 and RP23-378L12 clones and, therefore, this gene is located within the cytoband 7D1 but not within the cytoband 7C as it was mapped previously. The biggest open reading frame of *4933436H12Rik* is embedded into exon 2 (Fig. 2A) and can be translated into putative 110 aa polypeptide, which does not show significant homology with other mouse or human proteins and does not contain any conservative domains.

Search for TFII-I-binding motifs identified two TFII-I consensus sites together with multiple E-boxes within the 2 kb region upstream from the putative transcriptional start of *4933436H12Rik* (Fig. 2B). In addition, this genomic region contains TRA response element and consensus binding sites for SMAD, STAT, p300, and NFkB transcription factors.

The EST sequences of *4933436H12Rik* were found only in testis (UniGene Mm.410787). However, expression of this gene has been also detected in embryonic stem cells

differentiating into embryoid bodies *in vitro*. During two-week time course differentiation of mouse R1 ES cells (GDS2667 record), expression of *4933436H12Rik* gene showed temporal decrease just after induction of differentiation (Fig. 2C). Similar transient changes in *4933436H12Rik* expression were registered during analysis of differentiation of mouse CGR8 ES cells (GDS1882 record).

### 1110008P08Rik/Kbtbd7

*1110008P08Rik* transcript (GenBank Accession AK003597) was down-regulated 12- and 5-fold in *Gtf2i* and *Gtf2ird1* mutants, respectively. This gene is also known as *Kelch repeat and BTB (POZ) domain containing 7 (Kbtbd7)*; Entrez Gene ID 211255; MGI:2685141). It is located at 14D3 and has syntenic human ortholog *KBTD7* at 13q14.11.

Mouse *Kbtbd7* is intronless and encodes 690 aa protein containing three conservative domains: BTB/POZ domain (pfam00651), BACK domain (pfam 07707), and Kelch motif (pfam01344) (Fig. 3A). The BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger) domain is near the N-terminus. The BTP/POZ domain mediates homomeric dimerisation [9]. In some instances, this fold was adapted to several different modes of interactions with non-BTB proteins such as components of the histone deacetylase co-repressor or E3 ubiquitin ligase complexes [10].

The Kelch motif was initially discovered in the Kelch protein which contains six copies of this motif [11]. Many Kelch-repeat proteins are involved in organization of the cytoskeleton via interactions with actin and intermediate filaments. The BACK (for BTB and C-terminal Kelch) domain is present in the majority of proteins that contain both the BTB domain and kelch repeats [12]. The BACK domain is always found juxtaposed to the BTB/POZ domain (they are separated by as little as several residues) and these two domains are likely to cooperate in the protein-protein interactions.

The single TFII-I-binding consensus and multiple E-boxes were identified in the putative promoter region of the *Kbtbd7* gene (Fig. 3B). This genomic region also contains osmotic response elements and consensus binding sites for SMAD, STAT, and p300 transcription factors.

Although EST sequences of *Kbtbd7* were found in a number of mouse embryonic and adult tissues (UniGene Mm.284526), in terms of abundance of transcript signature, the large-scale transcriptome analysis [7] demonstrated highest expression in brain (GDS868 record). It is also true for expression of the human *KBTD7* ortholog (UniGene Hs.63841). Furthermore, analysis of expression profiles of five different brain regions from six inbred strains [13] showed that whereas most of genes displayed the brain region-specific patterns, *Kbtbd7* is among few genes exhibited a significant strain-specific expression pattern (Fig. 3C; GDS2917 record). In this connection it is worth to mention that knockout of inhibitor of protein phosphatase-1 inhibiting pathways involved in learning and memory resulted in the significant up-regulation of *Kbtbd7* in mouse hippocampus (GDS1920 record).

## Conclusions

All three genes show the highest down-regulation in *Gtf2i* and *Gtf2ird1* mutants and contain TFII-I-binding consensus motifs within their proximal promoter regions. Therefore these genes could be novel downstream targets of TFII-I transcription factors. *2410018M08Rik* and *4933436H12Rik* displayed dynamic expression pattern very early during embryonic development including differentiating ES cells. The third gene, *Kbtbd*, has unusual strain-specific expression pattern in mouse brain and is likely involved in learning and memory. Thus, our analysis suggests that impaired expression of these genes, resulting from the

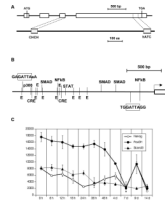
*GTF2IRD1* and *GTF2I* haploinsufficiency, could contribute to the complex Williams syndrome developmental and cognitive phenotype.

## Acknowledgments

This work was supported by the NIH grants K02 DE18412 and R01DE017205 to D.B.

## REFERENCES

- [1]. Tassabehji M. Williams-Beuren syndrome: a challenge for genotype-phenotype correlations. *Hum Mol Genet.* 2003; 12:229–237.
- [2]. Enkhmandakh B, Makeyev AV, Erdenechimeg L, Ruddle FH, Chimge N-O, Tussie-Luna MI, Roy AL, Bayarsaihan D. Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. *Proc. Natl. Acad. Sci. USA.* 2009; 106:181–186. [PubMed: 19109438]
- [3]. Nobrega MP, Bandeira SC, Beers J, Tzagoloff A. Characterization of COX19, a widely distributed gene required for expression of mitochondrial cytochrome oxidase. *J. Biol. Chem.* 2002; 277:40206–40211. [PubMed: 12171940]
- [4]. Essers L, Adolphs RH, Kunze R. A highly conserved domain of the maize activator transposase is involved in dimerization. *Plant Cell.* 2000; 12:211–214. [PubMed: 10662858]
- [5]. Vullhorst D, Buonanno A. Multiple GTF2I-like repeats of general transcription factor 3 exhibit DNA-binding properties: Evidence for a common origin as a sequence-specific DNA interaction module. *J. Biol. Chem.* 2005; 280:31722–31731. [PubMed: 15987678]
- [6]. Roy AL, Du H, Gregor PD, Novina CD, Martinez E, Roeder RG. Cloning of an inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. *EMBO J.* 1997; 16:7091–7104. [PubMed: 9384587]
- [7]. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George G, Eletr E, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 2000; 18:630–634. [PubMed: 10835600]
- [8]. Haillesellasse SK, Porter CJ, Palidwor G, Perez-Iratxeta C, Muro EM, Campbell PA, Rudnicki MA, Andrade-Navarro MA. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics.* 2007; 8:85. [PubMed: 17394647]
- [9]. Bardwell VJ, Treisman R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev.* 1994; 8:1664–1677. [PubMed: 7958847]
- [10]. Stogios PJ, Downs GS, Jauhal JJ, Nandra SK, Prive GG. Sequence and structural analysis of BTB domain proteins. *Genome Biol.* 2005; 6:R82. [PubMed: 16207353]
- [11]. Bork P, Doolittle RF. Drosophila kelch motif is derived from a common enzyme fold. *J. Mol. Biol.* 1994; 236:1277–1282. [PubMed: 8126718]
- [12]. Stogios PJ, Prive GG. The BACK domain in BTB-kelch proteins. *Trends Biochem. Sci.* 2004; 29:634–637. [PubMed: 15544948]
- [13]. Hovatta I, Zapala MA, Broide RS, Schadt EE, et al. DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. *Genome Biol.* 2007; 8:R25. [PubMed: 17324278]
- [14]. Rikitake Y, Moran E. DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol. Cell. Biol.* 1992; 12:2826–2836. [PubMed: 1534143]
- [15]. Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. *Cell Sci.* 2004; 117:5965–5973.
- [16]. Ferraris JD, Garcia-Perez A. Osmotically responsive genes: the mammalian osmotic response element (ORE). *Am. Zool.* 2001; 41:734–742.

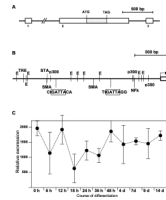


**Fig. 1. Mouse 2410018M08Rik/Scand3 gene**

A. Diagram of the exon-intron organization of the *2410018M08Rik/Scand3* gene and conservative domains of the Scand3 protein. The exons are shown as boxes and drawn to scale (scale bar shown), and the introns, which vary greatly in size, are represented in a uniform manner. CHCH - coiled-coil-helix-coiled-coil-helix motif (pfam06747); hATC - hAT family dimerisation motif (pfam05699).

B. Transcription factor binding sites in the proximal promoter of *Scand3*. The arrow indicates the putative transcription start site. Sequences are shown for the TFII-I-binding consensus BRGATTRBR; the core part is boxed. The relative positions are shown for the following consensus sequences: CRE — cyclic AMP response element (TGASSTCA); NFkB — binding site for activated NFkB complex; p300 - p300 binding site RGGAGT [14]; SMAD - the SMAD family binding site (CAGACA); STAT - STAT binding site (TTCNNGAA). The transcription factor binding sites (TFBS) on plus-strand are shown above each of the diagrams and TFBS on minus-strand are indicated below the diagrams.

C. The time-course of *Nanog*, *Pou5f1(Oct3/4)*, and *Scand3* expression in mouse R1 embryonic stem cells differentiating into embryoid bodies *in vitro* (GDS2667 record). The error bars represent standard deviation calculated from three independent arrays.



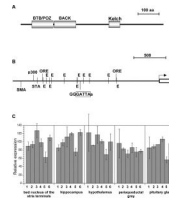
**Fig. 2. Mouse 4933436H12Rik gene**

A. Diagram of the exon-intron organization of 4933436H12Rik. The shaded box represents a putative open reading frame. The other notations are as in Fig. 1A.

B. The transcription factor binding sites (TFBS) in the proximal promoter of 4933436H12Rik. The notation and labeling are as in Fig. 1B. TRE - TPA response element, binding site of AP-1 [15].

C. The time-course of 4933436H12Rik expression in mouse R1 embryonic stem cells differentiating into embryoid bodies *in vitro* (GDS2667 record). The error bars represent standard deviation calculated from three independent arrays.





**Fig. 3. Mouse 1110008P08Rik/Kbtbd7 gene**

A. Diagram of the conservative domains of Kbtbd7 protein. BTB/POZ - BRC, Ttk and Bab domain or Pox virus and Zinc finger domain (pfam00651); BACK - BTB And C-terminal Kelch domain (pfam 07707); Kelch - Kelch1 motif (pfam01344).

B. The transcription factor binding sites in the proximal promoter of 4933436H12Rik. The notation and labeling as in Fig. 1B. TRE - TPE response element, binding site of AP-1 [15].

C. Analysis of the *Kbtbd7* gene expression in 5 different brain regions from 6 inbred strains (GDS2917 record). Inbred strains: 1 - 129S6/SvEvTac; 2 - A/J; 3 - C57BL/6J; 4 - C3H/HeJ; 5 - DBA/2J; 6 - FVB/NJ. The error bars represent standard deviation calculated from two independent arrays.



**Table 1**

Exon-intron structure of *4933436H12Rik* gene.

| Exon | Exon size (bp) | Exon-intron boundary  | Intron | Intron size (bp) | Intron-exon boundary |
|------|----------------|-----------------------|--------|------------------|----------------------|
| 1    | 106            | TGTGTA GAACgtaagatggg | 1      | 7063             | gatctgcagGTGTTTCATT  |
| 2    | 1141           | CAGTGAAAAGgtgggaaca   | 2      | 284              | ccctaacagGACTACTTGT  |
| 3    | 115            |                       |        |                  |                      |