Computational analysis of *VDR* gene and its product using Bioinformatics tools

Lokman VARISLI[1], Fuat DILMEC[2], Abdullah OZGONUL[3], Osman CEN[1]
[1]Harran University, Art and Science Faculty, Department of Biology, [2]Department of Medical Biology, [3]Department of General Surgery Sanliurfa

Biyoinformatik araçlarla *VDR* geni ve ürününün karşılaştırmalı analizi

Özet

Amaç: Bu çalışma, insan VDR molekülüne homolog protein dizileri, promotor dizileri üzerinde genel transkripsiyon faktör bağlanma yerleri, korunan bölgelerin filogenetik ilişkileri ve araştırılan ESTdb'lerin varlığında EST verilerine bağlı ekspresyon profilleri gibi genin ve onun ürünlerinin özelliklerini araştırmak amacı ile planlandı.

Gereç-Yöntem: Biz, farklı anlı türlerinde biyoinformatik araçları kullanarak ESTdb'den EST verilerine dayanarak *VDR* genin homolojisini, korunan bölgelerini, promotor ve ekspresyon profillerini inceledik.

Sonuçlar: Bu çalışmanın sonuçları, VDR molekülünün çalışılan tüm organizmalar arasında, örneğin *Macaca mulatta* (98%), *Saguinus oedipus* (98%) ve *Petromyzon marinus* (59%)'un insana benzediğini ve orta derecede korunduğunu gösterdi. ZnF_C4 (nükleer hormon reseptöründe C4 çinko ucu) ve HOLI (hormonların ligand bağlama alanı) alanlarına göre bir kaç benzer motif belirlendi. Ayrıca insanın normal ve kanserli dokuları arasında gen ekspresyonu bakımından anlamlı fark olduğu gösterildi.

Yorum: Bu çalışma, farklı canlı türlerinde *VDR* gen promotoru üzerinde herhangi bir genel transkripsiyon bağlama yerinin olmadığını kanıtladı. Yakınlık metodu (NJ)'na dayanarak oluşturulan filogenetik ağaçlar, ZnF_C4 ve HOLI dizilerinin yakın evrimsel ilişkisini gösterdi.

Anahtar kelimeler: VDR, Biyoinformatik, Karşılaştırmalı genomik, *in silico* Biyoloji

Abstract

Background: The aims of this study was to analyze some properties of this gene and its products; such as the homologous protein sequences to human VDR molecule, the common transcription factor binding sites on promoter sequences, phylogenetic relationships of the conserved domains, and *in silico* expression profiles based on EST data that presence in ESTdb in all investigated.

Methods: We investigated the homology, conserved domain, promoter and expression profiles of the *VDR* gene based on EST data from the ESTdb by using bioinformatics tools in different species.

Results: The results of this study indicated that VDR molecule is middling conserved among all studied organisms; for example, human similar to *Macaca mulatta* (98%) and *Saguinus oedipus* (98%), to *Petromyzon marinus* (59%). From the point of view of ZnF_C4 (C4 zinc finger in nuclear hormone receptors) and HOLI (Ligand binding domain of hormones) domains, it was obtained several similar motifs. Separately, it was seen that there was significantly difference gene expression in normal and cancer tissues of brain in human.

Conclusion: This study demonstrated that there was no common transcription factor binding sites in promoter *VDR* gene in different species. Phylogenetic trees constructed using the neighbor-joining method (NJ) revealed a close evolutionary relationship of ZnF_C4 and HOLI in various species.

Key words: VDR, Bioinformatics, Comparative genomics, *in silico* Biology

INTRODUCTION

The vitamin D (1, 25-dihydroxyvitamin D3) receptor (*VDR*) gene has located on 12q in the human and contains 11 exons. Besides, the *VDR* gene product is an intracellular hormone receptor that specifically binds the active form of vitamin D (1,25-dihydroxyvitamin D3 or calcitriol), interacts with target-cell nuclei to produce a variety of biologic effects, also functions as a receptor for the secondary bile acid lithocholic acid, which is hepatotoxic and a potential enteric carcinogen (1-6).

The VDR protein which is closely related to the thyroid hormone receptors contains a zinc-finger DNA-binding, transcriptional activation and a ligand-binding domains (1). *VDR* promoter DNA has a GC-rich region, several potential binding sites for the transcription factor SP1 and other activators, but not TATA (7, 8).

The overall percentage of identity between the entire human and mouse VDR genomic sequences is 28.8% and coding exons are found in a highly conserved region (86.5%–92.6% identity) (9). It has been shown that the *Xenopus laevis* VDR has a highly conserved DNA binding domain and a less conserved ligand-binding domain (LBD). The *Xenopus* VDR is 79%, 73%, 73%, and 75% identity at the amino acid level with the chicken, mouse, rat, and human VDRs, respectively (10-12).

In this study, we aimed to the determining of the phylogenetic trees, the transcription factors binding sites on promoters, the rates of

expression, and protein homologies of VDR of human using bioinformatics tools.

MATERIAL AND METHODS

Homology search

The homologous protein sequence searches, to VDR, was carried out using the BLASTp program (13, 14) at NCBI (http://www.ncbi.nlm.nih.gov) against the protein databases were performed using the amino acid sequences of Human VDR as query sequences. ZnF_C4 and HOLI conserved domain sequences of both human and other orthologous species proteins were downloaded and then amino acid sequence multiple alignments were performed using the ClustalW (15) program at EBI (http://www.ebi.ac.uk).

Promoter Analysis

We used Genomatix software (http://www.genomatix.de) for analysis of *VDR* gene promoters in various species. VDR promoters were compared among *Homo sapiens, Bos taurus, Canis familiaris, Rattus norvegicus, Mus musculus* and *Macaca mulatta* in Database of Genomatix software. We cross-examined *VDR* gene and then we were performed multiple alignment analysis by using ClustalW. Then *in silico* common transcription factor binding sites were searched through Dialign TF program in Genomatix software for all of *VDR* promoters that presence in database.

Evolutionary Analysis

We carried out amino acid sequences of ZnF_C4 and HOLI domains of VDR to construct of phylogenetic trees. The phylogenetic trees were constructed by using the neighbor-joining method (NJ) with Jones-Taylor-Thomton (JTT) distances. NJ searches were conducted by using Molecular Evolutionary Genetics Analysis (MEGA4) Software Version 4.0 (16). The reliability of internal branches was assessed by using 500 bootstrap replicates, and sites with gaps were ignored in this analysis.

In Silico Expression Analysis

We used DigiNorthern (17, 18) to analyze the expression profiles of *VDR* gene based on EST data. The DigiNorthern collects all ESTs for a query gene and categorizes these ESTs based on the types of tissues and their histological status. Pairwise comparisons of relative frequencies were performed with the Fisher's exact test using SPSS 11.0 for Windows.

RESULTS

Homology Search

BLASTp of human VDR revealed that it is found as a different protein in various vertebrate species. According to the results of BLASTp, it was determined that human VDR protein is very close to *Macaca mulatta* (98%), and *Saguinus oedipus* (98%), but *Crocodylus niloticus* (66%), and *Petromyzon marinus* (59%) (Table 1).

Table 1. BLASTp results of *VDR* gene

Species	AC Number	Protein name	Size of protein	Identity (%
Homo sapiens	NP00036	vitamin D (1,25-dihydroxyvitamin D3) receptor	427	100
Macaca mulatta	XP001095385	vitamin D (1,25-dihydroxyvitamin D3) receptor	487	98
Saguinus oedipus	AAK48863	vitamin D receptor	427	98
Canis familiaris	XP543714	similar to Vitamin D3 receptor (VDR) (1,25-dihydroxyvitamin D3 receptor)	448	92
Bos taurus	XP613129	Similar to Vitamin D3 receptor (VDR) (1,25-Dihydroxyvitamin D3 receptor)	426	90
Rattus norvegicus	NP058754	vitamin D (1,25-dihydroxyvitamin D3) receptor	423	89
Equus caballus	AAX47065	vitamin D receptor	356	89
Mus musculus	NP033530	vitamin D (1,25-dihydroxyvitamin D3) receptor	422	86
Gallus gallus	NP990429	vitamin D (1,25-dihydroxyvitamin D3) receptor	451	79
Bufo marinus	AAP22715	vitamin D receptor	342	73
Xenopus laevis	AAB58585	vitamin D (1,25-dihydroxyvitamin D3) receptor	422	72
Gekko gecko	AAP13096	vitamin D receptor	341	72
Elaphe sp.	CAC69541	putative vitamin D receptor	270	72
Oryctolagus cuniculus	AAP20884	vitamin D3 receptor	132	72
Cyprinus carpio	CAH045518	vitamin D receptor I	432	71
Danio rerio	NP570994	vitamin D receptor	453	71
Paralichthys olivaceus	BAA95016	vitamin D receptor a	420	71
Salmo salar	CAG47089	vitamin D receptor	420	70
Trachemys scripta elegans	CAC69550	putative vitamin D receptor	324	70
Crocodylus niloticus	CAB56417	vitamin D receptor	215	66
Petromyzon marinus	AAP05810	vitamin D receptor	406	59

ClustalW results of of VDR conserved domains indicated that ZnF_C4 and HOLI domains structures were especially well conserved in some species. It was determined that the ZnF_C4 domain in *Trachemys scripta elegans, Dicentrarchus labrax, Gekko gecko, Elaphe sp*. and *Bufo marinus* was the truncated. Besides, this domain is not in *Crocodylus niloticus*. In the same way, the HOLI domain is truncate in *Trachemys scripta elegans, Dicentrarchus labrax, Crocodylus niloticus, Elaphe sp*., *Bufo marinus* and *Equus caballus* (Figure 1). These similar areas in the other species probably functionally very important due to excellent conserved throughout evolution process.

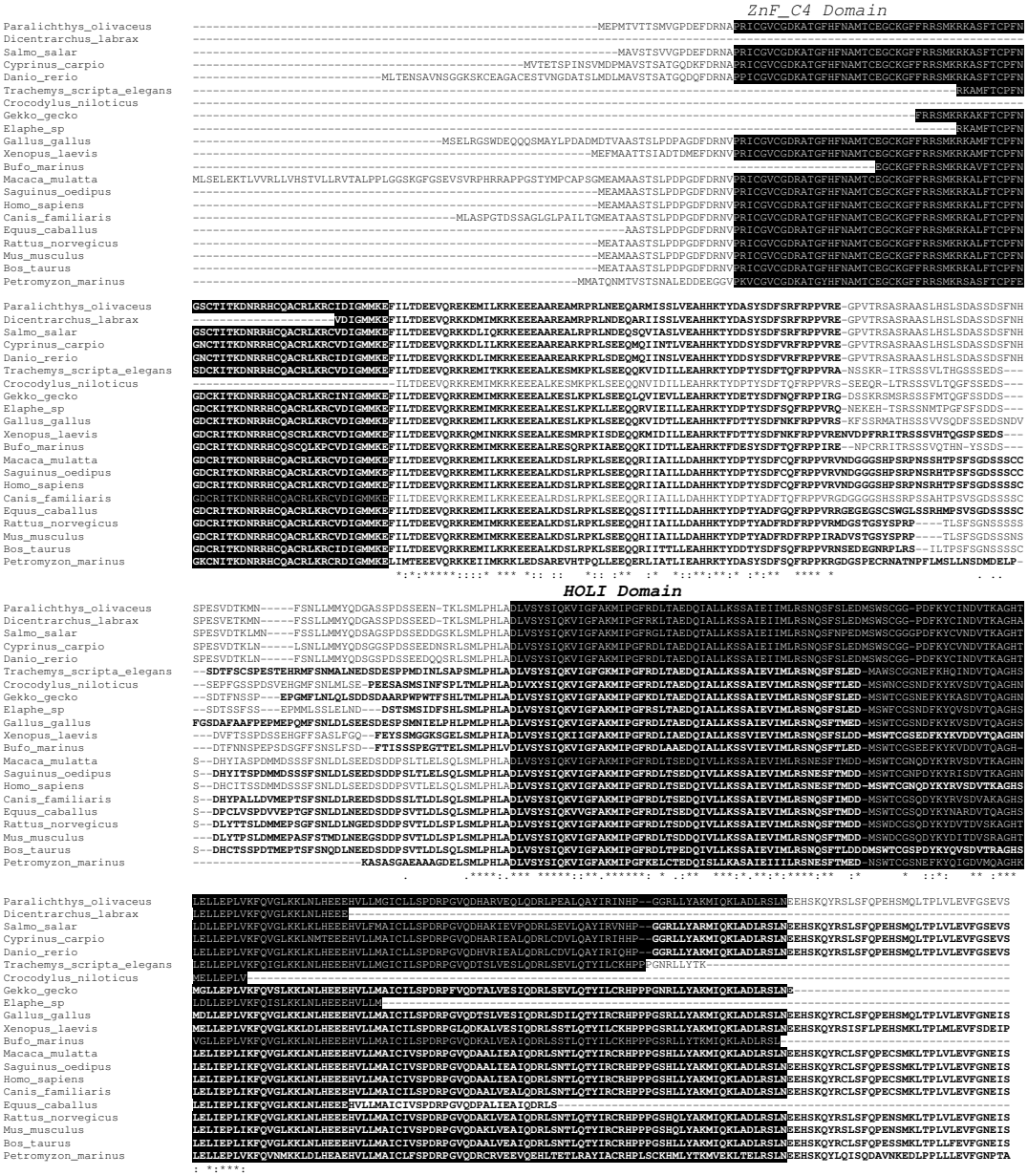

Figure 1. Multiple alignments of human VDR BLASTp results. The ZnF_C4 domain in the side of

N-terminus of VDR and the HOLI domain in the side of C-terminus of VDR are highlighted with black background. An asterisk shows the conserved amino acid residues and amino acids with similar properties are shown by under the alignment.

Promoter Analysis

By database investigations, we found that, the orthologous *VDR* gene promoters did not include any common transcription factor binding sites (TFBs) among *H. sapiens, B. taurus, C. familiaris, R. norvegicus, M. musculus, M. mullata, D. rerio* and *P. troglodytes* in Database of Genomatix software. Barely, it was found that the similarity (value 1.000) and the number of identical nucleic acids (in percentage of shorter sequence) was 86% between *Mus musculus* and *Rattus norvegius*, for each pairwise alignment, but it does not necessarily mean that these sequences are identical.

Evolutionary Analysis

From the phylogenetic trees reconstructed by MEGA3 (Figure 2) we found that VDR ZnF_C4 and HOLI domains are conserved among same organisms investigated when branch lengths take into consideration. We observed that the ZnF_C4 domains of *S. oedipus*, *C. familiaris, M. mulatta, H. sapiens, R. norvegicus, M. musculus,* and*, E. caballus* are localized in the same group (scale length 0.01) (Figure 2A). On the other hand, HOLI domain of human is similar to the HOLI domains of *M. mulatta* and *S. oedipus* in the same group (scale length 0.05) (Figure 2B). It was seen that *H. sapiens* ZnF_C4 and HOLI are far to *P. marinus.*

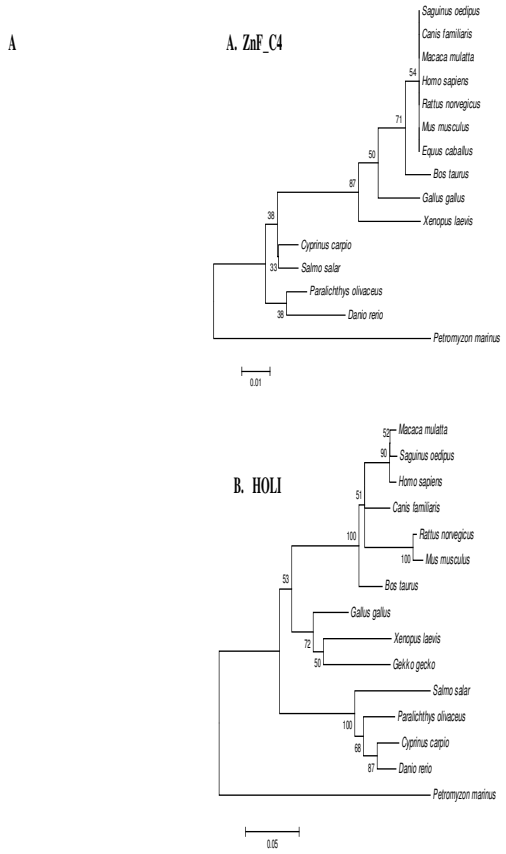

Figure 2. Phylogenetic tree of ZnF_C4 (A) and HOLI (B) domains of VDR. Phylogenetic trees were constructed by the MEGA3 program. Species names are indicated on the figure. Branch lengths indicate evolutionary relationship.

In Silico Expression Analysis

The distribution of VDR in the cDNA library database was analyzed using the DigiNorthern program. The expression profile of VDR gene in Brain tumor tissue was present at relatively higher frequency in cDNA libraries in comparison to corresponding brain normal tissues ($0.007<p$). Besides, *VDR* gene expression profiles were not significant frequencies in cDNA libraries among tumor and normal tissues of human, when these compared according to Fisher's exact test (p value <0.05) (Table 2).

Table 2. The distribution of VDR gene expression profiles in the cDNA library database

Tissue/Organ Type	Normal	Cancer	Fisher's exact test
Bone	0/7929 (0)	0/45730 (0)	>1.000
Brain	**0/257019 (0)**	**6/201219 (30)**	**0.007**
Cartilage	1/13369 (75)	3/39893 (75)	1.000
Cervix	1/1157 (864)	2/44671 (45)	0.074
Colon	0/28085 (0)	15/220946 (68)	0.401
Eye	2/85966 (23)	0/49827 (0)	0.535
Genitourinary	0/1687 (0)	0/39698 (0)	>1.000
Germ Cell	0/0 (0)	1/56605 (18)	>1.000
Head And Neck	0/55508 (0)	2/107902 (19)	0.551
Heart	7/69026 (101)	0/0 (0)	>1.000
Kidney	5/74917 (67)	1/96375 (10)	0.093
Lung	6/129822 (46)	6/207630 (29)	0.554
Lymph node	7/97096 (72)	0/54341 (0)	0.055
Lymphoreticular	0/15679 (0)	1/56791 (18)	1.000
Mammary gland	2/71315 (28)	2/124006 (16)	0.626
Muscle	2/90941 (22)	0/45799 (0)	0.554
Ovary	0/11587 (0)	17/109344 (155)	0.399
Pancreatic islet	12/95891 (125)	0/0 (0)	>1.000
Parathyroid	0/0 (0)	0/0 (0)	>1.000
Pituitary gland	1/16853 (59)	0/1598 (0)	1.000
Placenta	10/248276 (40)	0/43818 (0)	0.377
Pooled Tissue	7/373366 (19)	0/55060 (0)	0.606
Prostate	5/82545 (61)	4/81283 (49)	1.000
Salivary gland	0/414 (0)	1/20747 (0)	1.000
Skin	5/49729 (101)	4/137037 (29)	0.063
Stomach	0/26066 (0)	9/140405 (64)	0.371
Testis	2/122158 (16)	0/44649 (0)	1.000
Uncharacterized tissue	0/88784 (0)	4/105216 (38)	0.130
Uterus	2/36080 (55)	4/163186 (25)	0.298
Whole Body	2/73648 (27)	0/0 (0)	>1.000
Total number of ESTs	**79/2224913 (36)**	**68/2293776 (30)**	**0.284**

Relative frequencies are normalized per 10^6 cDNAs; p-values are for comparison of relative frequencies of VDR in normal versus tumor tissues, using the Fisher's exact test. The data for tissues with significant or suggestive significant higher or lower frequency of VDR in the tumor and normal tissues are shown in bold. (computed only for a 2x2 table, 2 cells (50.0%) have expected count less than 5. The minimum expected count is 0.05. (SPSS))

DISCUSSION

The vitamin D receptor (VDR) is a member of the nuclear hormone receptor family, which also includes retinoid, thyroid hormone, and steroid hormone receptors. These receptors function as ligand-inducible transcription factors by binding to specific DNA sequences known as hormone response elements in the promoters of the target genes (19-22).

The results of BLASTp of human VDR indicated that the VDR was found homologous to various species. According to the results of BLASTp, it was determined that human VDR protein (427 amino acids) was very close to the proteins of *Macaca mulatta* (487 amino acids) (98%) and *Saguinus oedipus* (427 amino acids) (98%), but *Petromyzon marinus* (406 amino acids) (59%) (Table 1).

On the other hand, we examined the phylogenetic trees of ZnF_C4 and HOLI conserved domains of VDR in different species using MEGA3 for Evolutionary Analysis (Figure 1). When the Figure 1 was investigated, it had been seen that human ZnF_C4 domain (at position 81-152) of VDR protein is closest to the domains of *Saguinus oedipus, Canis familiaris, Macaca mulatta, Mus musculus,* and *Equus caballus* (these species in the same group from the point of view of ZnF_C4 domain), most far to *Petromyzon marinus.* Likewise, we shown that human HOLI conserved domain (at position 294-457) was close to domains of *Macaca mulatta* and *Saguinus oedipus,* and to far *Petromyzon marinus.* These areas probably functionally were very important due to excellent conserved throughout evolution

process. At first, it has been examined VDR conserved domain among species and accused of the zinc finger and the ligand binding domains, that the VDR is well conserved throughout evolution. The DNA binding domain of the *Xenopus* VDR is 93% identical to the human VDR, whereas the rat, mouse and chicken VDRs are 99%, 99%, and 97% identical, respectively, are completely conserved (10). It is expressed that nuclear receptors (NRs) exhibited a common modular structure with an evolutionary highly conserved DNA binding domain (DBD) and a moderately conserved ligand-binding domain (LBD), which functions as a multi-functional domain (23). Our results concerning ZnF_C4 and HOLI in VDR were in agreement with previous studies demonstrating that VDR protein was characterized as conserved in moderate.

The distribution of human VDR in the cDNA library database was analyzed using the DigiNorthern program (Table 2). We designated to be different frequencies of expression of VDR mRNA in normal and cancer of tissues of human according to the results of Table 2. We compared these results with Fisher's exact test, we only determined that in the expression of human brain tumor tissue was significantly higher than normal tissue ($0.007<p$). In the other tissues, it was not statistically seen any differences between normal and cancer tissues of human. Analysis of gene refers to the detection and quantification of a gene transcript in different tissue/cells including those under different developmental, physiological and pathological conditions. The availability of the comprehensive data generated by high-throughput functional genomics approaches, mainly expressed sequence tag (EST) and serial analysis of gene expression (SAGE), provides the feasibility to study gene expression through in silico analysis (24). DigiNorthern can provide very useful preliminary result for guiding the design of further experimental analysis. Users should verify the results by experimental methods (17). A comparison of the normalized values have showed that Prostate derived ETS transcription factor (PDEF) is present at relative higher frequencies in the cDNA libraries from specific tumor types, especially from brain, breast and lung tumors with a suggestion of an increase in ovarian tumors, in comparison to those from the corresponding normal tissues (18)

We used Dialign TF program in Genomatix software for predict transcription factor binding sites (transcriptional elements) of all orthologous VDR promoters that present in the database. Dialign TF results indicated that VDR orthologous promoters had no common conserved transcriptional elements. The conservation of transcriptional elements in promoter sequences can evident of functionally conservation (25-27). The results of these studies have shown that transcriptional elements are different in various species. In promoter region of the human VDR, it is shown that there is a GC-rich island and did not a TATA box. In that respect, the human *VDR* gene is like certain other steroid receptor gene promoters (7, 28).

REFERENCES

1. Baker AR, McDonnell DP, Hughes M, et al. Cloning and expression of full-length cDNA encoding Human vitamin D receptor. Proc Nat Acad Sci. 1988; 85:3294-3298.
2. Faraco JH, Morrison NA, Baker A, Shine J, Frossard PM. *Apa*I dimorphism at the human Vitamin D receptor gene locus. Nucleic Acids Res. 1989; 17:2150.
3. Szpirer J, Szpirer C, Riviere M, et al. The Sp1 transcription factor gene (SP1) and the 1, 25-dihydroxyvitamin D (3) Receptor gene (VDR) are colocalized on human chromosome arm 12q and rat chromosome 7. Genomics. 1991; 11:168-173.
4. Labuda M, Fujiwara TM, Ross MV, et al. Two hereditary defects related to vitamin D metabolism map to the same region of human chromosome 12q13-14. J Bone Miner Res. 1992; 7:1447-1453.
5. Miyamoto K, Kesterson RA, Yamamoto H, et al. Structural organization of the human vitamin D Receptor chromosomal gene and its promoter. Molec Endocr. 1997; 11:1165-1179.
6. Makishima M, Lu TT, Xie W, et al. Vitamin D receptor as an intestinal bile acid sensor. Science. 2002; 296:1313-1316.
7. Kamei Y, Kawada T, Fukuwatari T, et al. Cloning and sequence of the gene encoding the mouse vitamin D receptor. Gene. 1995; 152:281-282.
8. Jurutka PW, Remus LS, Whitfield GK, et al. The polymorphic N terminus in human vitamin D receptor isoforms influences

transcriptional activity by modulating interaction with transcription factor IIB. Molec Endocr. 2000; 14:401-420.
9. Fang Y, van Meurs JBJ, d'Alesio A, et al. Promoter and 3-prime-untranslated-region haplotypes in the vitamin D receptor gene predispose to osteoporotic fracture: the Rotterdam Study. Am J Hum Genet. 2005; 77:807-823.
10. Li YC, Bergwitz C, Juppner H, Demay MB. Cloning and Characterization of the Vitamin D Receptor from *Xenopus laevis.* Endocrinology. 1997; 138:2347-2353.
11. Li Y, Lambert MH, Xu HE. Activation of nuclearreceptors: a perspective from structural genomics. Structure. 2003; 11:741-746.
12. Krasowski MD, Yasuda K, Hagey LR, Schuetz EG. Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I subfamily (vitamin D, pregnane X, and constitutive androstane receptors). Nuclear Receptor. 2005; 3:1-20.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403-410.
14. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 1997; 25:3389-3402.
15. Higgins D, Thompson J, Gibson T, et al. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673-4680.
16. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinf. 2004; 5:150-163.

Corresponding Author:
Fuat DILMEC, Harran University,
Medicine Faculty, Department of Medical Biology
Sanliurfa-Turkey
Tel: +90 414 312 84 56/24 06
Fax: +90 414 313 96 15
E-mail: fdilmec@harran.edu.tr

17. Wang J, Liang P. DigiNorthern, digital expression analysis of query genes based on ESTs. Bioinformatics. 2003; 19:653-654.
18. Ghadersohi A, Odunsi K, Lele S, et al. Prostate derived ETS transcription factor shows better tumor-association than other cancer-associated molecules. Oncology reports. 2004; 11:453-458.
19. Evans RM. The steroid and thyroid hormone receptor superfamily. Science. 1988; 240:889-895.
20. Glass CK, Holloway JM. Regulation of gene expression by the thyroid hormone receptor. Biochim Biophys. Acta. 1990; 1032:157-176.
21. De Luca LM . Retinoids and their receptors in differentiation, embryo-genesis, and neoplasia. FASEB J. 1991; 5:2924-2933.
22. Yu VC, Naar AM, Rosenfield MG. Transcriptional regulation by the nuclear receptor superfamily. Curr Opin Biotechnol. 1992; 3:597-602.
23. Yamada S, Yamamoto K, Masuno H. Structure-Function Analysis of Vitamin D and VDR Model. Current Pharmaceutical Design. 2000; 6:733-748.
24. Lash AE, Tolstoshev CM, Wagner L, et al. SAGEmap: a public gene expression resource. Genome Res. 2000; 10:1051-60.
25. Vanpoucke G, Goossens S, De Craene B, et al. GATA-4 and MEF2C transcription factors control the tissue specific expression of the alphaT-catenin gene CTNNA3. Nucleic Acids Res. 2004; 32:4155-4165.
26. Doan LL, Porter SD, Duan Z, et al. Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. Nucleic Acids Res. 2004; 32:2508-2519.
27. Cartharius K, Frech K, Grote K, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005; 21:2933-2942.
28. Burmester JK, Maeda N, DeLuca HF. Isolation and expression of rat 1,25-dihydroxyvitamin D receptor cDNA. Proc Natl Acad Sci USA. 1988; 85:1005-1009