Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

Trends in Bioinformatics 6 (3): 62-90, 2013
ISSN 1994-7941 / DOI: 10.3923/tb.2013.62.90

In-silico Study of Transcription Factor Binding Elements of Human PAX Gene Family Members

[1]Rashmi, [2]V.K. Singh, [3]A.N. Gangopadhyay, [1]G.L. Shah, [4]A. Khanna, [5]T.M. Mohapatra, [6]Om Shankar and [1]Royana Singh

[1]Department of Anatomy, Cytogenetics Unit,
[2]Centre for Bioinformatics, Faculty of Sciences, [3]Department of Pediatrics Surgery,
[4]Department of Gynecology and Obstetrics, [5]Department of Microbiology,
[6]Department of Cardiology, Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Corresponding Author: Royana Singh, Department of Anatomy, Cytogenetics Lab., Institute of Medical Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

ABSTRACT

PAX gene family members, tissue specific transcription factors mainly involved in the formation of tissues and organs during embryonic development and has important role in transcriptional regulation. The presence of consensus paired domain play significant role in DNA-binding transcription regulation with PAX domain. Regulatory behavior of PAX family members were determined using *cis*-acting elements study and repeat identification. The study helped in investigating the potential conserved motifs in the paired domain. Further, investigation of *cis*-acting elements was done to elucidate the function for each PAX members and then repeat analyses and their correlation with functional elements were done. The study illustrates that the *cis*-acting elements are involved in tissue specific developmental expression and transcriptional regulation of PAX family members. Further, based on physiochemical property study of these PAX gene family members it was found that they are mainly Ser, Pro, Gly and Ala rich amino acids. It was found that repeats containing functional DNA motifs interact with signature motifs of paired domain. The main six signature motifs NQLGG, NGRPLP, RPC, SR, GCVSKIL and PGAIGGSKP are involved in interaction. Altogether, this study provides new insights into the regulatory behavior of upstream region of each PAX members and its effect in transcriptional regulation and developmental expression of these PAX members with involvement in disease management.

Key words: Paired box, cis-acting elements, transcription factor, motifs

INTRODUCTION

PAX gene family belongs to the class of transcription factor (Walther *et al*., 1991). It comprises of a paired box domain (~128 amino acids long) which has a unique DNA binding motif (Noll, 1993). A homeodomain is present in some of the PAX genes, like PAX3, PAX4, PAX6 and PAX7, while an octapeptide is present in all PAX genes except PAX4 and PAX6 (Underhill, 2012; Walther and Gruss, 1996; Pierpont and Erickson, 1993; Underhill and Gros, 1997; Mansouri *et al*., 1996).

Members of PAX gene family has been classified into subgroups based on sequence homology as well as due to presence and absence of domains. PAX gene members have been classified into four subgroups: subgroup 1: PAX1/PAX9, subgroup 2: PAX2/PAX5/PAX8, subgroup 3: PAX3/PAX7, and subgroup 4: PAX4/PAX6 (Wang *et al*., 2010a; Balczarek *et al*., 1997).

Several defects occur due to insertion and deletion of functional motifs in these genes. An amino acid substitution in PAX1 gene is responsible for human neural tube defect mainly spina bifida (Hol *et al*., 1996). In mouse, haplo insufficiency in PAX1 is responsible for development of undulated phenotype with defect in vertebrae (Wilm *et al*., 1998).

Renal Coloboma Syndrome occurs due to frame-shift mutation in PAX2 gene (Quinlan *et al*., 2007). PAX2 gene has been postulated to be involved in growth and survival of cancer cells (Busse *et al*., 2009). PAX2 is also involved in papillary serous carcinoma of ovary (Tong *et al*., 2007). It regulates neuronal apoptosis inhibitory protein which is expressed in kidney (Dziarmaga *et al*., 2006). PAX2 mediates anchorage-independent cell growth of melanoma cells (Lee *et al*., 2011). A missense mutation in human PAX3 gene is associated with Waardenburg Syndrome (Zhang *et al*., 2012). Mutation in PAX3 gene is responsible for Waardenburg syndrome type 1 (Wang *et al*., 2010b). PAX3 is involved in melanoma cells (Medic *et al*., 2011). PAX3 has been involved in melanocyte specification and development (Medic and Ziman, 2010). PAX3 and PAX6 are involved in human embryogenesis (Terzic and Saraga-Babic, 1999). Frame shift mutation in PAX3 is responsible for spina bifida (Hol *et al*., 1995).

PAX4 plays role in tumor suppressor in human melanoma (Hata *et al*., 2008) as well as in hematologic malignancies (Li *et al*., 2006). Mutation in PAX4 gene is responsible for beta cell dysfunction (Kanatsuka *et al*., 2002). PAX5 is intricated in B cell differentiation (Pridans *et al*., 2008). PAX6 is responsible for development of eye (Chanas *et al*., 2009) and in pancreatic cell expression and differentiation (Gosmain *et al*., 2010). PAX7 has a role in neural crest development (Murdoch *et al*., 2012). PAX8 is involved in thyroid development (Ruiz-Llorente *et al*., 2012). Mutation in PAX9 leads to hypodontia and oligodontia (Das *et al*., 2002; Mostowska *et al*., 2003). PAX signature sequences are involved in binding to enhancer DNA sequence for transcriptional activation of target genes (Chi and Epstein, 2002). Since, PAX3 has a role in neural tube development, any mutation in PAX3 leads to neural tube defect. PAX3/PAX7 is involved in neural crest cell formation (Nikitina *et al*., 2008).

For computational analysis, the protein should belong to same family or super family because they have conserved sequence, structure and function. Genomic sequences from the existing database were used to determine the structural and functional aspect of PAX gene family. The search of regulatory element in promoter region of PAX3 and PAX7 will enable us to determine the development of neural crest. Physiochemical properties and transcription factor analysis of *cis*-acting elements helps in elucidating the functional properties of PAX gene family. Therefore, the present study was undertaken to elucidate the sequence-structure and functional aspect of the PAX gene family members. The purpose of the present study is to find out a correlation between sequence, structure and function of human PAX gene using a detailed study of PAX sequence, create the structure of PAX gene using homology modeling and thus help us to correlate the functions of these signatures at sequential and structural levels.

MATERIALS AND METHODS

All the available sequences of PAX proteins isolated from human were retrieved from NCBI.-4000 base pairs upstream to determine the repeats present in promoter region. Structures

for the entire PAX gene were modeled. Computational methods were used to determine the evolutionary relationship among PAX members. An interaction was checked by modeling and docking the repeats present in promoter region of each PAX gene.

Sequence analysis and phylogenetic classification: Protein sequences of all the genes of PAX family were retrieved from NCBI database (http://www.ncbi.nlm.nih.gov/; Pruitt *et al.*, 2007). The protein sequences of all PAX family members were used for MSA using ClustalW tool (http://www.ebi.ac.uk/Tools/msa/clustalw2; Thompson *et al.*, 1994) with default parameter. Phylogenetic tree was inferred using Neighbor joining method and MEGA 5.1 (Kumar *et al.*, 2012). The sequenced PAX gene family domains from NCBI were subjected to phylogenetic tree construction using MEGA with NJ Method at 1000 bootstrap replication value. For motif detection within entire gene family members, MEME Suite (Multiple EM for Motif Elicitation) Version 4.9.0 (http://meme.nbcr.net/meme/cgi-bin/meme.cgi; Bailey *et al.* (2009) was used. To identify core-conserved motifs in these sequences, maximum number of motif was set to 30 using minimum and maximum width of 30 and 50 amino acids, respectively. The PAX member was analyzed using InterProScan (http://www.ebi. ac.uk/Tools/pfa/iprscan/) which combines different protein signature recognition methods (Zdobnov and Apweiler, 2001).

Transcription factor binding sites (TFBS) and repeat analysis: Homology search was done using tBLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Gertz *et al.*, 2006; Altschul *et al.*, 1990) to find complete gene including start and end position. For study of *cis*-acting elements-4000 upstream regions were retrieved using NCBI. Cis-acting elements analysis were done using fgenesh tool of softberry server (http://linux1 .softberry.com/berry.phtml; Solovyev and Salamov, 1997). The positions of *cis*-acting elements were determined using enquiry of cis acting element database from transfac (Wingender *et al.*, 1996). In case of eukaryotic system it has been already reported that, cis-acting elements range from -19 (SF) to -1 (ST) (Kruse *et al.*, 1988) in rat -8 (SF) to 8 (ST) (Takiya *et al.*, 1990) in *Bombyx mori* to at -10150 (Herbst *et al.*, 1989) upstream region having most of the transcription factor binding sites in -8 to -4000 upstream region. Based on the above evidence-4000 upstream regions elements retrieved sequences have been inputted in Prom (Solovyev and Shahmuradov, 2003), TSSW and TSSG tools (Solovyev and Salamov, 1997) for *cis*-acting elements. Further repeat sequences were identified in upstream region using SSRIT tool (http://www.gramene.org/db/markers/ssrtool; Bilgen *et al.*, 2004; Cardle *et al.*, 2000) of Gramene database.

Study of physicochemical properties: The physicochemical properties of all the genes were calculated using Protparam tool available in Expasy server (www.web.expasy.org/protparam/; Wilkins *et al.*, 1999). This server helps to determine the physical and chemical properties of a protein. Amino acid comparative analysis was done for the entire PAX gene family member and graphics were generated using Microsoft Office Excel 2007 (Elliott *et al.*, 2006; Dixon *et al.*, 2009).

Homology modeling of PAX genes: For three dimensional structure predictions, target protein was taken for PDB BLAST search (http://www.rcsb.org/pdb/home/home.do; Berman *et al.*, 2000) to select template structure. Discovery studio version 3.1 was used for homology modeling of target protein (http://www.accelrys.com Discovery Studio 3.1, 2011; Joshi *et al.*, 2013). Structures of all members of PAX were modeled using Discovery Studio 3.1; except PAX5 (1K78) and PAX8 (2K27).

Structure refinement was done using RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php; Lovell *et al.*, 2003) and Errat version 2 (www.nihserver.mbi.ucla.edu/ERRATv2; Colovos and Yeates, 1993). CATH (www.cathdb.info; Orengo *et al.*, 1997) and SCOP database (www.scop.mrc-lmb.cam.ac.uk/scop; Murzin *et al.*, 1995) were used for structural classification of predicted models. Secondary structure components of the PAX sequences were analyzed using SWISS-PDB viewer (http://spdbv.vital-it.ch; Guex and Peitsch, 1997). The modeled structure of PAX gene family members were superimposed at backbone level and give the structural differences among them. This superimposition was done through Australis (http://eds.bmc.uu.se/eds/australis.php; Kleywegt, 1996).

DNA-protein interaction: Hex 6.3 (http://www.hex.loria.fr/dist/index.php; Ritchie and Venkatraman, 2010) was used to determine the DNA protein interaction. First docking was done between the protein modeled and already available DNA from different PDB ID. Visualization of interaction was done in Discovery studio version 3.1. Structure modeling of DNA was done using DNA Sequence to Structure tool available online (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp; Arnott *et al.*, 1976). Hex 6.3 tools were used for DNA-protein interaction study.

RESULTS

Sequence analysis and phylogenetic classification: All the protein sequences of PAX family members were aligned to find the extent of similarity present among the sequences of same family which enjoy a common phylogeny. Multiple sequence alignment of PAX protein sequence revealed a conserved sequence (NQLGG, NGRPLP, GCVSKIL, IGGSKP and VSSI). The sequences at N-terminal are highly conserved. The six signature motifs (NQLGG, NGRPLP, RPC, SR, GCVSKIL and PGAIGGSKP) were essentially conserved in all the PAX members with minor alterations (Fig. 1). In PAX6, a signature sequence of thirteen amino acid residue (HADAKVQVLDNQN) was

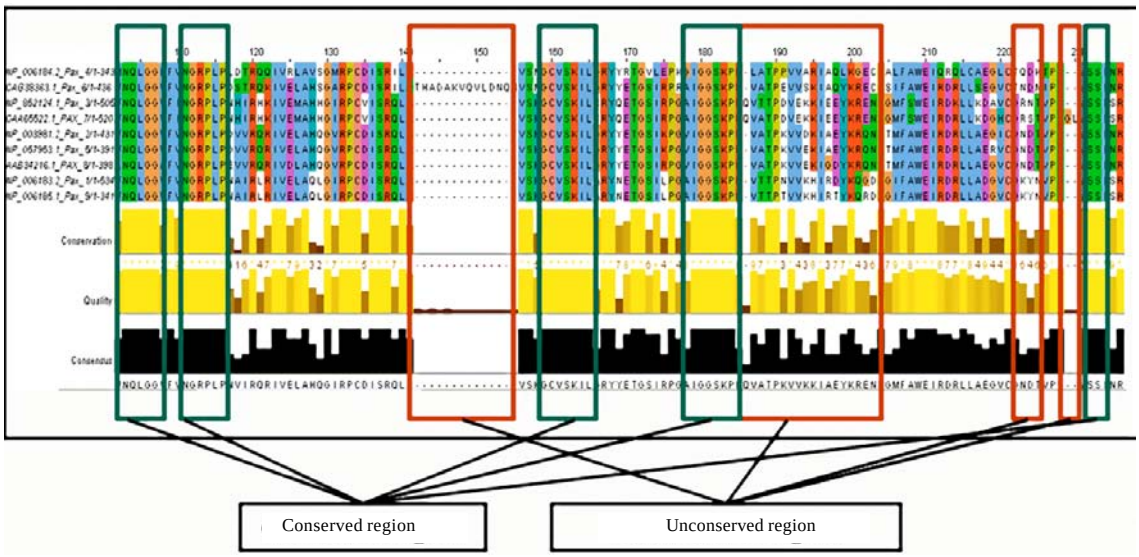

Fig. 1: Multiple sequence alignment obtained using ClustalW2. The region marked in red represents the differences observed in the sequence of PAX family. The region marked in green denotes the conserved region

observed which is not available in other family members (Fig. 1). Presence of Gly, Leu at 198,199 position distinguished PAX7 from PAX3 (Fig. 1). Insertion of Gln at 96 position, substitution of His at 118, 121 and 129, Val at 135, Cys at 156, Gln at 169, Asp at 191, Glut 193 and 197, Ser at 208, Lys at 220, Arg at 223 distinguished PAX3 and PAX7 with other family members. The non-conserved regions were also identified (marked red Fig.1). Interproscan determined that the identified motif belongs to paired domain (IPR001523), homeodomain (IPR009057) and winged helix-turn-helix transcription repressor DNA-binding (IPR011991).

In human nine different members of PAX gene were identified and classified. The clusters obtained were divided into four groups group 1: PAX1, PAX 9, group 2: PAX2, PAX5, PAX8 and group 3: PAX3, PAX7 and group 4: PAX4, PAX6 based on changes in motif (Fig. 2). Twenty-seven conserved motifs were observed in proteins of PAX gene family members. The motif-1 and motif-2 with multilevel consensus sequence VNQLGGVFVNGRPLPNHIRQKIVELAHHGIRPCDIS RQLRVSHGCVSKIL and KPKVATPKVVKKIAEYKRENPTMFAWEIRDRLLAEGVCDNDTVP SVSSID were uniformly conserved among all the PAX gene family domain sequences (Table 1). Motif-1 and motif-2 both represented the paired box domain and had a diverse range of DNA-binding domains that had a helix-turn-helix motif. Motif 3 was present in all domains except PAX1 and PAX9 representing homeobox in functional domain region. This family also contained a diverse range of DNA-binding domains that had a helix-turn-helix motif. Motif 5 was present only in PAX2 and PAX5 representing Paired-box protein 2_C terminal. Motif 4, 7, 8, 11, 12 and 15 were present only in PAX3 and PAX7 (Table 1). Motif 9 is present in all members except PAX3, PAX4 and PAX7 (Table 1). Motif 14 is present only in PAX2 and PAX5 representing Paired-box protein 2_C terminal (Table 1). Motif 13 is present in PAX4 and PAX9, motif 16 in PAX1 and PAX8, motif 17 in PAX6 and PAX7, motif 18 in PAX1 and PAX9 and motif 19 in PAX6 and PAX8 (Table 1). Motif 20 is present at the N-terminal in PAX1 and at C-terminal in PAX4 (Table 1). Whereas Motif 21 at C-terminal of PAX1 was present at N-terminal in PAX6 representing paired domain (Table 1). Motif 22 present at N-terminal of PAX1 was present at C-terminal of PAX6 (Table 1). Motif 23 is present in PAX4 and PAX6 (Table 1). Motif 24 and 25 are present in PAX1 and PAX2. Motif 26 present at N-terminal of PAX4 was present at C-terminal in PAX6 (Table 1). Motif 27 present at N-terminal of PAX8 was present at C-terminal of PAX9 (Table 1).

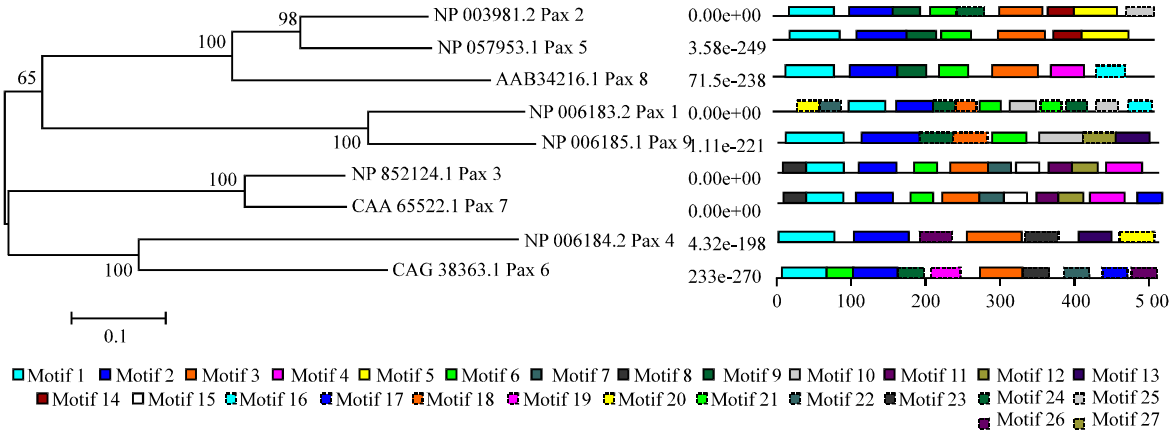

Fig. 2: Unrooted Phylogenetic tree construction of PAX gene family domain sequences using neighbor-joining method. Bootstrap values are indicated against each branch. Bootstrap similarities are >98% and the tree were built after 1000 replication. Schematic distribution of respective conserved motifs identified by means of MEME software

Table 1: Multilevel consensus sequences for the MEME defined motifs among PAX gene family domains

Name of motif	Name of PAX gene	Motif sequence
Motif 1	All	VNQLGGVFVNGRPLPNHIRQKIVELAHHGIRPCDISRQLRVSHGCVSKIL
Motif 2	All	KPKVATPKVVKKIAEYKRENPTMFAWEIRDRLLAEGVCDNDTVPSVSSIN
Motif 3	All except PAX1 PAX9	FTQEQLEELEKEFERTHYPDIFTREELAKKEKLTEYRVQVWFSNRRAKWR
Motif 4	PAX3 PAX7	CSQRADHMKPGDSLPTSQAYCPPTYSTTGYSMDPVAGYQYGQYGQSEC
Motif 5	PAX2 PAX5	YPPHVPPAGQGSYPAPTLAGMVPGSEFSGNPYSHPQYTAYNEAWRFPN
Motif 6	All except PAX4 PAX6	DEWKSSHSINGILGIRSFMEDKNKMDEYEG
Motif 7	PAX3 PAX7	KQAGANQLMAFNHLIPGGFPPTGMPTLPPYQL
Motif 8	PAX3PAX7	VPRMMRPGPGQNYPRTGFPLEVSTPLGQGR
Motif 9	All except PAX3 PAX4 PAX7	RIIRNKVQQPFQQGMDGAYKQIQMTPHHTL
Motif 10	PAX1 PAX9	NNFPFPAPHAVNGLEKGALEQEIKYGQAPNGLPAVGGF
Motif 11	PAX3 PAX7	YCAYGARHGFSSYTDSFMNPAGPSNHMNPV
Motif 12	PAX3 PAX7	NGLSPQVMGILGNHGGVPHQPQADFAISPLHGGLE
Motif 13	PAX4 PAX9	HCWQHCWGTALERCNCDIPPKACFKGCWGH
Motif 14	PAX2 PAX5	TNPEIGSNVPGPQTYPIVTGRDMASTTLPG
Motif 15	PAX3 PAX7	DQPYPIPQIVQDGGSTVHRPQPLPPSTMHQ
Motif 16	PAX1PAX8	EEWWGPGCRDAHPASPQADRCAWPDLPHFAWW
Motif 17	PAX6 PAX7	CMFMESYKVNGGWYDTITQMEKQKHMNMEQF
Motif 18	PAX1 PAX9	HNHIYPYPSPITPAGAKMGTHPGVPGIPGH
Motif 19	PAX6PAX8	WHQGYPVVGQPPFWICQKQEGGGENTMPFPMNGEC
Motif 20	PAX4 PAX1	DSGLRCRPCRVSHCHLGRRGGRQAGAWPGC
Motif 21	PAX6 PAX1	DNQNNQNGCYSKIGGGYYEPGPIWPRAIGG
Motif 22	PAX6 PAX1	SMGGGGDQALPDCYGPLPGMPGFPMANNLP
Motif 23	PAX4 PAX6	KEKWEMQRRGASNGLTHIRIAPGFITAQYQ
Motif 24	PAX2 PAX1	GDGVHIHGGEGAHAMWFKHDVREGSVPNGD
Motif 25	PAX2PAX1	CPLRPRPMRGTSYRGDHRKERQDDFGAHIC
Motif 26	PAX6 PAX4	QGLPCTGLRSPGVLVPVQVPGHEGDMTQYW
Motif 27	PAX8 PAX9	MPHNMIRYGHGGQNQLYMAFVNGRPGYEVG

Identification of *cis*-acting elements: In case of PAX1, seven potential *cis*-acting elements were identified in which HS$HH4_02 (-296), HS$PK_04 (-290), MOUSE$IAP (-208), HS$CDC25C (-152) are novel cell cycle-regulated repressor elements involved in transcription. In case of PAX2, best eight *cis*-acting elements were identified. The predicted *cis*-acting elements are HS$ALBU_03 (-222), MOUSE$MT1 (-152), HS$PK_04 (-322), HS$TGFB1_0 (-144), HS$EG_06 (-294), HS$FN_08 (-142), PH$PAL_04 (-91) and HS$GG_36 (-262). PAX3 has best thirteen *cis*-acting elements which are HS$BAC_03 (-1268), HS$GG_12 (-1272), HS$GG_17 (-1268), HS$HH4_02 (-1176), HS$IGF2_02 (-1020), HS$DPOLA_0 (-1192), HS$TK_01 (-1277), HS$EG_07 (-1269), HS$APOA2_0 (-1142), HS$CLASE_04 (-1142), HS$FN_06 (-1269). In case of PAX4, four *cis*-acting elements were identified in which HS$TK_01 (-1158), HS$EG_08 (-1094), HS$CLASE_0 (-2783) are involved in growth and development. PAX5 has six elements which are HS$EGFR_12 (-354), HS$GMCSF_0 (-527), (-392), HS$GRH_03 (-561), HS$HH1_01 (-334), HS$PL_02 (-335), HS$PR264_0 (-420). In case of PAX6, five *cis*-acting elements were identified HS$APOE_09 (-672), HS$GG_12 (-752), HS$TK_01 (-797), HS$FN_08 (-716), HS$CDC2_09 (-760) and all of them are involved in developmentally expressed gene. In case of PAX7, eleven *cis*-acting elements HS$GG_12 (-629), HS$GMCSF_0 (-706), HS$GRH_03 (-504), HS$HH4_02 (-567), HS$PK_04 (-741), HS$EG_06 (-691), HS$EG_08 (-507) and HS$CLASE_04 (-685) were identified. In case of PAX8, twelve *cis*-acting elements HS$CDC2_06 (-263), HS$CDC2_09 (-196), HS$CDC25C_01

(-133), HS$PL_12 (-222), HS$GP2B_09 (-407), HS$EG_06 (-171), HS$TK_01 (-353), HS$INS_04 (-199), HS$BG_01 (-166), HS$GG_13 (-167), MOUSE$MT1_01 (-139) are identified. In case of PAX9, best nine *cis*-acting elements were identified. The predicted elements are HS$CDC2_06 (-1577), HS$FN_08 (-1465), HS$GP2B_13 (-1468), HS$EG_06 (-1465), HS$GMCSF_0 (-1548), HS$GG_17 (-1464), HS$GG_12 (-1484), HS$BG_01 (-1543), HS$BAC_03 (-1268).

Repeat identification in upstream of PAX gene family members: Repeat identification in upstream of PAX gene family members represents the presence of a number of repeats available in promoter region. GT, CTCC, AG, AC, TG, GA, CCT, AC and AGGG repeats were observed in all PAX genes upstream except PAX4. PAX1 had $(GT)^{22}$ repeat at position (-269), PAX2 had $(CTCC)^{5}$ repeat at position (-916). PAX3 had $(AG)^{5}$ repeat at position (-3755), $(AC)^{5}$ repeat at position (-2823) and $(AC)^{25}$ repeat at position (-2752). PAX5 had $(GT)^{6}$ repeat at position (-1035) and $(AC)^{15}$ repeat at position (-432). PAX6 had $(TG)^{5}$ repeat at position -74. PAX7 had $(GA)^{5}$ repeat at position (-3449), $(TG)^{5}$ repeat at position (-2440), $(CCT)^{10}$ repeat at position (-799). PAX8 $(AC)^{8}$ repeat at position (-2131). PAX 9 had $(AGGG)^{5}$ repeat at position (-742). Table 2 representing the details of repeats along with position, name of promoter, sequential region and their respective functions.

Table 2: The position of repeat and its function present in promoter region of PAX family

Gene name	Promoter position	Promoter name	Sequence	Function
PAX1 GT	-138	CHICK$BG_0	GGGTGGGG	CHICK$BG_0 act as promoter/enhancer binding protein in NF-E4 (Gallarda *et al*., 1989)
	-138	PUF$CONS	GGGTGGG	PUF$CONS plays an important role in regulation of c-myc by acting as transcription factor binding sites (Postel *et al*., 1989)
	-139	HS$BG_17	GGTGGGG	Presence of conserved motif is responsible for protein interaction (Strauss and Orkin, 1992)
	-139	HS$BG_22	GGTGG	Presence of conserved motif is responsible for protein interaction (Strauss and Orkin, 1992)
	-131	HS$AG_12	cctgtggGGGTGGa	The regulatory element has binding site for nuclear factor GATA-1 and AP-1 binding sites and also act as primary element for controlling globin gene expression (Jarman *et al*., 1991)
alpha-				
PAX2 CTCC	-422	HS$NPY_04	CCCCTCC	The Cis-acting sequences are involved in the expression of the human neuropeptide Y (NPY) gene and has a DNA binding site (Minth and Dixon, 1990)
PAX3 AG	-71	PA$PY_12	AGAGG	These cis acting elements are pathogenic in nature and have DNA binding sites (Piette and Yaniv, 1986)
PAX5 AAC	-44	MOUSE$RAS1	ACAACA	The presence of glucocorticoid plays important role in transcription of the mouse Ha-ras promoter region (Strawhecker *et al*., 1989)
PAX7 GA	-71	RAT$POMC_0	CAGAG	The cis acting element is responsible for repression of gene transcription due to two DNA-binding proteins (Drouin *et al*., 1989)
PAX7 GA	-70	DROME$EVE_	gctGAGAGcagcac	These play an important role in regulation of gene expression in developmental genes (Read *et al*., 1990)
PAX8 AC	-127	HS$HH1_01	AAACACA	The cis acting element has two classes of trans acting factor that is specific for histone gene promoter and human gene promoter (Van Wijnen *et al*., 1988)
AC	-141	XENLA$AC_0	tgtgctgCACCTGtctactc	It has DNA binding site for muscle development (Taylor *et al*., 1991)
PAX9 AGGG	-742	Nil	Nil	Nil

Physiochemical properties: Analysis of physiochemical properties of the PAX gene family showed that PAX gene members were rich in Ser, Pro, Gly and Ala (Fig. 3). Based on negative gravy index it was found that PAX gene family members were mainly hydrophilic in nature. The instability index was above 40 in all the genes of PAX family indicating that these family members were highly unstable. This hydrophilic nature facilitates the structural properties determination of PAX gene family. The paired domain in PAX gene showed the presence of turns and helixes.

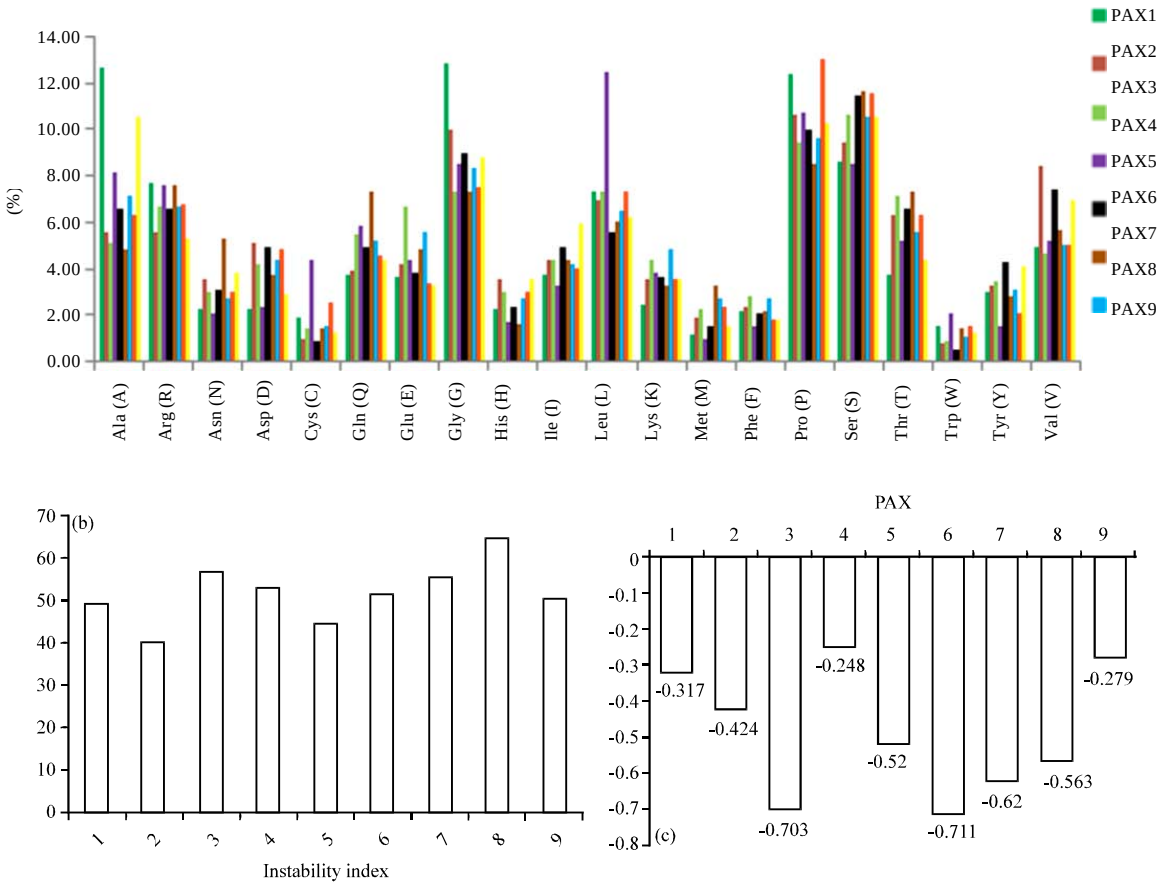

Fig. 3(a-c): Physiochemical properties of PAX gene family members (a) Graph representing the percentage of amino acid in different PAX gene family. Name of amino acid is on x axis and percentage on Y axis, (b) Graph representing the instability index in PAX gene family. Instability index is present on Y axis and name of PAX family members on X axis and (c) Graph representing the Gravy index in PAX gene family. Gravy index is present on Y axis and name of PAX family members on X axis

Structure analysis of PAX gene family: The best-refined model structures were submitted in PMDB database. The PMDB ID for PAX1, PAX2, PAX3, PAX4, PAX6, PAX7 and PAX9 were PM0078614, PM0078622, PM0078618, PM0078619, PM0078621, PM0078620 and PM0078623, respectively. The designed structure contains 123, 124, 127, 126, 147,124 and 123 amino acid residues for PM0078614, PM0078618, PM0078619, PM0078620, PM0078621, PM0078622 and PM0078623, respectively. Secondary elements analysis of determined structures showed the presence of turns and helix in PAX1, PAX2, PAX4 and PAX9 and only β-sheets and turn

Table 3: Superimposition of modeled structure of PAX family. Here, superimposition was done between the PAX proteins of same group

Name of gene	PMDB ID	Sequeuce identity (%)	Z score	RMSD	Relative RMSD	Normalized RMSD	SAS (A°)
PAX1	PM0078614	52.85	1.9170E+01	0.19	0.01767	0.245 A	0.296
PAX9	PM0078623						
PAX3	PM0078618	54.84	1.9742E+01	0.38	0.03351	0.470 A	0.558
PAX7	PM0078620						
PAX4	PM0078619	42.52	1.9961E+01	0.94	0.05079	1.070 A	1.195
PAX6	PM0078621						

Table 4: Superimposition of all the protein modeled using PAX1 as a template

Name of gene	Protein modeled	RMSD
PAX1	PM0078614	Used as a template
PAX2	PM0078622	12.338
PAX3	PM0078618	16.401
PAX4	PM0078619	4.51
PAX6	PM0078621	3.312
PAX7	PM0078620	7.27
PAX9	PM0078623	10.408

arrangement in PAX3, PAX6 and PAX7 structures. The sequences of the modeled structure of PAX protein were analyzed using Interproscan to investigate the functional domain region. The domain present in the structure mainly belongs to paired box domain with IPR001523, homeodomain like with IPR009057 and winged helix-turn-helix transcription repressor DNA binding with IPR011991. Structural classification using CATH resulted that this super family represents winged helix repressor DNA binding domain (CATH ID: 1.10.10.10). The Z-score obtained for different superimposition signifies that they are good. The Z-score obtained was close to zero indicating that the structures belong to same family having same fold type but major differences in loop/turn region. Table 3 represents the Z-score, RMSD, sequence identity, relative RMSD and normalized RMSD of modeled structures. Using PAX1, as a template rests of the structures were superimposed using Discovery studio version 3.1. The RMSD retrieved was much higher indicating that although the structures are of same family but changes are occurring. In other words, superimposition study helps in determining that the major changes are observed in turn/loop region. PAX members overlapped with each other in an overall similar deviation with some differences. These differences observed in turn/loop regions of all PAX gene members (Fig. 4). Figure 4 representations in green denotes that turn region is not similar at structural level. Table 4 represents the RMSD obtained by superimposing all the modeled structure.

DNA-protein interaction: The modeled PAX family members were used for interaction with repeats present in promoter region. Table 5 and Fig. 5 represent the interaction and function of PAX proteins except PAX4, PAX5 and PAX8 with their respective DNA repeats. During the interaction RMS obtained for all the members was -1 representing that the interaction of DNA with PAX protein was good and binding energy also represent best affinity of binding between DNA and protein models. By this investigation, it was found that all repeats are binding with respective core signature residues of functional domain of each PAX gene members for stabihzing this PAX family. Based on repeats study revealed that GT, CTCC, AG, AAC, GA, AC TG, CCT and AGGG repeats

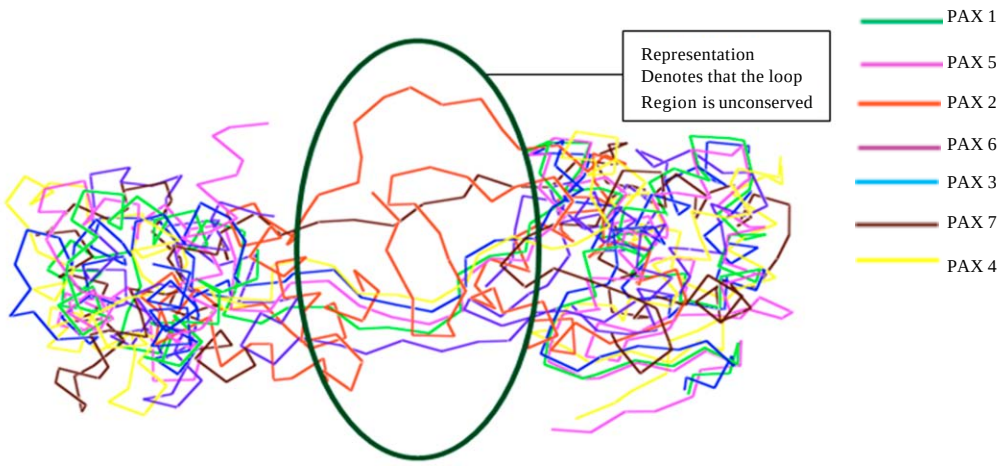

Fig. 4: Superimposition of the modeled structure of PAX gene family

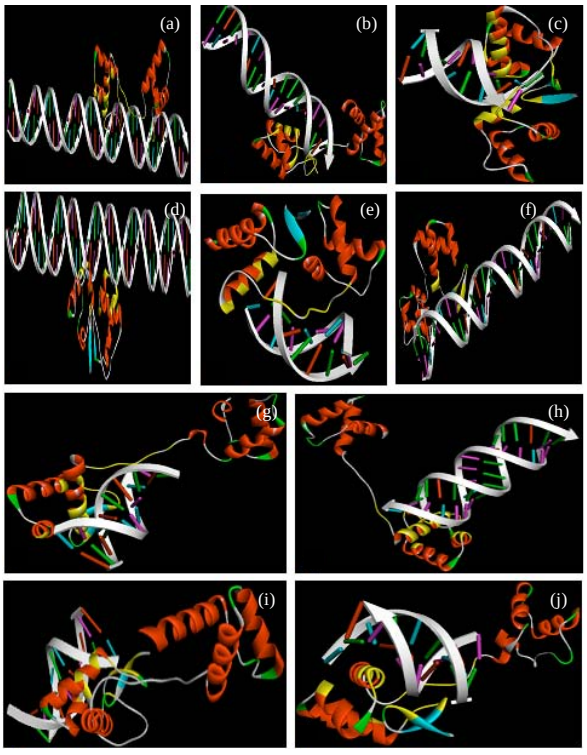

Fig. 5(a-j): Protein-DNA interaction of different PAX gene with DNA modeled for repeats using Hex 6.3 (a) PAX1-GT repeat DNA complex, (b) PAX2-CTCC repeat DNA complex, (c) PAX3-AC repeat DNA complex, (d) PAX3-AC repeat DNA complex. The repeat of AC was retrieved for 25 times, (e) PAX3-AG repeat DNA complex, (f) PAX6-TG repeat DNA complex, (g) PAX7-CCT repeat DNA complex, (h) PAX7-TG repeat DNA complex, (i) PAX7-Showing interactions with GA repeat DNA and (j) PAX9-AGGG repeat DNA complex

Table 5: Interaction of DNA repeats with the modeled structure of PAX protein amino acids along with positions

Gene name	Repeat name	Interacted residues	Function of repeats based on *cis*-acting binding elements	RMS	Energy obtained	Function of repeats based on literature review
PAX1	GT repeat	Asn103, Gln104, Leu105, Phe109, Val110, Asn111, Gly112, Arg113, Pro114, Leu115, Arg120, Arg132, Pro133, Cys134, Arg138, Val142, Ser143, His144, Gly145, Cys146, Ser148, Lys149, Arg153, Pro162, Gly163, Ala164, Ile165, Gly166, GLY167, Ser168, Arg171	The beta-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. A Nuclease-Hypersensitive Element of the Human c-myc. Promoter Interacts with a Transcription Initiation Factor. *In vivo* protein-DNA interactions at hypersensitive site 3 of the human beta-globin locus control region. Characterization of the major regulatory element upstream of the human alpha-globin gene cluster	-1.00	-5.595 017e+002	PMID:21945498, PMID:21556890 PMID:20796278. PMID:20708634 PMID:20175935, PMID:15235025 PMID:11754397, PMID:11513560 PMID:11333219, PMID:11179974 PMID:10701898, PMID:11102993 PMID:10050298, PMID:9920050 PMID:9630081, PMID:8557253, PMID:8125493, PMID:8125490 PMID:8225327, PMID:1595886 PMID:1370810, PMID:1754418
PAX2	CTCC repeat	Gly19, Asn21, Gln22, Leu23, Phe27, Val28, Asn29, Gly30, Arg31, Arg50, Pro51, Arg59, Val60, Ser61, His62, Gly63, Ser66, Lys67, Gly70, Lys87	Expression of the Human Neuropeptide Y Gene	-1.00	4.862 507e+002	PMID:1346778 PMID:1370809
PAX3	AC repeat (5 times)	Asn39, Gln40, Leu41, Phe45, Ile46, Arg49, Pro50, Leu51, Arg68, Pro69, Val78, Ser79,	Nil	-1.00	-4.862 507e +002	PMID:9501299

Table 5: Continue

Gene name	Repeat name	Interacted residues	Function of repeats based on *cis*-acting binding elements	RMS	Energy obtained	Function of repeats based on literature review
		His80, Gly81, Cys82, Ser84, Lys85, Leu87, Cys88, Gln91, Glu92, Pro106, Lys107, Gln108, Val109, Thr110, Ser152, Ser153, Ser155, Arg156, Ile157, Leu158				
PAX3	AC repeat (25 times)	Gln40, Leu41, Ala63, His64, His65, Gly66, Ile67, Arg68, Pro69, Cys70, Arg74, Ser84, Leu87, Cys88,Gln91, Gly127, Phe129, Ser130, Trp131, Glu132, Val151, Ser155, Leu158	Nil	-1.00	-5.033 719e +002	Nil
PAX3	AG repeat	His80, Gly81, Cys82, Val83, Ser84, Lys85, Arg89, Glu92, Thr93, Ala100, Ile101, Gly102, Ser104, Lys105, Pro106, Lys107, Gln108, Val109, Pro112	Molecular analysis of the interaction between an enhancer binding factor and its DNA target.	-1.00	-5.226 558e +002	PMID:3027657
PAX6	TG repeat	Ser3, Gly4, Val11, Phe12, Val13, Pro19, Ser21, Thr22, Lys25, Arg41 ,Leu43, Gln44,His46, Asn58	Nil	-1.00	-3.268 263e+002	Nil
PAX7	CCT repeat	Gln35, Gly36, Arg37, Val38, Gln40, Pro52, Asn53, His54, Ile55, Arg77	Nil	-1.00	-3.993 109e +002	Nil

Table 5: Continue

Gene name	Repeat name	Interacted residues	Function of repeats based on *cis*-acting binding elements	RMS	Energy obtained	Function of repeats based on literature review
PAX7	GA repeat	Asn39, ASN47, GLY48, ARG49, PRO50, LEU51, ARG56, CYS70, SER79, HIS80, GLY81, CYS82, SER84, LYS85, ARG89, PRO98, GLY99, ALA100, ILE101, GLY102, GLY103, LYS105	Glucocorticoid receptor binding to a specific DNA sequence is required for hormone-dependent repression of Pro-Opiomelanocortin gene transcription. The Drosophila Even-Skipped Promoter is transcribed in a Stage-Specific Manner *in vitro* and Contains Multiple, Overlapping Factor-Binding Sites	-1.00	-5.565 139e +002	PMID:18533279 PMID:9149848 PMID:7849732
PAX7	TG repeat	ASN39, GLN40, ASN47, GLY48, ARG49, PRO50, LEU51, ARG56, PRO69, CYS70, SER79, HIS80, GLY81, CYS82, SER84, LYS85, ARG89, PRO98, GLY99, ALA100, ILE101, GLY102,GLY103, LYS105	Nil	-1.00	-5.521 989e +002	PMID:17234733 PMID:12391723 PMID:9589597 PMID:9298742 PMID:9156330 PMID:2395673
PAX9	AGGG repeat	GLN10, LEU11, PHE15, VAL16, ASN17, GLY18, ARG19, LEU21, ARG26, PRO39, CYS40, SER43, SER49, HIS50, GLY51, CYS52, SER54, LYS55, LEU57, ALA58, ARG59, ASN61, GLU62, GLY69, ALA70, ILE71	Nil	-1.00	-5.513 411e+002	Nil

Table 6: The interaction of DNA with the modeled structure of PAX

S. No	PAX1GT repeat	PAX2 CTCC repeat	PAX3 AC repeat	PAX3 AC repeat (25 times)	PAX3 AG repeat	PAX6 TG repeat	PAX7 CCT repeat	PAX7 TG repeat	PAX7 GA repeat	PAX9 AGGG repeat
1	Asn[5], Gln[6], Leu[7]	Asn[5], Gln[6], Leu[7]	Asn[5] ,Gln[6], Leu[7]	Gln[6], Leu[7]			Gln[6]	Gln[6], Leu[7]	Asn[5]	Gln[6], Leu[7]
2	Phe[9], *Val[10]*, Asn[11],Gly[12], Arg[13], Pro[14], Leu[15]	Phe[9], Val[10], Asn[11], Gly[12], Arg[13], *Pro[14]*, *Leu[15]*	Phe[11], Ile[12], *Asn[13]*, *Gly[14]*, *Arg[15]*, Pro[16], Leu[17]		Arg[15]	Phe[12], Val[13], *Asn[14]*, *Gly[15]*, *Arg[16]*, Pro[17]		*Phe[12]*, *Ile[13]*, Asn[14], Gly[15], Arg[16], Pro[17], Leu[18]	*Phe[11]*, *Ile[12]*, Asn[13], Gly[14], Arg[15], Pro[16], Leu[17]	Phe[15], *Val[16]*, Asn[17], Gly1[8], Arg[19], *Pro[20]*, Leu[21]
3	Arg[31], Pro[32], Cys[33]	Arg[31], Pro[32]	Arg[34], Pro[35]	Arg[34], Pro[35], Cys[36]				Pro[36], Cys[37]	Cys[35]	Pro[32], Cys[33]
4	Val[41], Ser[42], *His[43]*, Gly[44], Cys[45]	Val[42], Ser[43], *His[44]*, Gly[45]	Val[44], Ser[45], *His[46]*, Gly[47], Cys [48]		Val[44], Ser[45], *His[46]*,Gly[47], Cys [48]			Ser[46], His[47], Gly[48], Cys[49]	Ser[45], *His[46]*, Gly[47], Cys[48]	Ser[42], *His[43]*, Gly[44], Cys[45]
5	Ser[47], Lys[48]	Ser[48], Lys[49]	Ser[50], Lys[51]	Ser[50]	Ser[50], Lys[51]			Ser[51], Lys[52]	Ser[50], Lys[51]	Ser[47], Lys[48]
6	Pro[61], *Gly[62]*, *Ala[63]*, *Ile[64]*, Gly[65], Gly[66], Ser[67]				*Gly[65]*, *Ala[66]*, *Ile[67]*, Gly[68], Gly[69], Ser[70]			Pro[65], *Gly[66]*, *Ala[67]*, *Ile[68]*, Gly[69], Gly[70], *Ser[71]*, Lys[72]	Pro[64], *Gly[65]*, *Ala[66]*, *Ile[67]*, Gly[68], Gly[69], *Ser[70]*, Lys[71]	Gly[62], Ala[63], Ile[64]

These interactions are compared with the MEME result and the Multiple Sequeuce Alignment result. Amino acid marked italic denotes that interaction was not observed. Amino acid in bold determines the substitution of amino acid

found in upstream region of PAX gene members. Function of each repeats was also determined based on literature review along with the PMID no. represented in the Table 5. Figure 6 represents the interaction between repeats and their respective paired domains.

Based on multiple sequence alignment (Fig. 1), motif identification (Table 1) DNA-Protein interaction of different repeats (Fig. 5, Table 6) in PAX family members it was determined that the signature motif NQL were present in the entire family members except in PAX3 with AG repeat. The motif FVNGRPL was conserved in the entire family members except PAX3 with $(AC)^{25}$ repeat and PAX7 with CCT repeat. RPC represent another consensus sequence present in the entire family members of PAX family except PAX3 with AG repeat and PAX7 with CCT repeat. VSHG was also conserved in the entire family members except PAX3 with $(AC)^{25}$ repeat and PAX7 with CCT repeat.

Motif SK was conserved in the entire family member except PAX7 with CCT repeat and S was observed in PAX3 with AC repeat 25 times. PGAIGG also act as a signature motif with minor exception in all the PAX family members. The same interaction of repeat was present in all the members except in PAX2 (CTCC repeat), PAX3 (AC repeat), PAX3 $(AC)^{25}$ repeat and PAX7 (CCT repeat). In PAX1 (GT repeat), PAX3 (AG repeat), PAX7 (TG repeat) and PAX7 (GA repeat) no interaction was observed in GAI from the signature motif PGAIGG. In PAX6 (TG repeat) only the motif FVNGRPLP is involved in which FV and P are involved in DNA-protein interaction. FV represents β strand and P represents turn region. *In-silico* study revealed that TG repeat was present in both PAX6 and PAX7. However, the TG repeat of PAX6 only shows interaction with the motif FVNGRPL whereas the TG repeat of PAX7 shows interaction with all the motifs (NQL, FVNGRPL, RPC, SK, PGAIGG and VSHGC) (Table 5, Fig. 6).

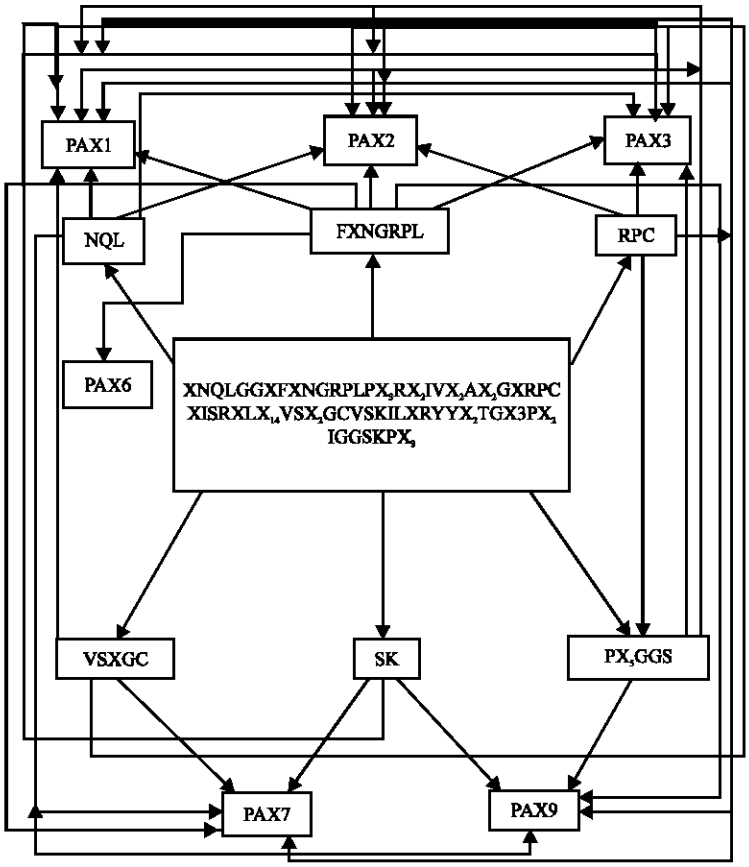

Fig. 6: Flowchart represents the conserved motif and their interaction with different PAX family members

The motif NQL, FVNGRPL, PGAIGG represents turn region whereas motif SK represent helix. RPC has R as turn, PC as helix, VSHGC has VS as turn and GC represents helix region. In PAX3 and PAX7 in the signature motif FVNGRPL, I replaced V and the FI present denotes β strand involved in interaction. However, the results represent the same signature motif in all the members with minor differences. These results indicate the conservation of motifs in the entire family member of PAX gene family and even the interaction were observed in all the members with different repeat.

These interactions were compared with the MEME result and the MSA result. Amino acid marked red denotes that interaction was not observed. Amino acid in green determines the substitution of amino acid.

In the PAX paired domain and DNA interaction the loops and helix are the major player. Yet β-strand also play role in interaction in PAX3, PAX6 and PAX7. In PAX3 and PAX7 in the signature motif FVNGRPL, I (12, 13) replaced V (10, 13 and 16). This change is responsible for structural change resulting in β-strand in place of loop. In PAX6 however, V (10, 13, 16) was present but they represent β-strand. The signature motifs NQL, FXNGRPL, RPC, VSXGC and PX_2IGGS were highly conserved and play an important role in the interaction with DNA. The same signature motif was present in interaction with different repeats with minor variation (Table 6).

DISCUSSION

Here, the study report on the use of this bioinformatics tools and techniques in identification, classification and visualization of the PAX family members in the genetic model *Homo sapiens*. The prediction through sequence alignment, phylogenetic analysis, motif identification, functional annotation of *cis*-acting binding elements, physiochemical analysis and structure analysis by homology modeling, along with interaction of repeats with modeled structure, elucidated PAX members having paired domain. The proposed study showed that the interaction occurred in paired domain elucidating that PAX members has paired domain which are involved in DNA protein interaction which is similar to the study observed by Chi and Epstein (2002). It was observed that the three dimensional structural analysis of PAX members facilitated in determining that they belong to the paired domain.

Previous analysis reports that PAX family has a paired domain (Noll, 1993; Walther and Gruss, 1996; Pierpont and Erickson, 1993; Underhill and Gros, 1997; Mansouri *et al*., 1996). However, in this study, the sequential changes present in the paired domain of PAX members and the interaction of this paired domain with the repeats present in upstream region was observed.

Multiple Sequence Alignment shows the sequence conservation among different members of PAX family in humans. Along with the conservation, there are differences that are sufficient to support the variations subsequently reflected at the structural and functional levels (Fig. 1). For example, conserved motifs NQLGG, NGRPLP, RPC, SR, GCVSKIL and PGAIGGS which forms the paired domain of the PAX protein, are directly involved in protein-DNA interactions, are essentially conserved in all human sequences previously identified by Balczarek *et al*. (1997). In addition to the conservation of signature motifs, there are some un-conserved regions as HADAKVQVLDNQN is present in PAX6 which is unconserved in other family members. Presence of Gly, Leu at 198,199 position helps in distinguishing PAX7 with PAX3. Insertion of Gln at 96 position, substitution of His at 118, 121 and 129, Val at 135, Cys at 156, Gln at 169, Asp at 191, Glu at 193 and 197, Ser at 208, Lys at 220, Arg at 223 distinguishes PAX3 and PAX7 with other family members.

In human, the nine different members of PAX gene family were previously identified and classified into four groups (Wang *et al*., 2010a; Balczarek *et al*., 1997). Phylogenetic tree results outlined the clustering of PAX gene members into different groups (Fig. 2). The classification of PAX family has been done on the presence and absence of signature motif. PAX2, PAX5 and PAX8 contain similar motif in the same sequential order hence kept in same cluster. In PAX2 and PAX5 motif 14 was present, motif 24 and 25 were present only in PAX2 and in PAX8 motif 16, 18 and 27 were present representing that PAX2 and PAX5 are similar with minor differences hence they are closest to each other. The presence of motif 16, 18 and 27 in PAX8 distinguishes PAX8 from PAX2 and PAX7. In PAX1 and PAX9, similar motif were observed in same order with presence of motif 13 and 27 in PAX9 and motif 16, 20, 21, 22, 24 and 25 in PAX1. Presence of motif 16 in PAX1 determines that it is close to PAX8. PAX3 and PAX7 represent approximately the same motif except presence of motif 17 in PAX7. PAX4 and PAX6 contain similar motif in same order with presence of motif 11, 13 and 20 in PAX4 and motif 17, 19, 21 and 22 in PAX6. Presence of conserved motifs in domain of PAX family represents the evolutionary similarity and diversity within these family members (Fig. 2). Motif-1and motif-2 both represents the paired box domain, DNA-binding domains that contain a helix-turn-helix motif (Table 1). Motif 3 is present in all domains except PAX1 and PAX9 representing homeobox domain and DNA-binding domains having a helix-turn-helix motif (Table 1). Motif 5 present in PAX2 and PAX5 represents Paired-box protein 2_C terminal which is transcription factor, involved in embryonic development and organogenesis (Table 1). Motif

4,7,8,11,12 and 15 are present only in PAX3 and PAX7 whose function has not been determined (Table 1). In motif 11 the conserved sequence represent PAX7 having role in paired-box-containing transcription factor involved in the control of developmental processes (Table 1). Motif 9 is present in all members except PAX3, PAX4 and PAX7 (Table 1). Motif 14 is present only in PAX2 and PAX5 representing Paired-box protein 2_C terminal. This family having the C terminal of the paired-box protein 2 which is a transcription factor, involved in embryonic development and organogenesis (Table 1). Motif 21 is present in PAX6 and PAX1 representing paired box domain (Table 1). This family contains a diverse range of mostly DNA-binding domains that contain a helix-turn-helix motif. Protein functional analysis of these domains using Pfam and IPR confirmed their identity to PAX family like proteins (Table 1). Motif 22 present at N-terminal of PAX1 was present at C-terminal of PAX6 (Fig. 2). Motif 26 present at N-terminal of PAX4 was C-terminal in PAX6 (Fig. 2). Motif 27 present at N-terminal of PAX8 was present at C-terminal in PAX9 (Fig. 2). Pfam and IPR determines that these motif belongs to paired type domain (IPR001523) having winged helix-turn-helix transcription repressor DNA-binding (IPR011991) site.

The main focus has been on the analysis of *cis*-acting elements in upstream of PAX family members where the function of upstream elements present in PAX members were elucidated using animal *cis*-acting database.

In PAX1 the possible function of HS$HH4_02 is observed during DNA binding activity as cells progress toward division (Arnott *et al*., 1976). HS$PK_04 is associated with proenkephalin and act as enhancer binding factors (La Bella and Heintz, 1991), MOUSE$IAP is reported as novel enhancer core-binding component (Mermod *et al*., 1988). HS$CDC25C is associated with cell cycle-regulated repressor element (Falzon and Kuff, 1989). The promoter HSV1$IE3_0 acts as promoter-regulatory domains of alpha genes of herpes simplex virus type which shows pathogenic properties for PAX1 gene (Lucibello *et al*., 1995). The predicted *cis*-acting element HS$APOE_08 is involved in regulation by multiple positive and negative elements to control gene expression (Kristie and Roizman, 1986).

HS$ALBU_03 and MOUSE$MT1 plays a significant role in interaction with signature domain of PAX2 gene (Smith *et al*., 1988; Paik *et al*., 1988). HS$PK_04 acts as an enhancer binding factors (La Bella and Heintz, 1991). HS$TGFB1_0 is related to promoter sequences of the human transforming growth factor (Seguin and Hamer, 1987). HS$EG_06 acts as binding site for promoter-enhancer interaction (Kim *et al*., 1989). Element HS$FN_08 can play as responsive element for dependent transcriptional activation (Gong *et al*., 1991). PH$PAL_04 is involved in the responses to both UV irradiation and elicitor application (Bernath *et al*., 1990). HS$GG_36 acts as silencer sequences in the human and play important role in stage-specific silencing of PAX2 gene (Lois *et al*., 1989).

In PAX3, HS$BAC_03 (-1268) is involved in efficient transcription of gene in non-muscle cells and expression is developmentally regulated (Gumucio *et al*., 1992). The element HS$GG_12 (-1272) acts as repressor at two region at a distance of about 60-75 nucleotides in mutation of Hereditary Persistence of Fetal Hemoglobin (Frederickson *et al*., 1989). Element HS$GG_17 (-1268) is associated with DNA-protein interaction (Mantovani *et al*., 1988a). The HS$HH4_02 (-1176) element has DNA binding properties of transcription factor involved in cell cycle regulation (Ikuta and Kan, 1991). HS$IGF2_02 (-1020) acts as repressor of IGF-II (insulin-like growth factor 2). HS$DPOLA_0 (-1192) plays an important role in cell cycling (La Bella and Heintz, 1991). HS$TK_01 (-1277) is responsible for cell-cycle-regulated expression (Kim and Lee, 1991). HS$EG_07 (-1269) acts as transcription factor for human gene promoter (Gong *et al*., 1991).

HS$APOA2_0 (-1142) has an important role in gene regulation (Ogami *et al.*, 1991). HS$GP2B_13 (-1272) has two domain in promoter region and deletion of these domain leads to loss of promoter activity and presence of domain leads to tissue specific expression (Uzan *et al.*, 1991). HS$CLASE_04 (-1142) controls both up and down regulation of gene (Konig *et al.*, 1992). HS$FN_06 (-1269) is responsible for inhibition of gene expression in newly developed granulose cells (Bernath *et al.*, 1990). HSV1$IE3_0 (-1021) are present in HSV-1 and are pathogenic in nature (Kristie and Roizman, 1986). HS$PK_04 (-1237) are enhancer for AP-4 and AP-1 together they activate SV-40 late transcription (Mermod *et al.*, 1988). All predicted *cis*-acting elements showed that PAX3 gene can involved in these types of functions.

Promoter region of PAX4 shows that HS$TK_01 (-1158) controls the expression of cell cycle regulation (Kim and Lee, 1991). HS$EG_08 (-1094) act as binding site for erythroid specific enhancer (Gong *et al.*, 1991). HS$CLASE_04 (-1142) is involved in up and down regulation (Kim and Lee, 1991). HS$PR264_0 6 (-1246) has a DNA binding domain and it has a physiological target for transcription factor family (Sureau *et al.*, 1992).

The promoter region of PAX5 HS$EGFR_12 (-354) involved in DNA-protein interaction (Xu *et al.*, 1993). HS$GMCSF_0 (-527), (-392) acts in controlling the expression of gene in bone marrow microenvironment (Nimer *et al.*, 1988). HS$GRH_03 (-561) functions as DNA binding activity for growth hormone (Lefevre *et al.*, 1987). HS$HH1_01 (-334) has two target site for protein binding in promoter region for regulation of cell cycle (Van Wijnen *et al.*, 1988). HS$PL_02 (-335) functions in expression of human prolactin gene (Peers *et al.*, 1990). HS$PR264_0 (-420) plays role in expression of transcription factor family (Sureau *et al.*, 1992).

The promoter region of PAX6 HS$APOE_09 (-672) is involved in multiple upregulation and downregulation of gene expression (Paik *et al.*, 1988). The predicted *cis*-acting elements HS$GG_12 (-752) have effect on binding of erythroid-specific factor related gene (Mantovani *et al.*, 1988b). The predicted *cis*-acting elements HS$TK_01 (-797) determines expression of cell cycle regulation (Kim and Lee, 1991). The element HS$FN_08 (-716) plays role in inhibition of gene expression in newly developed granulose cells (Bernath *et al.*, 1990). The *cis*-acting elements HS$CDC2_09 (-760) acts as repressor of cell cycle interacting with gene which have similarity with upstream activating sequence (Zwicker *et al.*, 1995).

In PAX7, HS$GG_12 (-629) have effect on binding of erythroid-specific factor related gene (Mantovani *et al.*, 1988a). HS$GMCSF_0 (-706) stimulates the formation of granulocyte, macrophages and eosinophil. Promoter region has role in T-cell specific expression and are involved in expression of those cell where it is not present, also has DNA binding properties (Shannon *et al.*, 1988). The possible *cis*-acting element HS$GRH_03 (-504) plays role in expression of human growth hormone and their recognition is characterized by trans-acting elements (Lefevre *et al.*, 1987). The *cis*-acting elements HS$HH4_02 (-567) is a multigene, has role in DNA synthesis and transcription regulation (Dailey *et al.*, 1986, 1988; Bernath *et al.*, 1990). HS$PK_04 (-741) has functional *cis*-element for activating early and late transcription and has DNA binding factor (Mermod *et al.*, 1988). HS$EG_06 (-691) and HS$EG_08 (-507) elements are mainly predicted to be expressed during embryonic development and has a role in protein-DNA interaction (Gong *et al.*, 1991; Yu *et al.*, 1991). HS$CLASE_04 (-685) is involved in up and down regulation of gene (Konig *et al.*, 1992). HCMC$IE1_1 (-699) forms nucleoprotein complex during transcription regulation (Ghazal *et al.*, 1987; Niller and Hennighausen, 1991). The HS$PR264_06 (-750) acts as target for transcription factor (Sureau *et al.*, 1992). Y$ADH2_01 (-475) acts as upstream activating system of gene (Eisen *et al.*, 1988).

In PAX8, HS$CDC2_06 (-263) is involved in cell cycle progression (Wen *et al.*, 1995). HS$CDC2_09 (-196) shows repression of activated transcription due to presence of cell cycle dependent element (Zwicker *et al.*, 1995). HS$CDC25C_01 (-133) is involved in cell cycle progression (Lucibello *et al.*, 1995). HS$PL_12 (-222) had homeo and POU domain (Peers *et al.*, 1991). HS$GP2B_09 (-407) has sequence implicated in DNA-protein interaction (Uzan *et al.*, 1991). HS$EG_06 (-171) is predicted to be expressed during embryonic development and in protein-DNA interaction (Gong *et al.*, 1991; Yu *et al.*, 1991). HS$TK_01 (-353) plays role in expression of cell cycle regulation (Kim and Lee, 1991). HS$INS_04 (-199) shows expression of human gene by multiple Trans acting factors (Boam *et al.*, 1990). HS$BG_01 (-166) plays role in erythroid and embryonic development for transcriptional regulation and are homologous to NF-1 which are involved in DNA protein interaction (Mantovani *et al.*, 1988b; Jones *et al.*, 1987; Schule *et al.*, 1988). The *cis*-acting element HS$GG_13 (-167) is involved in switching from fetal-adult hemoglobin (Gumucio *et al.*, 1988). MOUSE$MT1_01 (-139) binds nuclear factor with upstream metal regulatory element of the gene (Kristie and Roizman, 1986). HS$APOB_04 (-150) element has a DNA binding site (Kardassis *et al.*, 1990; Paulweber *et al.*, 1991).

In PAX9, the transcription factor binding site HS$CDC2_06 (-1577) is involved in cell cycle progression progression (Wen *et al.*, 1995). Element HS$FN_08 (-1465) can play as responsive element for dependent transcriptional activation (Bernath *et al.*, 1990). Element HS$GP2B_13 (-1468) has sequence implicated in DNA protein interaction (Xu *et al.*, 1993). HS$EG_06 (-1465) is predicted to be expressed during embryonic development and has a role in protein-DNA interaction (Gong *et al.*, 1991; Yu *et al.*, 1991). HS$GMCSF_0 (-1548) stimulates the formation of granulocyte, macrophages and eosinophil. Promoter region plays significant role in T-cell specific expression and are involved in expression of those cell where it is not present, also has DNA binding properties (Shannon *et al.*, 1988). HS$GG_17 (-1464) elements has the property of DNA-protein interaction (Ikuta and Kan, 1991). HS$GG_12 (-1484) have effect on binding of erythroid-specific factor (Mantovani *et al.*, 1988a). HS$BG_01 (-1543) was found in erythroid and embryonic development for transcriptional regulation and are homologous to NF-1 which are involved in DNA protein interaction (Mantovani *et al.*, 1988b; Wen *et al.*, 1995). HS$BAC_03 (-1268) is required for efficient transcription of gene in non-muscle cells and expression is developmentally regulated (Frederickson *et al.*, 1989). The *cis*-acting elements study revealed that main possible function of these PAX gene members act as transcription regulators that has significant role in cell growth and differentiation.

Based on bioinformatics analysis above; it was observed that some elements are common to other members of PAX gene family. HS$APOE_08, HS$APOE_09 are involved in multiple up regulation and down regulation of gene expression common to PAX1 and PAX6. The element HS$HH4_02 act as for histone He pHu4A gene reported during DNA binding activity as cells progress toward division found in PAX1, PAX3 and PAX7. In PAX1 and PAX3 the element HSV1$IE3_0 can act as promoter-regulatory domains of alpha genes of herpes simplex virus type. HS$PK_04 determines enhancer binding factors are common to PAX1, PAX2, PAX3, PAX7. HS$CDC25C_functions as cell cycle-regulated repressor element in PAX1 and PAX8. MOUSE$MT1_is involved in interaction with signature domain in PAX2 and PAX8. The element HSEG_06, HSEG_07, HS$EG_08 has DNA binding sites in PAX2, PAX3, PAX4, PAX7, PAX8, PAX9. HSFN_08, HSFN_06 plays role in transcription activation of PAX2, PAX6, PAX9. HSGG_36, HSGG_12, HSGG_17, HSGG_12 are present in erythroid specific DNA-protein interaction of PAX2, PAX3, PAX6, PAX7, PAX8 and PAX9. HS$BAC_03 determines developmental

regulation of muscle in PAX3 and PAX9. Element HS$TK_01 has significant role in cell cycle regulation of PAX3, PAX4, PAX6 and PAX8. The element HS$GP2B_13, HS$GP2B_09 can act as DNA-protein interaction in upstream region of PAX3, PAX8 and PAX9. The element HS$CLASE_04 controls the up and down regulation of PAX3, PAX4, PAX7. HS$PR264_06 has a DNA binding domain involved in transcriptional regulation of PAX4, PAX5 and PAX7. The element HS$GMCSF_0 has been detected in PAX5, PAX7 and PAX9 has role in regulation of gene in bone marrow microenvironment for formation of granulocyte, macrophages. HS$GRH_03 has DNA binding activity for growth hormone in PAX5 and PAX7. HS$PL_02 has homeodomain similar to PAX5, PAX8. HS$CDC2_06, HS$CDC2_09 has role in cell cycle regulation and are present in PAX8, PAX9. In PAX8 and PAX9, HS$BG_01 was found which has role in erythroid specific regulation and DNA-protein interaction of NF gene. The expression which are similar suggests that dysfunction of a gene can be regulated by other gene. Despite of the similarity and differences in each PAX members resulted that all the members have the property of DNA-protein interaction, development and transcriptional regulation. These elements have motifs which are present in other gene and are present in PAX members.

The analysis of probable repeats in the upstream of PAX family members and their functions showed that the elements predicted in GT repeat of PAX1 were CHICKBG_0, PUFCONS, HSBG_17, HSBG_22, HS$AG_12 and CHICK$BG_0. These elements functions as transcriptional enhancer (Gallarda *et al*., 1989) of CHICK$BG_0 and regulator (Postel *et al*., 1989) of PUF$CONS. HS$BG_17 and HS$BG_22 is responsible for protein interaction (Strauss and Orkin, 1992). HS$AG_12 is involved in transcriptional gene expression (Jarman *et al*., 1991). CTCC repeat of PAX2 has HS$NPY_04 element which is involved in expression of NPY gene having DNA binding sites (Minth and Dixon, 1990). The AG repeat of PAX3 has DNA binding sites that are pathogenic in nature (Piette and Yaniv, 1986). The AAC repeat of PAX5 has important role in transcriptional regulation (Strawhecker *et al*., 1989). The GA repeat of PAX7 functions as up and down regulation of gene transcription (Drouin *et al*., 1989; Read *et al*., 1990). The AC repeat of PAX8 regulates human and histone gene promoter (Van Wijnen *et al*., 1988; Taylor *et al*., 1991). The repeats present in upstream region of PAX gene family may play important role in disease management.

Thus, it is suggested that these motifs in PAX family may show the same functions as that of other gene. The study further illustrates that the *cis*-acting elements present in PAX gene is involved developmental expression and transcriptional regulation of PAX family members. The same element present in repeats and promoter region of PAX5 (HS$HH1_01), PAX8 and PAX9 (HS$BG_01) has important role in cell cycle regulation and transcriptional regulation, respectively.

Physiochemical properties determine hydrophilic nature of PAX gene family. This shows the presence of turns and helix in paired box domain of PAX gene family. These members are rich Ser, Pro, Gly and Ala and are unstable in nature.

The modeled structures have 90% refinement in favored region. Best predicted models were deposited in PMDB database. The PMDB ID for PAX1, PAX2, PAX3, PAX4, PAX6, PAX7 and PAX9 were PM0078614, PM0078622, PM0078618, PM0078619, PM0078621, PM0078620 and PM0078623, respectively. The structural analysis in revealed that the paired domain is mainly conserved in all the members (Underhill, 2012; Walther and Gruss, 1996; Pierpont and Erickson, 1993; Underhill and Gros, 1997; Chi and Epstein, 2002). The modeled structure of PAX9 has close similarity to the PDBID 2k27A, 6paxA, 1pdnC and 1r1uA with Genbank accession no. NP_006185.1. The overall structure is similar for different PAX genes with slight change in

backbone specially loop region due to difference in amino acids in signature region. During the superimposition the Z-score obtained was close to zero indicating that the structures belong to same family having same fold type but major differences in loop region. Superimposition of PAX members helps in elucidating that helix was conserved and major changes were observed in loop/turn region.

DNA-protein interaction showed that the structural analysis exhibits the presence of helix and turns in all the members except in PAX3, PAX6 and PAX7 where β-strands are present. However, whatever variations were present in PAX family at structural and functional levels, the paired domain present were found to conserve the 3D structural folds that distinguishes them as paired domains. DNA-Protein interactions occur mainly in the paired domain. In this study, it has been elucidated that repetitive DNA structures interact with signature part of functional or structural domain. It was observed that during the interaction analysis of the modeled structure with the DNA repeats (functional repeat motifs) help in stabilization of the PAX members. DNA binding sites are mainly present in loop region. The signature motifs NQL, FXNGRPL, RPC, VSXGC and PX_2IGGS were highly conserved and play an important role in the interaction with DNA. The same signature motif was present in interaction with different repeats with minor variation. This variation may be due to presence of different repeat in the promoter region. This may also be the cause of different function played by these family members. The functional diversity of PAX family can be determined due to variation in structure and sequence of this family.

GT repeat polymorphism play an important role in Heme oxygenase (HO)-1 which is responsible for BP regulation (Wu *et al*., 2011), in case of (GT)n repeat in HO-1 gene promoter prevents carotid arthesclerosis (Wu *et al*., 2010). In dyslipidemia patients GT repeat in heme oxygenase-1 gene promoter protects from ischemic stroke (Bai *et al*., 2010), GT repeat in monoamine oxidase B (MAOB) may play an important role in Parkinson's disease (Mellick *et al*., 2000). The insertion of two short CTCC repeat leads to deletion of 27-bp in allele of the type 3 collagen gene (COL3A1) in a family with Ehlers-Danlos syndrome type IV (Richards *et al*., 1992). Polymorphic GA repeat in SH2D2A gene leads to arousal of chronic inflammatory demyelinating polyradiculoneuropathy (Notturno *et al*., 2008). Polymorphism in intronic TG repeat observed in the human proteasome core particle PROS-27K gene act as a marker for genetic linkage study of common diseases (Sjakste *et al*., 2002). TG repeat polymorphism in neuronal nitric oxide synthase is not responsible for hypertension (Takahashi *et al*., 1997).

This study suggest that the regulatory elements and repeats sequences will help in regulation based on switch on and off condition of each gene members of PAX family. The repeats interact with the signature motifs of paired domain. These signature functional motifs are present in the upstream region of some other genes hke gamma globin, neuropeptide Y, histone etc. and PAX gene may show the same function as those of other genes. The above all results elucidate that repeats found in promoter regions of PAX Gene family members will help in disease management.

CONCLUSION

In silico study was done to determine the functional motifs at sequential and structural level in members of PAX gene family. The *cis*-acting elements study in -4000 bp upstream region of PAX gene family members has been investigated to elucidate the functional role and behaviour of each members of this family.

Sequential similarity and differences in each PAX members concluded that all the members have the property of DNA-protein interaction and involved in tissue specific developmental and transcriptional regulation. The identified elements have certain roles i.e., HS$HH4_02 has DNA

binding activity as cells progress toward division in PAX1, PAX3 and PAX7. HSEG_06, HSEG_07 and HS$EG_08 elements has DNA binding activity and available in PAX2, PAX3, PAX4, PAX7, PAX8and PAX9. HS$GRH_03 element has mainly play important role during tissue growth and development and available in PAX5 and PAX7 members.

The analysis of probable repeats in the upstream of PAX family members and their functions showed that the elements predicted in GT repeat of PAX1 functions as transcriptional enhancer of CHICK$BG_0 and regulator of PUF$CONS. HS$BG_17 and HS$BG_22 is responsible for DNA protein interaction. HS$AG_12 is involved in transcriptional gene expression. CTCC repeat of PAX2 has HS$NPY_04 element which is involved in expression of NPY gene having DNA binding sites. The AG repeat of PAX3 in element PA$PY_12 has DNA binding sites that are pathogenic in nature. The AAC repeat of PAX5 has element MOUSE$RAS1 and play important role in transcriptional regulator in promoter region. The GA repeat of PAX7 in elements RAT$POMC_0 and DROME$EVE_functions as up and down regulation of gene transcription. The AC repeat of PAX8 in element HS$HH1_01 regulates human and histone gene promoter. The repeats present in upstream region of PAX gene family may play important role in disease management during transcriptional regulation of developmental genes involved in neurogenesis. The regulatory elements present in PAX3 and PAX7 are suggested to play role in transcriptional regulation of neural crest development. Structural alignment within PAX proteins derived that paired box helix fold type are core conserved and changes occur in loop/turn region which may affect the behavior of each members of this family. DNA-Protein interactions occur mainly in the paired domain region contain helices and turns type secondary elements. In this investigation it has been found that functional repetitive DNA elements interact with signature part of functional domain, mainly DNA binding sites are present in loop region. Interaction analysis of the modeled structures with the DNA repeats (functional repeat motifs) help in stabilization of PAX gene members. The regulatory elements and repeats identification in upstream sequences will help in understanding of gene regulation of each gene members of PAX family.

ACKNOWLEDGMENT

The authors are thankful to the Coordinator, Center of Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi, India, for facilities to do the present study in the department.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman, 1990. Basic local alignment search tool. J. Mol. Biol., 215: 403-410.

Arnott, S., P.J. Campbell-Smith and R. Chandrasekaran, 1976. Nucleic Acids. In: Handbook of Biochemistry and Molecular Biology, Fasman, G.P. (Ed.). Vol. 2, CRC Press, Cleveland, USA., pp: 411-422.

Bai, C.H., J.R. Chen, H.C. Chiu, C.C. Chou, L.Y. Chau and W.H. Pan, 2010. Shorter GT repeat polymorphism in the heme oxygenase-1 gene promoter has protective effect on ischemic stroke in dyslipidemia patients. J. Biomed. Sci., Vol. 17. 10.1186/1423-0127-17-12

Bailey, T.L., M. Boden, F.A. Buske, M. Frith and C.E. Grant *et al.*, 2009. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res., 37: W202-W208.

Balczarek, K.A., Z.C. Lai and S. Kumar, 1997. Evolution of functional diversification of the paired box (Pax) DNA-binding domains. Mol. Biol. Evol., 14: 829-842.

Berman, M.H., J. Westbrook, Z. Feng, G. Gilliland and T.N. Bhat *et al.*, 2000. The protein data bank. Nucl. Acids Res., 28: 235-242.

Bernath, VA., A.F. Muro, A.D. Vitullo, M.A. Bley and J.L. Baranao and A.R. Kornblihtt, 1990. Cyclic AMP inhibits fibronectin gene expression in a newly developed granulosa cell line by a mechanism that suppresses cAMP-responsive element-dependent transcriptional Activation. J. Biol. Chem., 265: 18219-18226.

Bilgen M., M. Karaca, A.N. Onus and A.G. Ince, 2004. A software program combining sequence motif searches with keywords for finding repeats containing DNA sequences. Bioinformatics, 20: 3379-3386.

Boam D.S., A.R. Clark and K. Docherty, 1990. Positive and negative regulation of the human insulin gene by multiple trans-acting factors. J. Biol. Chem., 265: 8285-8296.

Busse, A., A. Rietz, S. Schwartz, E. Thiel and U. Keilholz, 2009. An intron 9 containing splice variant of PAX2. J. Transl. Med., Vol. 7. 10.1186/1479-5876-7-36

Cardle, L., L. Ramsay, D. Milborne, M. Macaulay, D. Marshall and R. Waugh, 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics, 156: 847-854.

Chanas, S.A., J.M. Collinson, T. Ramaesh, N. Dora, D.A. Kleinjan, R.E. Hill and J.D. West, 2009. Effects of elevated Pax6 expression and genetic background on mouse eye development. Invest. Ophthalmol. Vis. Sci., 50: 4045-4059.

Chi, N. and J.A. Epstein, 2002. Getting your pax straight: Pax proteins in development and disease. Trends Genet., 18: 41-47.

Colovos, C. and T.O. Yeates, 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci., 2: 1511-1519.

Dailey, L., S.B. Roberts and N. Heintz, 1988. Purification of the human histone He gene-specific transcription factors H4TF-1 and H4TF-2. Genes Dev., 2: 1700-1712.

Dailey, L., S.M. Hanly, R.G. Roeder and N. Heintz, 1986. Distinct transcription factors bind specifically to two regions of the human histone He promoter. Proc. Natl. Acad. Sci., 83: 7241-7245.

Das P., D.W. Stockton, C. Bauer, L.G. Shaffer, R.N. D'Souza, T. Wright and P.I. Patel, 2002. Haploinsufficiency of PAX9 is associated with autosomal dominant hypodontia. Hum. Genet., 110: 371-376.

Dixon, M.R., J.W. Jackson, S.L. Small, M.J. Horner-King, N.M.K. Lik, Y. Garcia and R. Rosales, 2009. Creating single-subject design graphs in Microsoft Excel™ 2007. J. Applied Behav. Anal., 42: 277-293.

Drouin, J., M.A. Trifiro, R.K. Plante, M. Nemer, P. Eriksson and O. Wrange, 1989. Glucocorticoid receptor binding to a specific DNA sequence is required for hormone-dependent repression of pro-opiomelanocortin gene transcription. Mol. Cell. Biol., 9: 5305-5314.

Dziarmaga, A., P.A. Hueber, D. Iglesias, N. Hache and A. Jeffs *et al.*, 2006. Neuronal apoptosis inhibitory protein is expressed in developing kidney and is regulated by PAX2. Am. J. Physiol. Renal. Physiol., 291: F913-F920.

Eisen, A., W.E. Taylor, H. Blumberg and E.T. Young, 1988. The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activating sequence of ADH2. Mol. Cell Biol., 8: 4552-4556.

Elliott, A.C., L.S. Hynan, J.S. Reisch and J.P. Smith, 2006. Preparing data for analysis using microsoft Excel. J. Invest. Med., 54: 334-341.

Falzon, M. and E.L. Kuff, 1989. Isolation and characterization of a protein fraction that binds to enhancer core sequences in intracisternal a-particle long terminal repeats. J. Biol. Chem., 264: 21915-21922.

Frederickson, R.M., M.R. Micheau, A. Iwamoto and N.G. Miyamoto, 1989.. 5' Flanking and first intron sequences of the human beta-actin gene required for efficient promoter activity. Nucleic Acids Res., 17: 253-270.

Gallarda, J.L., K.P. Foley, Z.Y. Yang and J.D. Engel, 1989. The beta-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. Genes Dev., 3: 1845-1859.

Gertz, E.M., Y.K. Yu, R. Agarwala, A.A. Schaffer and S.F. Altschul, 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biol. Vol. 4 10.1186/1741-7007-4-41

Ghazal, P., H. Lubon, B. Fleckenstein and L. Hennighausen, 1987. Binding of transcription factors and creation of a large nucleoprotein complex on the human cytomegalovirus enhancer. Proc. Natl. Acad. Sci. USA., 84: 3658-3662.

Gong, Q.H., J. Stern and A. Dean, 1991. Transcriptional role of a conserved GATA-1 site in the human epsilon-globin gene promoter. Mol. Cell. Biol., 11: 2558-2566.

Gosmain, Y., E. Marthinet, C. Cheyssac, A. Guerardel and A. Mamin *et al.*, 2010. Pax6 controls the expression of critical genes involved in pancreatic{alpha} cell differentiation and function. J. Biol. Chem., 285: 33381-33393.

Guex, N. and M.C. Peitsch, 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis, 18: 2714-2723.

Gumucio, D.L., K.L. Rood, T.A. Gray, M.F. Riordan, C.I. Sartor and F.S. Collins, 1988. Nuclear proteins that bind the human gamma-globin gene promoter: Alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. Mol. Cell. Biol., 8: 5310-5322.

Gumucio, D.L., H. Heilstedt-Williamson, T.A. Gray, S.A. Tarle and D.A. Shelton *et al.*, 1992. Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. Mol. Cell. Biol., 12: 4919-4929.

Hata, S., J.I. Hamada, K. Maeda, T. Murai and M. Tada *et al.*, 2008. PAX4 has the potential to function as a tumor suppressor in human melanoma. Int. J. Oncol., 33: 1065-1071.

Herbst R.S., N. Friedman, J.E. Darnell Jr. and L.E. Babiss, 1989. Positive and negative regulatory elements in the mouse albumin enhancer. Proc. Natl. Acad Sci. USA., 86: 1553-1557.

Hol, F.A., B.C. Hamel, M.P. Geurds, R.A. Mullaart, F.G. Barr, R.A. Macina and E.C. Mariman, 1995. A frameshift mutation in the gene for PAX3 in a girl with spina bifida and mild signs of Waardenburg syndrome. J. Med. Genet., 32: 52-56.

Hol, F.A., M.P. Geurds, S. Chatkupt, Y.Y. Shugart and R. Balling *et al.*, 1996. PAX genes and human neural tube defects: An amino acid substitution in PAX1 in a patient with spina bifida. J. Med. Genet., 33: 655-660.

Ikuta, T. and Y.W. Kan, 1991. *In vivo* protein-DNA interactions at the beta-globin gene locus. Proc. Natl. Acad. Sci. USA., 88: 10188-10192.

Jarman, A.P., W.G. Wood, J.A. Sharpe, G. Gourdon, H. Ayyub and D.R. Higgs, 1991. Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. Mol. Cell. Biol., 11: 4679-4689.

Jones, K.A., J.T. Kadonaga, P.J. Rosenfeld, T.J. Kelly and R. Tjian, 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell, 48: 79-89.

Joshi, R.S., D.M. Jamdhade, S.M. Sonawane and A.P. Giri, 2013. Resistome analysis of *Mycobacterium tuberculosis*: Identification of aminoglycoside 2'-N-acetyltransferase (AAC) as co-target for drug desigining. Bioinformation, 9: 174-181.

Kanatsuka, A., Y. Tokuyama, O. Nozaki, K. Matsui and T. Egashira, 2002. Beta-cell dysfunction in late-onset diabetic subjects carrying homozygous mutation in transcription factors NeuroD1 and Pax4. Metabolism, 51: 1161-1165.

Kardassis, D., M. Hadzopoulou-Cladaras, D.P. Ramji, R. Cortese, V.I. Zannis and C. Cladaras, 1990. Characterization of the promoter elements required for hepatic and intestinal transcription of the human apoB gene: Definition of the DNA-binding site of a tissue-specific transcriptional factor. Mol. Cell Biol., 10: 2653-2659.

Kim, S.J., K.T. Jeang, A.B. Glick, M.B. Sporn and A.B. Roberts, 1989. Promoter sequences of the human transforming growth factor-beta 1 gene responsive to transforming growth factor-beta 1 autoinduction. J. Biol. Chem., 264: 7041-7045.

Kim, Y.K. and A.S. Lee, 1991. Identification of a 70-base-pair cell cycle regulatory unit within the promoter of the human thymidine kinase gene and its interaction with cellular factors. Mol. Cell. Biol., 11: 2296-2302.

Kleywegt, G.J., 1996. Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr. D: Biol. Crystallogr., 52: 842-857.

Konig, H., H. Ponta, H.J. Rahmsdorf and P. Herrlich, 1992. Interference between pathway-specific transcription factors: Glucocorticoids antagonize phorbol ester-induced AP-1 activity without altering AP-1 site occupation *in vivo*. EMBO J., 11: 2241-2246.

Kristie, T.M. and B. Roizman, 1986. DNA-binding site of major regulatory protein alpha 4 specifically associated with promoter-regulatory domains of alpha genes of herpes simplex virus type 1. Proc. Natl. Acad. Sci., 83: 4700-4704.

Kruse, F., C.T. Komro, C.H. Michnoff and R.J. MacDonald, 1988. The cell-specific elastase I enhancer comprises two domains. Mol. Cell. Biol., 8: 893-902.

Kumar, S., G. Stecher, D. Peterson and K. Tamura, 2012. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. Bioinformatics, 28: 2685-2686.

La Bella, F. and N. Heintz, 1991. Histone gene transcription factor binding in extracts of normal human cells. Mol. Cell. Biol., 11: 5825-5831.

Lee, S.B., K. Doberstein, P. Baumgarten, A. Wieland and C. Ungerer *et al.*, 2011. PAX2 regulates ADAM10 expression and mediates anchorage-independent cell growth of melanoma cells. PLoS ONE, Vol. 6. 10.1371/journal.pone.0022312

Lefevre, C., M. Imagawa, S. Dana, J. Grindlay, M. Bodner and M. Karin, 1987. Tissue-specific expression of the human growth hormone gene is conferred in part by the binding of a specific trans-acting factor. EMBO J., 6: 971-981.

Li, Y., H. Nagai, T. Ohno, H. Ohashi, T. Murohara, H. Saito and T. Kinoshita, 2006. Aberrant DNA demethylation in promoter region and aberrant expression of mRNA of PAX4 gene in hematologic malignancies. Leukemia Res., 30: 1547-1553.

Lois, R., A. Dietrich, K. Hahlbrock and W. Schulz, 1989. A phenylalanine ammonia-lyase gene from parsley: Structure, regulation and identification of elicitor and light responsive cis-acting elements. EMBO J., 8: 1641-1648.

Lovell, S.C., I.W. Davis, W.B. Arendall, P.I. de Bakker and J.M. Word *et al.*, 2003. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. Proteins, 50: 437-450.

Lucibello, F.C., M. Truss, J. Zwicker, F. Ehlert, M. Beato and R. Muller, 1995. Periodic cdc25c transcription is mediated by a novel cell cycle-regulated repressor element (CDE). EMBO J., 14: 132-142.

Mansouri, A., M. Hallonet and P. Gruss, 1996. Pax genes and their roles in cell differentiation and development. Curr. Opin. Cell Biol., 8: 851-857.

Mantovani, R., N. Malgaretti, S. Nicolis, A. Ronchi, B. Giglioni and S. Ottolenghi, 1988a. The effects of HPFH mutations in the human gamma-globin promoter on binding of ubiquitous and erythroid specific nuclear factors. Nucleic Acids Res., 16: 7783-7797.

Mantovani, R., N. Malgaretti, S. Nicolis, B. Giglioni and P. Comi *et al.*, 1988b. An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. Nucleic Acids Res., 16: 4299-4313.

Medic, S. and M. Ziman, 2010. PAX3 expression in normal skin melanocytes and melanocytic lesions (naevi and melanomas). PLoS One, Vol. 5. 1371/journal.pone.0009977

Medic, S., H. Rizos and M. Ziman, 2011. Differential PAX3 functions in normal skin melanocytes and melanoma cells. Biochem. Biophys. Res. Commun., 411: 832-837.

Mellick, G.D., D.D. Buchanan, P.A. Silburn, D.K.Y. Chan and D.G. Le Couteur *et al.*, 2000. The monoamine oxidase B gene GT repeat polymorphism and Parkinson's disease in a Chinese population. J. Neurol., 247: 52-55.

Mermod, N., T.J. Williams and R. Tjian, 1988. Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription *in vitro*. Nature, 332: 557-561.

Minth, C.D. and J.E. Dixon, 1990. Expression of the human neuropeptide Y gene. J. Biol. Chem., 265: 12933-12939.

Mostowska, A., A. Kobielak, B. Biedziak and W.H. Trzeciak, 2003. Novel mutation in the paired box sequence of PAX9 gene in a sporadic form of oligodontia. Eur. J. Oral Sci., 111: 272-276.

Murdoch, B., C. DelConte and M.I. Garcia-Castro, 2012. Pax7 lineage contributions to the mammalian neural crest. PLoS One, Vol. 7. 10.1371/journal.pone.0041089

Murzin, A.G., S.E. Brenner, T. Hubbard and C. Chothia, 1995. Scop A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol., 247: 536-540.

Nikitina, N., T. Sauka-Spengler and M. Bronner-Fraser, 2008. Dissecting early regulatory relationships in the lamprey neural crest gene network. Proc. Natl. Acad. Sci., 105: 20083-20088.

Niller, H.H. and L. Hennighausen, 1991. Formation of several specific nucleoprotein complexes on the human cytomegalovirus immediate early enhancer. Nucleic Acids Res., 19: 3715-3721.

Nimer, S.D., E.A. Morita, M.J. Martis, W. Wachsman and J.C. Gasson, 1988. Characterization of the human granulocyte-macrophage colony-stimulating factor promoter region by genetic analysis: Correlation with DNase I footprinting. Mol. Cell. Biol., 8: 1979-1984.

Noll, M., 1993. Evolution and role of *Pax* genes. Curr. Opin. Genet. Dev., 3: 595-605.

Notturno, F., M. Pace, M.V. De Angelis, C.M. Caporale, A. Giovannini and A. Uncini, 2008. Susceptibility to chronic inflammatory demyelinating polyradiculoneuropathy is associated to polymorphic GA repeat in the SH2D2A gene. J. Neuroimmunol., 197: 124-127.

Ogami, K., D. Kardassis, C. Cladaras and V.I. Zannis, 1991. Purification and characterization of a heat stable nuclear factor CIIIB1 involved in the regulation of the human ApoC-III gene. J. Biol. Chem., 266: 9640-9646.

Orengo, C.A., A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells and J.M. Thornton, 1997. CATH-a hierarchic classification of protein domain structures. Structure, 5: 1093-1109.

Paik, Y.K., D.J. Chang, C.A. Reardon, M.D. Walker, E. Taxman and J.M. Taylor, 1988. Identification and characterization of transcriptional regulatory regions associated with expression of the human apolipoprotein E gene. J. Biol. Chem., 263: 13340-13349.

Paulweber, B., M.A. Onasch, B.P. Nagy and B. Levy-Wilson, 1991. Similarities and differences in the function of regulatory elements at the 5'end of the human apolipoprotein B gene in cultured hepatoma (HepG2) and colon carcinoma (CaCo-2) cells. J. Biol. Chem., 266: 24149-24160.

Peers, B., M.L. Voz, P. Monget, M. Mathy-Hartert, M. Berwaer, A. Belayew and J.A. Martial, 1990. Regulatory elements controlling pituitary-specific expression of the human prolactin gene. Mol. Cell. Biol., 10: 4690-4700.

Peers, B., P. Monget, M.A. Nalda, M.L. Voz, M. Berwaer, A. Belayew and J.A. Martial, 1991. Transcriptional induction of the human prolactin gene by cAMP requires two cis-acting elements and at least the pituitary-specific factor Pit-1. J. Biol. Chem., 266: 18127-18134.

Pierpont, J.W. and R.P. Erickson, 1993. Facts on PAX. Am. J. Human Genet., 52: 451-454.

Piette, J. and M. Yaniv, 1986. Molecular analysis of the interaction between an enhancer binding factor and its DNA target. Nucleic Acids Res., 14: 9595-9611.

Postel, E.H., S.E. Mango and S.J. Flint, 1989. A nuclease-hypersensitive element of the human c-myc promoter interacts with a transcription initiation factor. Mol. Cell Biol., 9: 5123-5133.

Pridans, C., M.L. Holmes, M. Polli, J.M. Wettenhall and A. Dakic *et al.*, 2008. Identification of Pax5 target genes in early B cell differentiation. J. Immunol., 180: 1719-1728.

Pruitt, K.D., T. Tatusova and D.R. Maglott, 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res., 35: D61-D65.

Quinlan, J., M. Lemire, T. Hudson, H. Qu and A. Benjamin *et al.*, 2007. A common variant of the PAX2 gene is associated with reduced newborn kidney size. J. Am. Soc. Nephrol., 18: 1915-1921.

Read, D., T. Nishigaki and J.L. Manley, 1990. The Drosophila even-skipped promoter is transcribed in a stage-specific manner *in vitro* and contains multiple, overlapping factor-binding sites. Mol. Cell Biol., 10: 4334-4344.

Richards, A.J., J.C. Lloyd, P. Narcisi, P.N. Ward, A.C. Nicholls, A. De Paepe and F.M. Pope, 1992. A 27-bp deletion from one allele of the type III collagen gene (COL3A1) in a large family with Ehlers-Danlos syndrome type IV. Hum. Genet., 88: 325-330.

Ritchie, D.W. and V. Venkatraman, 2010. Ultra-fast FFT protein docking on graphics processors. Bioinformatics, 26: 2398-2405.

Ruiz-Llorente, S., E.C. de Pau, A. Sastre-Perona, C. Montero-Conde and G. Gomez-Lopez *et al.*, 2012. Genome-wide analysis of Pax8 binding provides new insights into thyroid functions. BMC Genomics, Vol. 13, 10.1186/1471-2164-13-147

Schule, R., M. Muller, H. Otsuka-Murakami and R. Renkawitz, 1988. Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. Nature, 332: 87-90.

Seguin C. and D.H. Hamer, 1987. Regulation *in vitro* of metallothionein gene binding factors. Science, 235: 1383-1387.

Shannon, M.F., J.R. Gamble and M.A. Vadas, 1988. Nuclear proteins interacting with the promoter region of the human granulocyte/macrophage colony-stimulating factor gene. Proc. Natl. Acad. Sci. USA., 85: 674-678.

Sjakste, T., L. Lauberte, Y. Collan, M.L. Savontaus, A. Bajare, K. Scherrer and N. Sjakste, 2002. Identification of an intronic TG repeat polymorphism in the human proteasome core particle PROS-27K gene. DNA Seq., 13: 139-143.

Smith, J.D., A. Melian, T. Leff and J.L. Breslow, 1988. Expression of the human apolipoprotein e gene is regulated by multiple positive and negative elements. J. Biol. Chem., 263: 8300-8308.

Solovyev, V. and A. Salamov, 1997. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol., 5: 294-302.

Solovyev, V.V. and I.A. Shahmuradov, 2003. PromH: Promoters identification using orthologous genomic sequences. Nucleic Acids Res., 31: 3540-3545.

Strauss, E.C. and S.H. Orkin, 1992. *In vivo* protein-DNA interactions at hypersensitive site 3 of the human beta-globin locus control region. Proc. Natl. Acad. Sci. USA., 89: 5809-5813.

Strawhecker, J.M., N.A. Betz, R.Y. Neades, W. Houser and J.C. Pelling, 1989. Binding of the 97 kDa glucocorticoid receptor to the 5' upstream flanking region of the mouse c-Ha-ras oncogene. Oncogene, 4: 1317-1322.

Sureau, A., J. Soret, M. Vellard, J. Crochet and B. Perbal, 1992. The PR264/c-myb connection: Expression of a splicing factor modulated by a nuclear protooncogene. Proc. Natl. Acad. Sci. USA., 89: 11683-11687.

Takahashi, Y., T. Nakayama, M. Soma, J. Uwabo, Y. Izumi and K. Kanmatsuse, 1997. Association analysis of TG repeat polymorphism of the neuronal nitric oxide synthase gene with essential hypertension. Clin. Genet., 52: 83-85.

Takiya, S., C.C. Hui and Y. Suzuki, 1990. A contribution of the core-promoter and its surrounding regions to the preferential transcription of the fibroin gene in posterior silk gland extracts. EMBO J., 9: 489-496.

Taylor, M.V., J.B. Gurdon, N.D. Hopwood, N. Towers and T.J. Mohun, 1991. Xenopus embryos contain a somite-specific, MyoD-like protein that binds to a promoter site required for muscle actin expression. Genes Dev., 5: 1149-1160.

Terzic, J. and M. Saraga-Babic, 1999. Expression pattern of PAX3 and PAX6 genes during human embryogenesis. Int. J. Dev. Biol., 43: 501-508.

Thompson, J.D., D.G. Higgins and T.J. Gibson, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680.

Tong, G.X., L. Chiriboga, D. Hamele-Bena and A.C. Borczuk, 2007. Expression of PAX2 in papillary serous carcinoma of the ovary: Immunohistochemical evidence of fallopian tube or secondary Mullerian system origin? Mod. Pathol., 20: 856-863.

Underhill, D.A. and P. Gros, 1997. The paired-domain regulates DNA binding by the homeodomain within the intact Pax-3 protein. J. Biol. Chem., 272: 14175-14182.

Underhill, D.A., 2012. PAX proteins and fables of their reconstruction. Crit. Rev. Eukaryot. Gene. Expr., 22: 161-177.

Uzan, G., M. Prenant, M.H. Prandini, F. Martin and G. Marguerie, 1991. Tissue-specific expression of the platelet GPIIb gene. J. Biol. Chem., 26: 8932-8939.

Van Wijnen, A.J., K.L. Wright, R.F. Massung, M. Gerretsen, J.L. Stein and G.S. Stein, 1988. Two target sites for protein binding in the promoter region of a cell cycle regulated human H1 histone gene. Nucl. Acids Res., 16: 571-592.

Walther, C., J.L. Guenet, D. Simon, U. Deutsch and B. Jostes *et al.*, 1991. Pax: A murine multigene family of paired box-containing genes. Genomics, 11: 424-434.

Walther, R. and P. Gruss, 1996. Pax and vertebrate development. Int. J. Dev. Biol., 40: 369-377.

Wang, J., S. Li, X. Xiao, P. Wang, X. Guo and Q. Zhang, 2010a. PAX3 mutations and chinical characteristics in Chinese patients with Waardenburg syndrome type 1. Mol. Vis., 16: 1146-1153.

Wang, W., J. Zhong and Y.Q. Wang, 2010b. Comparative genomic analysis reveals the evolutionary conservation of Pax gene family. Genes. Genet. Syst., 85: 193-206.

Wen, S.C., D.H. Ku, A. De Luca, P.P.Claudio and A. Giordano *et al.*, 1995. ets-2 regulates cdc2 kinase activity in mammalian cells: coordinated expression of cdc2 and cyclin A. Exp. Cell. Res., 217: 8-14.

Wilkins, M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez and K.L. Williams *et al.*, 1999. Protein identification and analysis tools in the ExPASy server. Meth. Mol. Biol., 112: 531-552.

Wilm, B., E. Dahl, H. Peters, R. Balling and K. Imai, 1998. Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. Proc. Natl. Acad. Sci. USA., 95: 8692-8697.

Wingender, E., P. Dietze, H. Karas and R. Knuppel, 1996. TRANSFAC: A database on transcription factors and their DNA binding sites. Nucleic. Acids. Res., 24: 238-241.

Wu, M.M., H.Y. Chiou, C.L. Chen, L.I. Hsu and L.M. Lien *et al.*, 2011. Association of heme oxygenase-1 GT-repeat polymorphism with blood pressure phenotypes and its relevance to future cardiovascular mortality risk: An observation based on arsenic-exposed individuals. Atherosclerosis, 219: 704-708.

Wu, M.M., H.Y. Chiou, T.C. Lee, C.L. Chen and L.I. Hsu *et al.*, 2010. GT-repeat polymorphism in the heme oxygenase-1 gene promoter and the risk of carotid atherosclerosis related to arsenic exposure. J. Biomed. Sci., 17: 70-70.

Xu, J., K.L. Thompson, L.B. Shephard, L.G. Hudson and G.N. Gill, 1993. T3 receptor suppression of Sp1-dependent transcription from the epidermal growth factor receptor promoter via overlapping DNA-binding sites. J. Biol. Chem., 268: 16065-16073.

Yu, C.Y., K. Motamed, J. Chen, A.D. Bailey and C.K. Shen, 1991. The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element *in vitro* and *in Vivo*. J. Biol. Chem., 266: 8907-8915.

Zdobnov, E.M. and R. Apweiler, 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17: 847-848.

Zhang, H., H. Chen, H. Luo, J. An and L. Sun *et al.*, 2012. Functional analysis of Waardenburg syndrome-associated PAX3 and SOX10 mutations: Report of a dominant-negative SOX10 mutation in Waardenburg syndrome type II. Hum Genet., 131: 491-503.

Zwicker, J., F.C. Lucibello, L.A. Wolfraim, C. Gross, M. Truss, K. Engeland and R. Muller, 1995. Cell cycle regulation of the cyclin A, cdc25C and cdc2 genes is based on a common mechanism of transcriptional repression. EMBO J., 14: 4514-4522.