American Journal of Bioinformatics Research 2019, 9(2): 56-67
DOI: 10.5923/j.bioinformatics.20190902.02

In-Silico Analysis of Three Transcription Factors Contributing to Repair and Regeneration of Lung Epithelial Progenitor Cells

Mohamed L. Salem[1,2,*], Ahmed M. Azlohairy[3], Ismail Atia[2,4], Heba Wassfy[2,5], Abdel-Aziz A. Zidan[2,6], Gábor Gyulai[7]

[1]Department of Zoology, Faculty of Science, Tanta University, Tanta, Egypt
[2]Center of Excellence in Cancer Research, Medical Campus, Tanta University Teaching Hospital, Tanta, Egypt
[3]Genetics Department, Faculty of Agriculture, Zagazig University, Egypt
[4]Zoology Department, Faculty of Science, Alazhar University, Assuite, Egypt
[5]Faculty of Engineering, Tanta University, Tanta, Egypt
[6]Faculty of Science Damanhur University, Damanhur, Egypt
[7]Institute of PhD Schools, St. Stephanus University, Gödöllö, H, Hungary

Abstract Cell differentiation processes of stem cells have unique pathways controlled by certain genes including *EYA*1 (Eyes Absent), *SIX*1 (Sine oculis homeobox), and *SOX*9 (Sry-related high mobility group BOX). It was reported that these genes act as transcriptional factors (TFs) for the repair and regeneration of the progenitor cells as well as of damaged tissues, however the function of these genes are not fully characterized and understood. Different bioinformatics tools were used in this study to analyze four transcriptional factors (TFs of *EYA*1, *SIX*1 and *SOX*9) in human and compare them with their orthologs in eleven different organisms by tools to provide more knowledge about their functions in stem cells. The present investigations showed that *SOX*9 was neither present in *Ceanorhabditis elegans* (nematode; roundworm) nor in *Cavia porcellus* (domestic guinea pig). We applied multiple sequence alignments (MSA) for selection isoforms of *SOX*9 protein showed conserved domains among different species. *EYA*1 protein showed wide range of alternative isoforms in *Homo sapiens*. High levels of *SIX*1 were detected in all human body tissues. Gene expressions were investigated to determine the expressions of genes in body tissues. The interaction between TFs proteins was confirmed by gene interaction network (GIN) which showed close relations between *EYA*1 and *SIX*1. The results of this study shed a light on the differential gene expression level of studied genes in different tissues damaged by cancer and their different repair mechanisms. The presence or absence of these TF genes may be an indicator of the developmental differentiation between invertebrates and vertebrates.

Keywords Bioinformatics' analysis, Progenitor cells, Transcription factors, Lung stem cells

1. Introduction

Stem cells are defined as undifferentiated cells which have the potential for self-renewal and develop into various cell types. In particular, lung stem cells are capable of carrying out different functions because of their ability of self-renewal and pluripotency. The lung has a much slower turnover of putative lung progenitor cells, unlike other epithelial tissues that undergo rapid regeneration [1,2].

Transcription factors plays an essential roles in regulation of gene expression [3]. TF genes of *EYA*1, *SIX*1, *SOX*9 have recently attracted the attention due to their multiple roles as transcriptional factors (TF) in controlling the transcription of DNA to mRNA by promoting or blocking the recruitment of RNA polymerase to specific genes [4]. *EYA*1-4 and *sine oculis* family genes (*SO*; which is a member of the *SIX* family of homeobox transcription factors, are unable to initiate eye development in non-retinal tissues) [5] exhibited interactions to regulate the development of many organs [6-9]. The *eyes absent* proteins act as transcriptional co-activators [10].

*EYA*1 switches *SIX*1 function from suppression to activation condition in the nucleus, resulting in transcriptional activation through recruitment of co-activators regulating precursor cell proliferation and survival during organogenesis [11]. *EYA*1 and *SIX*1 are important transcription factors for the lung epithelial

* Corresponding author:
Mohamed.labib@science.tanta.edu.eg (Mohamed L. Salem)
Published online at http://journal.sapub.org/bioinformatics

stem/progenitor cells maintenance. Mice, deficient in *EYA*1 or *SIX*1, do not have the epithelial progenitor cell markers; they highly express differentiation markers in their lungs. In addition, their lungs are markedly hypo plastic with decreased epithelial branching capacity and intensive mesenchymal cellularity [12,13].

*SIX*1 are responsible for developing skeletal muscles and supports gene expressions related to functioning of developing progenitor cells. *SIX*1 is a critical TF in skeletal muscle development, and also in eye and kidney development. Numerous studies have demonstrated that *SIX* gene family play an important role in organogenesis [11,14-16]. Moreover, *SIX*1 is able to drive the transformation of slow-twitch fibers to the fast-twitch phenotype in synergy with its cofactor *EYA*1 in muscles of adult mice [5].

*SOX*9 is a member of high mobility group-box class DNA binding protein family of transcription factors, it is required for branching morphogenesis in the lungs and for controlling alveolar epithelial progenitor cell proliferation [17,18]. Upregulation of *SOX*9 expression was reported to increase cell proliferation [19-21] associated with a poor lung ADC (Antibody-Drug Conjugate) survival [1]. However, the functional role of *SOX*9 in lung cancer has not been fully elucidated. Furthermore, the upstream oncogenic molecular pathways regulating *SOX*9 overexpression in lung ADC have not been completely delineated.

*SOX*9 is highly expressed in the distal epithelial progenitors in embryonic lung, nonetheless conditional deletion of *SOX*9 does not affect progenitor cell behavior in lung [17,22,23].

In this study, genes and proteins sequences of identified *EYA*1, *SIX*1, *SOX*9 transcriptional factors in twelve different organisms were analyzed. Constructed Phylogenetic trees of TFs protein sequences showed their cladistics relation based on their structure and function. Moreover, TFs protein domains, alternative splicing [24], expression patterns, and interaction network were analyzed to elucidate these TFs genes in human tissues. This investigations shwed the importance of these genes in stem cell differentiation and their relations in many syndromes and functional disorders. Those genes were chosen to provide more knowledge about their role not only in human but also in other living organisms to investigate their functional roles. The results of this investigation may open a new venue to repair lung defects due to cancer or other disorders.

2. Material and Methods

Data resources

The *EYA*1, *SIX*1, and *SOX*9 TF protein sequences used for analysis and phylogenetic tree construction were obtained from the Uniprot database (http://www.uniprot.org//). Transcript data used for analyzing the alternative splicing and the functional protein domains were obtained from NCBI database (http://www.ncbi.nlm.nih.gov). Expression analysis datasets were obtained from BioGPS database (http://Biogps.org/#goto).

Protein interaction networks were obtained from Gene Card human gene database (http://www.genecards.org/).

Identification of protein sequences and phylogenetic tree construction

*EYA*1, *SIX*1 and *SOX*9 TF protein sequences were obtained from the Uniprot database (http://www.uniprot.org//) and used as templates from twelve different spices To understand how different evolutionary changes of the structure did not affect conservative function of Homo sapiens (Human), *Pan troglodytes* (common chimpanzee), *Macaca mulatta* (Rhesus monkey), *Gorilla gorilla* (western gorilla), *Mus musculus* (mouse), *Rattus norvegicus* (Norway rats), *Cavia porcellus* (domestic guinea pig), *Canis lupus* (gray wolf); bird *Gallus gallus* (red junglefowl); frog *Xenopus laevis* (African clawed frog); fish *Danio rerio* (zebrafish); and roundworm *Caenorhabditis elegans* (nematode). Data were downloaded, analyzed, and phylogenetic trees were constructed using MEGA6 by the neighbor-joining method.

Protein Domains analysis and identification of alternative splicing

The functional TF protein domains of *EYA*1, *SIX*1 and *SOX*9 were obtained from NCBI database and analyzed by using the Pfam sequence tool (protein family database which contains information about protein domains and families) (https://www.ncbi.nlm.nih.gov/Structure/cdd). Alternative splicing of each gene was obtained from NCBI database.

Expression analysis in human body tissues

The gene expression analysis for TF *EYA*1, *SIX*1and *SOX*9 was obtained from BioGPS database (http://Biogps.org/#goto) representing samples of Human tissues at normal phase. More than 76 samples of tissues were measured.

Protein interaction network construction

Protein interaction network were obtained from Gene Card (Human gene database) (http://www.genecards.org/), and were illustrated the interacting proteins searched for. This search was based on published papers experimentally proven and accepted.

3. Results

Identification of TF proteins in species and phylogenetic tree construction

Different methods were used with all sequences to generate a multiple sequence alignment using [25,26]. In order to analyze TF gene sequences, we used MEGA6 program [27] to further analyze the function and conservation levels of these transcription factors, and to explain the evaluation between species. By using Neighbor-joining (NJ) method we constructed the phylogenetic tree of these genes. We assumed that the three

TF exist in all species, however, in contrast to the expectation, *SOX*9 was not found in *Caenorhabditis elegans* and *Cavia porcellus.*The same results were obtained using Phylip [28]. This observation indicated that the function of the *SOX*9 gene in this species may be regulated by other genes, or may be lost during differentiation due to their live cycle. The role of gene as sex determinate not as transcriptional factor may indicate developmental differences between vertebrates and invertebrates (Table 1).

Table 1. Accession numbers of the three transcription factor genes investigated obtained from Uniprot database

Organism	EYA1	SIX1	SOX9
Caenorhabditis_elegans	O17670	Q20938	NA
Canis_familiaris	F1Q2Y0	J9P3S7	Q7YRJ7
Cavia_porcellus	H0UUJ3	H0VTL6	NA
Danio_rerio	E9QJ32	A5WVX0	Q9DFH1
Gallus_gallus	F1NDB9	Q3C2H5	B48434
Gorilla_gorilla	G3QL19	A1YER0	G3RLX8
Homo_sapiens	Q99502	Q15475	P48436
Macaca_mulatta	F6SSJ2	A2D628	P61753
Mus_musculus	P97767	Q62231	Q04887
Pan_troglodytes	H2QWA5	A2T6X0	Q9BG89
Rattus_norvegicus	D3ZVI5	G3V970	F1LYL9
Xenopus_laevis	Q98SL6	Q9W7H9	B7ZR65

In *EYA*1, the constructed protein phylogenetic tree showed that *Rattus norvigicus* was the most distant vertebrate from *Homo sapiens*, which reveals that the function of *EYA*1 may be depleted in rat in contrast to other mammals (Fig. 1a). However, from all studied species *Caenorhabditis elegans EYA*1 protein was showed the most distal phylogenetic differences from *Homo sapiens, Gorilla gorilla* showed the closest similarity to *Homo sapiens* followed by *Pan troglodytes.*

In *SIX*1, the phylogenetic tree showed no difference from *EYA*1 phylogenetic tree except that *Danio rerio* was the most different from *Homo sapiens*. *Caenorhabditis elegans* also showed large difference from *Homo sapiens*, which may indicate that the mechanism of *SIX*1 as regulator can participate in skeletal muscle development, which may depleted in fishes and invertebrates, or regulated by other genes, which construction is confirmed by conserved domains (Fig. 1b).

The *SOX*9 phylogenetic tree showed different results. *Pan troglodytes, Macaca mulatta* were the closest to *Homo sapiens; Danio rerio, Xenopus laevis* and *Gallus gallus,* showed the largest difference from *Homo sapiens* (Fig. 1c), which may be caused due to the evolutionary differences among these species, for the evolution of mammals made them to have sex determining genes. In contrast, *Cavia porcellus* and *Caenorhabditis elegans* do not have *SOX*9 gene but we found that the function of *SOX*9 is regulated by another gene from *SOX* family. e.g SOX10 which is a gene from the *SRY* (sex determining region Y-box) group. gene function were found to be changed during evolution of this two spices, which reveals that SOX9 function may be regulated by another gene. This observation can be applied as developmental differences between vertebrates and invertebrates; however this needs more studies to confirm this hypothesis.

Protein domains analysis and identification of alternative splicing

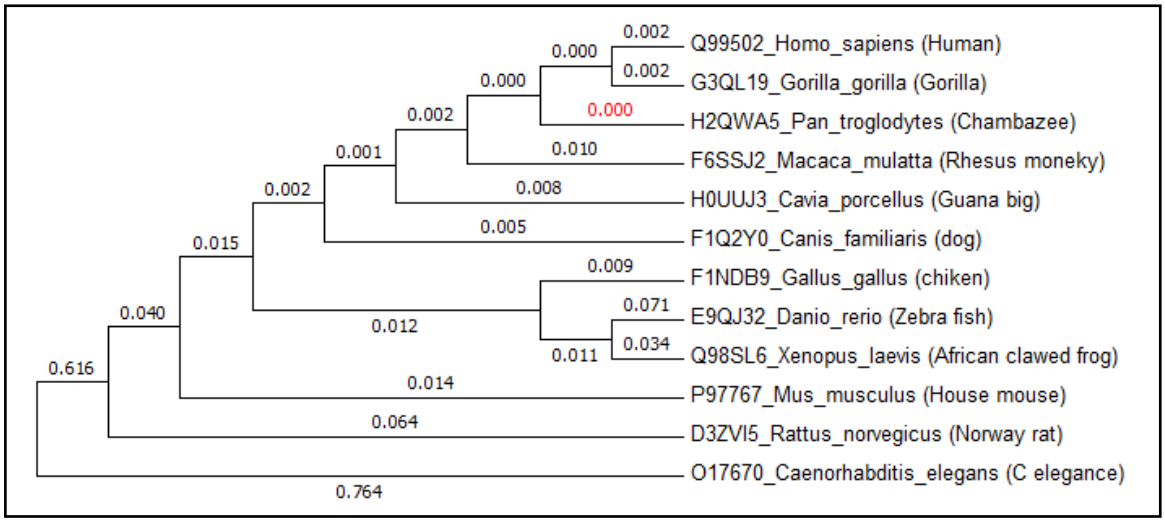
Protein domains are conserved parts of a given protein sequence that often form a functional unit [29]. Many proteins consist of several structural domains, and a given domain may appear in a variety of different proteins. Protein domains vary in length from about 25 amino acids (aa) up to 500 aa. In a multiple domain protein, each domain may fulfill its own function independently [30]. Molecular evolution research uses domains as building blocks and may recombine in different arrangements to form proteins with different functions.

Role of EYA1, SIX1 and SOX9 transcription factors in DNA repair and regeneration of lung epithelial progenitor cells

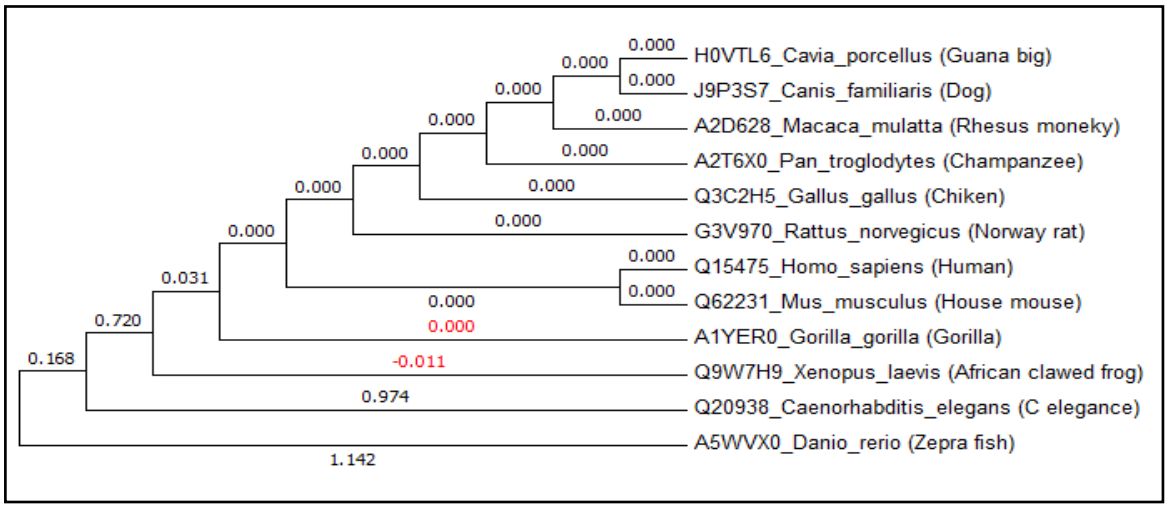
In *EYA*1, the conserved domain which shared in all species studied was found to be *EYA*-cons_domains super family [d11769]). Its length was 274 aa, and was common in all *EYA* homologs. Metazoan *EYE*'s also contain a variable N-terminal domains, consisting largely of low-similarity sequences to *Homo sapiens,* which had two conserved domains (*EYA*-cons_domains super family [d11769]) and (duf1421) (Fig. 2a).

In *SIX*1, conserved domains were found to common in the majority of species studied (Homeodomain[cd00086]), with 59 aa in length. In *Danio rerio,* it had one conserved domain (ham1[cd00515]) super family. In *Caenorhabditis elegance* there were other two conserved domains (pcl[Pfam01399] and Rpn6[cog5159]). These differences may cause the different phylogenetic tree constructions. In *Homo sapiens* a conserved domain of (Homeodomain [cd00086]) was identified (Fig. 2a).

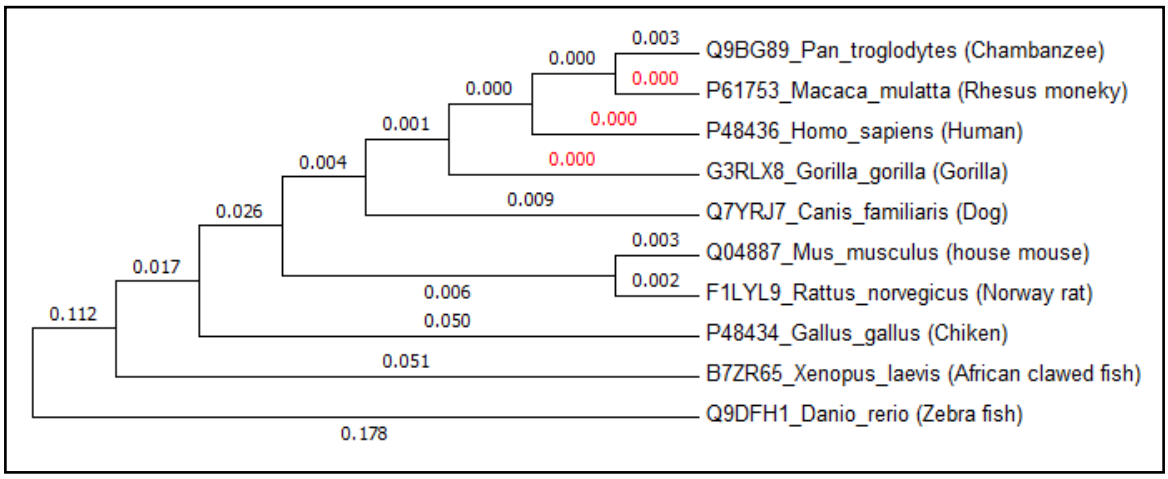
*SOX*9 conserved domains (*SOX*_N [Pfam12444] and *SOX*-TCF_*HMG*-box [cd01388]) were found to be shared in all studied species. The (*SOX*-TCF_*HMG*-box [cd01388]), with length = 72 aa, included *SRY* and its homologs in insects and vertebrates as a class I member of the *HMG*-box superfamily of DNA-binding proteins. These proteins contain a single *HMG* box, and bind to the minor groove of DNA in a highly sequence-specific manner. The ([*SOX*_N [Pfam1244]) with length = 77 aa was found in eukaryotes. This family was found in association with Pfam00505. There were two conserved sequence motifs of *YDW* and *PVR* (not presented). This family contains *SOX*8, *SOX*9 and *SOX*10 proteins which have structural similarity to *Homo sapiens* conserved domains (Fig. 2a).



(a)

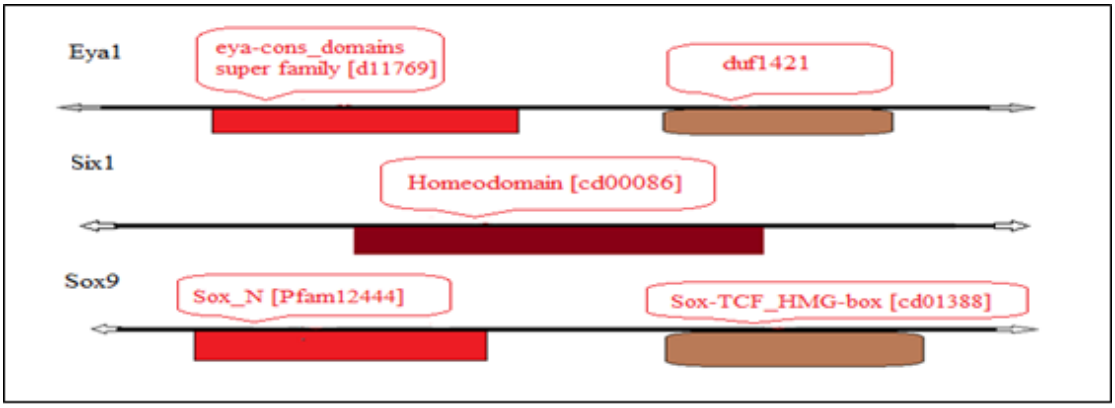


(b)

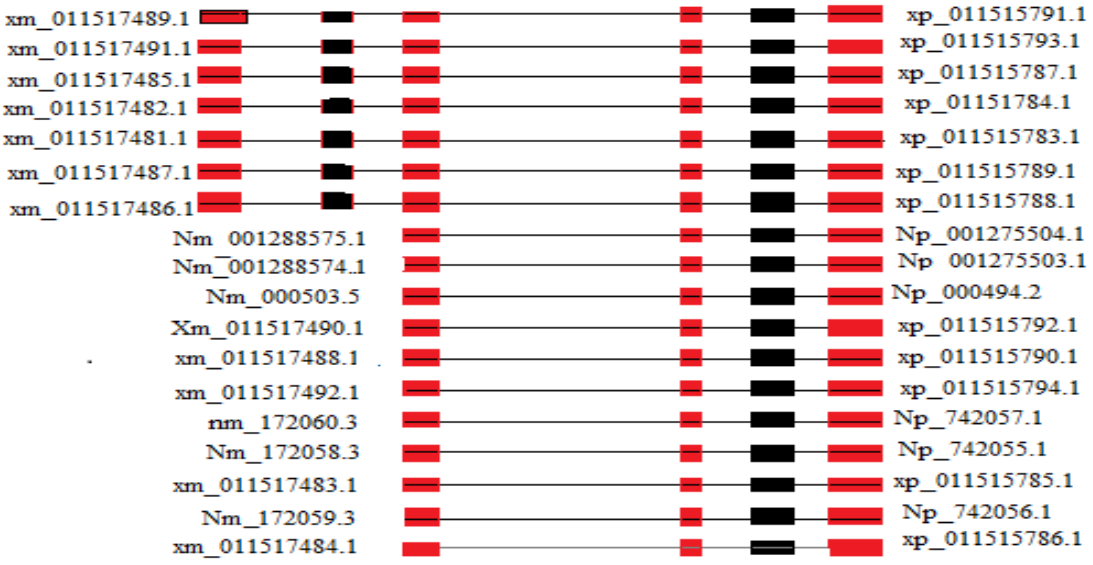


(c)

Figure 1. Protein phylogenetic trees of transcription factors of (A) *EYA*1, (B) *SIX*1 and (C) *SOX*9 of the taxa studied. The evolutionary history was inferred using the Neighbor-Joining method (37). The optimal tree with branch lengths (the sum = 0.46071988) is shown (next to the branches). The evolutionary distances were computed using the Poisson correction method (38) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 395 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (27). As *SOX*9 is missing from *Caenorhabditis elegans* and *Cavia porvellus* the tree C shows ten of the twelve taxa



(a)



(b)

(c)

(d)

Figure 2. (A) Conserved domains of transcription factors (A) *EYA*1, (B) *SIX*1, and (C) *SOX*9 of *Homo sapiens*. Functional protein domains were obtained from NCBI database, and analyzed by using the Pfam sequence tool. (B), (C) and (D) Alternatively spliced isoforms of Human TFs of (B) *EYA*1, (C) S*IX*1 and (D) *SOX*9 (data were obtained from *NCBI* database). Each line represents an individual transcript. Exons that have at least one alternative splicing are indicated by red boxes; black boxes represent unspliced exons that had no change during transcription; introns are indicated by lines. Gene Bank Accession numbers of alternatively spliced exons are indicated

Analysis of Alternative splicing

Our alternative splicing analysis of EYA1 showed that each studied species expressed a wide range of isoforms and that EYA1 genes code for multiple proteins. Of note, most of the EYA1 isoforms in Homo sapiens have the same conserved domains. The close homology of EYA1 and SIX1 could explain the wide distribution of EYA1 in Homo sapiens (Fig. 2b). Alternative isoforms of *EYA1* in *Homo sapiens* were analyzed to investigate the conserved domains of this gene. *SIX*1 showed less conserved domains in the isoforms and In *SOX*9 the alternative splicing was completely different with less isomorphs, which have been detected in species but more conserved domains were detected in these isoforms. That revealed that *SIX*1and *SOX*9 have less isoforms and less distributed proteins (table 2 a, b and c).

Table 2. The accession number of conserved domain of EYA1 isoforms, SIX1 isoforms and SOX9 isoforms obtained from Uniprot database

EYA1 isoforms	Conserved domains
xp_011515794.1 Nm_000503.5 Nm_172058.3 Np_001275504.1 Xp_011515784.1	[*EYA*-cons_domains super family (cl11769)]
Nm_172060.3,Np_000494.2,Np_742057.1,Xp_011515783.1, Xp_011515786.1,Xp_011515788.1,Xp_011515790.1, Xp_011515792.1,Np_742055.1,Nm_001288574.1, Nm_001275503.1,Xp_011515785.1,Xp_011515787.1, Xp_011515789.1, Xp_011515791.1, and Xp_011515793	[*EYA*-cons domain super family (cl11769) *EYA*-cons domain duf1421 (pfam07223)].
Xm_011517481.1, Xm_011517482.1, Xm_011517485.1, Xm_011517486.1, Xm_011517488.1, and Xm_011517492	[*EYA*-cons_domain super family (cl11769), and *EYA*-cons_domain super family (cl11769)]
Xm_011517491.1	[*EYA*-cons_domain super family (cl11769), *EYA*-cons_domain super family (cl11769), and duf1421 (pfam07223)].
In Np_742056.1	[*EYA*-cons_domain super family (cl11769), duf1421 (pfam07223), and Hydrolysis (pfam00702)].
Xm_011517491.1	[*EYA*-cons_domain super family (cl11769), duf1421 (pfam07223), and *EYA*-cons_domain super family (cl11769)]
Nm_172059.3,	[*EYA*-cons_domain super family (cl11769), Dll-N-superfamily (cl13801), and Hydrolysis (pfam00702)].
Xm_011517483.1	[*EYA*-cons_domain super family (cl11769), Dll-N-super family (cl13801), and duf1421 (pfam01223)]
Xm_011517487.1	[*EYA*-cons_domain super family (cl11769), Dll-N-super family (cl13801), and *EYA*-cons_domain super family (cl11769)].
Nm_001288575.1	[*EYA*-cons_domain super family (cl11769), Dll-N-super family (cl13801), and pha03307 (pha03307)].
Xm_011517490.1 Xm_011517490.1	[*EYA*-cons_domain super family (cl11769), duf1421 (pfam07223), and *EYA*-cons_domain super family (cl11769)].

SIX1 isoforms	Conserved domains
Np_005973.1,	[homo domain (cd00086)].
Nm_005982.3,	four conserved domains of [phao3307 (pha03307), Gt-Gtf-like (cd03784), homodomain (cd00086), and pha03307 (pha03307)].

SOX9 isoforms	Conserved domains
Nm_000346.3,	[*SOX*-tcf_*hmg*-box(cd01388),*SOX*_N(pfam12444), YppG(pfam14179), Med25_sd1(pfam11235), Duf3984super family(cl16124), pha02666super family(cl14378), Hpap super family(cl22974), TT_ORF1 super family(cl20238), Apc_basic(pfam05956), CLF(cd00832), mod_HExxH super family(cl17000), p450_rel_Gt_act(tigr04515), B_an_ocin(tigr03601) and Hpap super family(cl22974)].
In Np_000337.1	[SOX-TCF_HMG-box(cd01388)and SOX_N(pfam12444)]

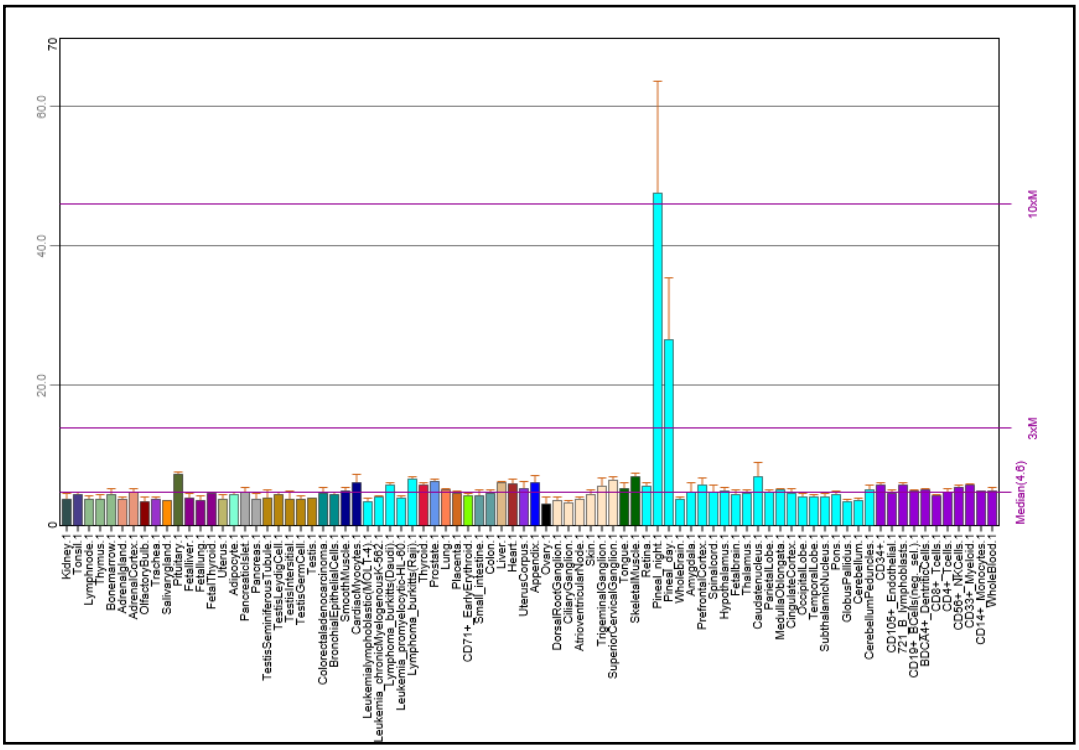
Microarray analysis of the three TFs expressed in Human tissues

We conducted systematical investigations on the expression of the three TFs in human body tissues using specific patterns of mRNA expression, which can indicate the important clues about gene functions. For microarray analysis of mRNA samples, more than 79 Human tissues from different organs were analyzed. The data was obtained from BioGPS database to explain the expression patterns.

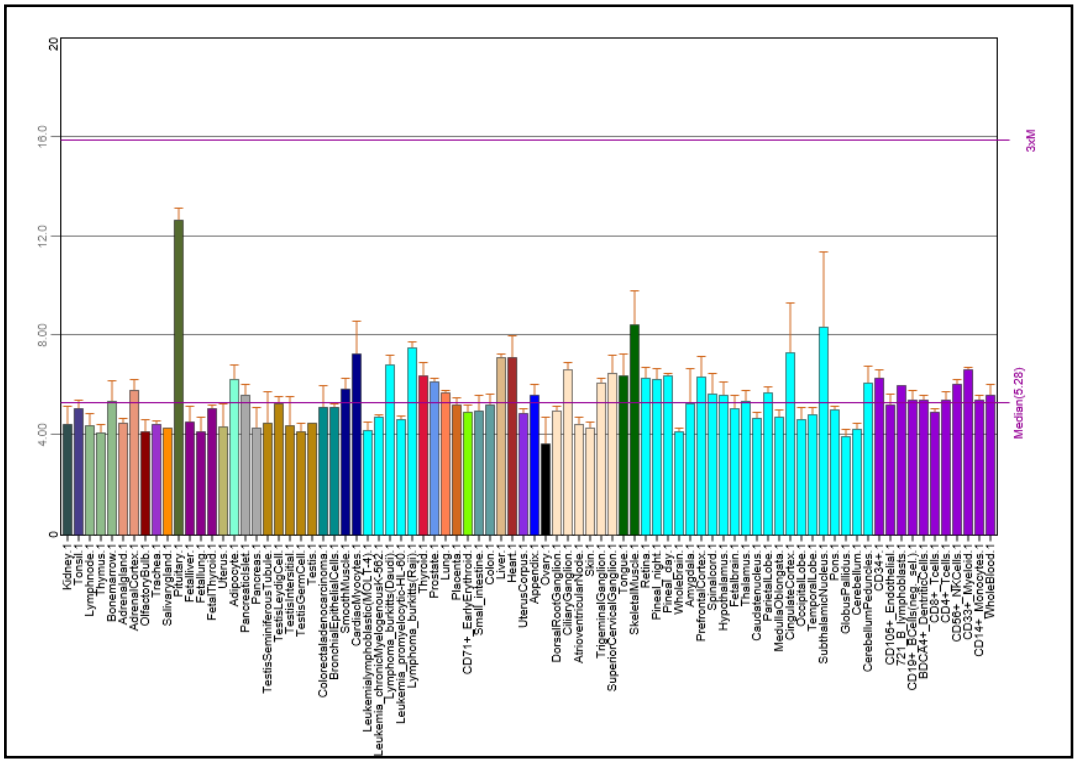
The expression *of EYA*1 gene in Human tissues and organs with Median level (M) 4.6, showed only one exception in Pineal gland, it is above 10xM (10 times above the M) in night; and above 3xM in day (Fig. 3a).

The *SX*1 expression showed balanced levels in all tissues with Median level 5.28, and with some slight exceptions in Pituitary gland (below 3xM), Skeletal Muscles (about 2xM), and Subthalamic nucleus of the brain (below 2xM) (Fig. 3b). The *SOX*9 expression showed the most complex patterns with Median level 10.1, and with several extremely high expression levels in lung's Trachea (above 10xM),

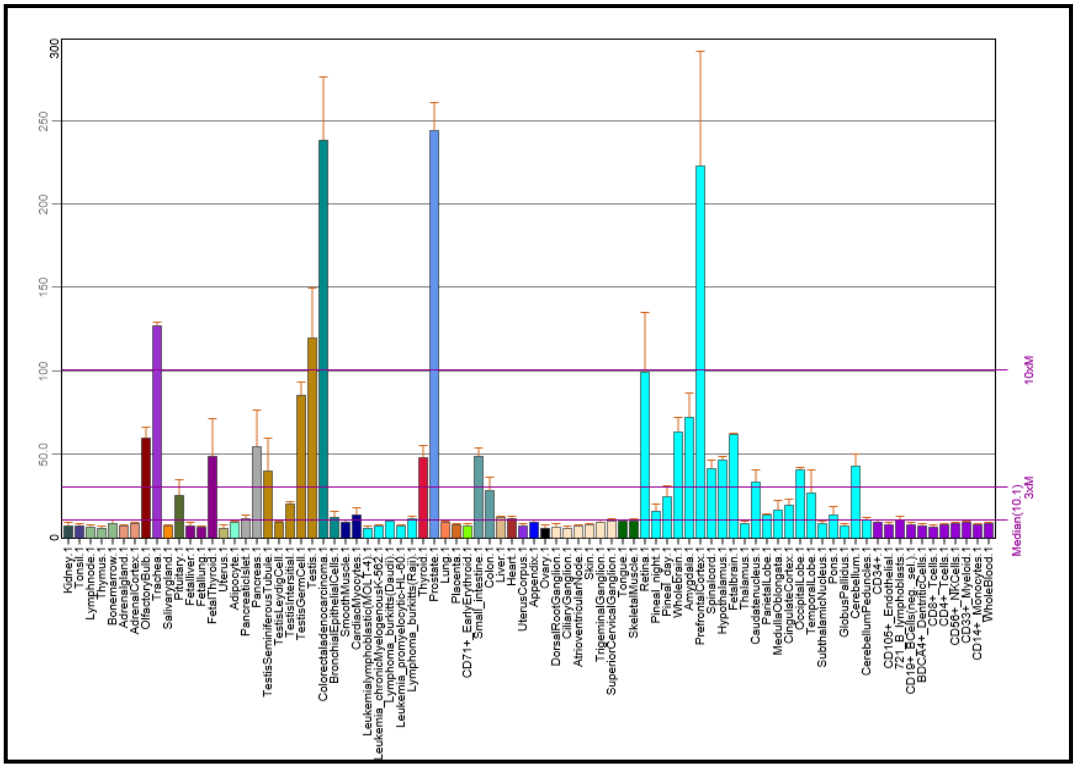
Colorectalaadenocarcinoma (above 20xM), Prostate (above 20xM), Retina (10xM), and Prefrontal Cortex of the Brain (above 20xM) (Fig. 3c). *Of note, we found high levels of SOX9 in male reproductive organs and tissues (above 10xM in testis) but with lower levels in other tissues. These data are in line with the notion that SOX9 gene contributes to sex determination in Human and that it functions with other genes to produce AMH (Anti-Mullerian Hormone) in Sertoli cells to inhibit the development of female reproductive tissues in particular ovary* (Fig. 3c).



(a)

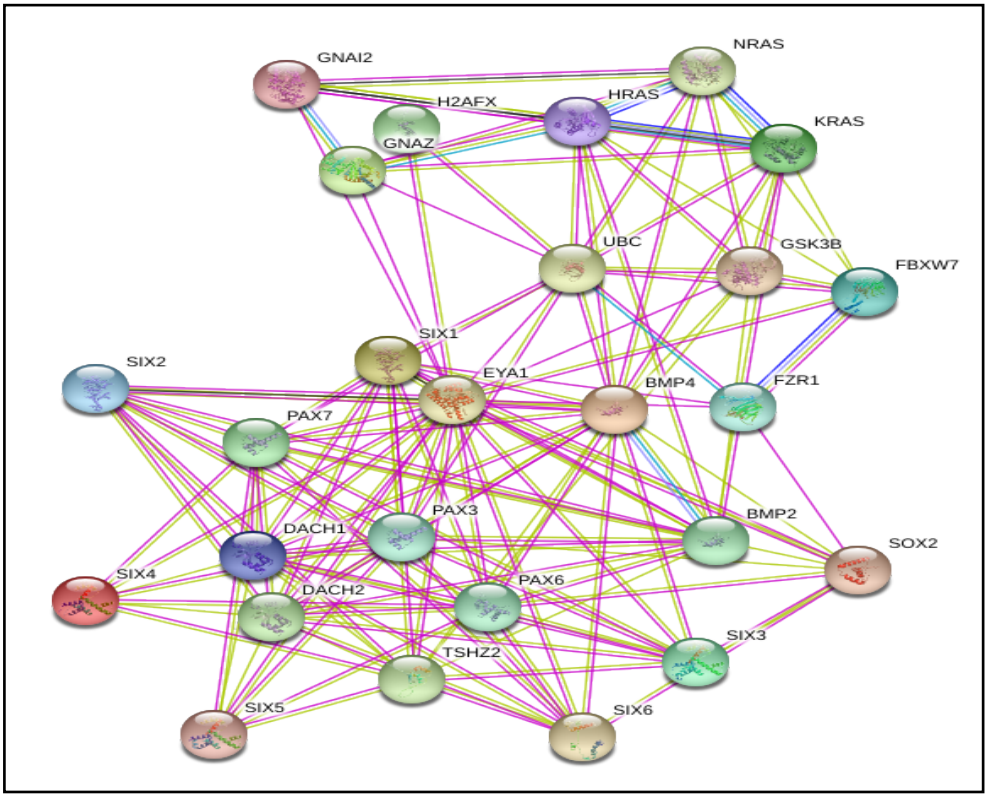


(b)

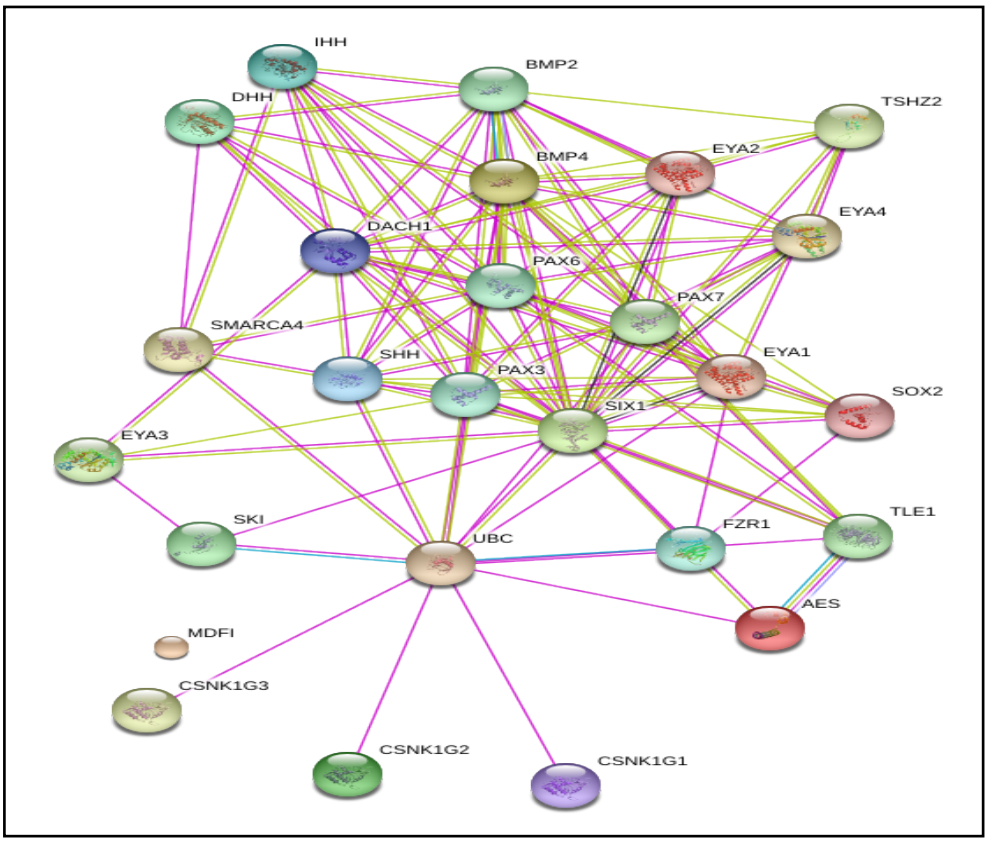


(c)

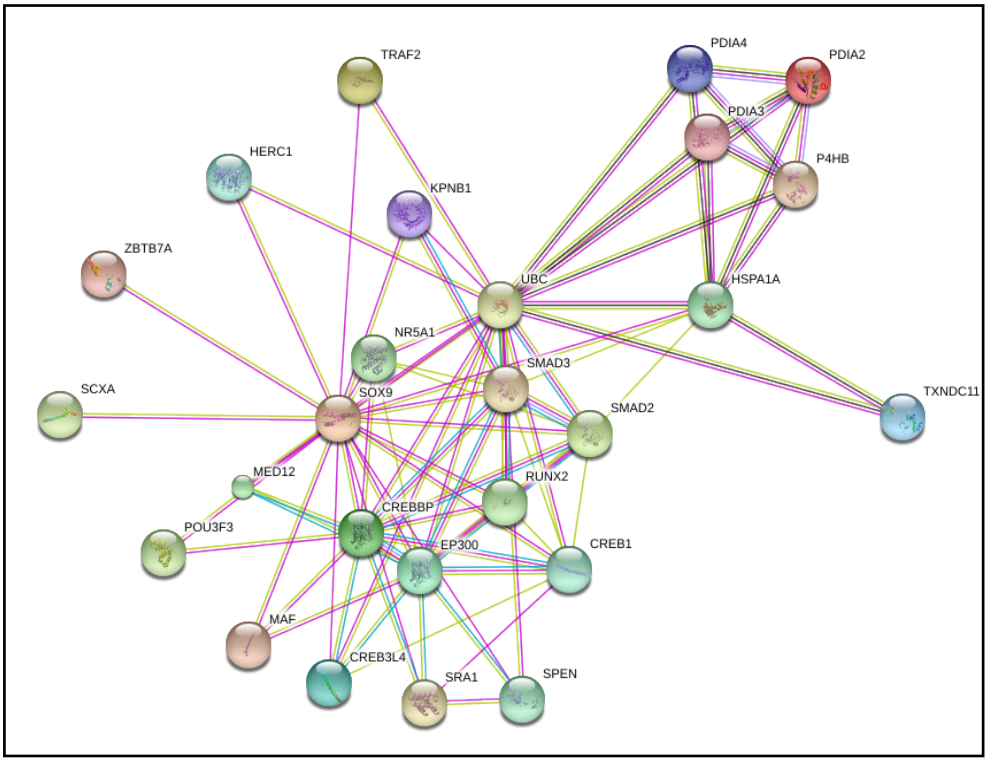
Figure 3. Microarray data of gene expression levels of the three TFs studied in human body tissues and organs of (A) *EYA*1, (B) *SIX*1, and (C) *SOX*9. The expression analysis was obtained from BioGPS database representing samples of human tissues at normal phase. More than 76 samples of Human tissues were analyzed *(39)*



(a)



(b)



(c)

Figure 4. The TF proteins interaction networks of (A) *EYA*1, (B) *SIX*1, and (C) *SOX*9. Data were obtained from Gene Card (human gene database) and illustrated the interacting proteins

Protein Interaction network constructions

The function and activity of a protein is often modulated by other proteins, that are interacting and cooperating together to perform their functions. *EYA*1 and *SIX*1 were found to have the strongest interactions.

We found that the switch of SIX1, SIX2-4, and SIX6 transcription factors from the suppressive forms form into the activated forms depends on EYA1 gene family, including EYA2, EYA3 and EYA4. For instance, EYA1 activates SIX1 function in the nucleus and turns it from the repression to activation through recruitment of co-activators. It provides a mechanism for activation of specific gene targets, including those regulating precursor cell proliferations during organogenesis [11].

*SOX*9 interaction network was found widely distributed through different families of genes of *PDIA*2 (Protein Disulfide Isomerase-Associated 2), *HERC*1 (HECT and RLD domain containing E3 ubiquitin protein ligase family member 1), *TRA*F2 (TNF receptor (TNFR) associated factor 2), *SMAD*2 (Similar to the gene products of the *Drosophila* gene 'mothers against decapentaplegic' (Mad) and the *C. elegans* gene Sma), *SRA*1 (Steroid Receptor RNA Activator 1), and *POU3F*3 (POU Class 3 Homeobox 3). The *HERC*1 were found to have different roles ranging from transcriptional activator and co-activator to intra-cellular signal transducer and transcriptional modulator (Fig. 4c). The search was based on earlier works experimentally proven and accepted [33,34].

4. Discussion

The phylogenetic trees mainly followed the evolutionary relationships among the species. We found that *SOX*9 doesn't exist in *Cavia porcellus, Caenorhabditis elegans*, which may indicate that the function of *SOX*9 is regulated by other TFs in these organisms. Conserved domains of *EYA*1, *SIX*1, and *SOX*9 were found. *EYA*1 showed more than 36 alternative isoforms, which indicated that alternative splicing is a very important source of protein diversity, and accounts for the majority of different protein functions [31,32]. Expression analysis of genes in human body tissue showed normal levels detected in *EYA*1, and *SOX*9. However, high levels of expressions were detected in sex organs. High levels of *SIX*1 expression were detected in all body tissues, which indicate that *SIX*1 is mainly responsible for all organ development in order to its somatic function. These results support the relation with function of developing progenitor cells associated with tumor progression and metastasis, which provides a mechanism for activation of specific gene targets, including those regulating precursor cell proliferation during organogenesis [11,14,35,36]. Interaction network of the three TF genes showed strong interactions, and confirmed that *EYA*1 and *SIX*1 interact in a complex way, and *SIX*1 interacts with *EYA*1 in the development of several organs in the human body.

5. Conclusions

The three transcription factors were found to have crucial roles in the development and function of organs in animal bodies. Focusing to lung, our results may provide new knowledge about these TF genes and open new prospects to further research which can help to understand repairs of lung defects in order to cancer or other diseases.

ACKNOWLEDGEMENTS

This work has been supported by a grant (ID# 5245) funded from the Science and Technology Development Fund (STDF), Ministry of Scientific Research, Egypt to prof. Mohamed L. Salem, the Principal investigator of this project.

REFERENCES

[1] Lu Y, Okubo T, Rawlins E, Hogan BL. Epithelial progenitor cells of the embryonic lung and the role of microRNAs in their proliferation. Proceedings of the American Thoracic Society. 2008; 5(3): 300-4.

[2] Green MD, Huang SX, Snoeck HW. Stem cells of the respiratory system: from identification to differentiation into functional epithelium. BioEssays: news and reviews in molecular, cellular and developmental biology. 2013; 35(3): 261-70.

[3] Saleh MM, Alzohairy, A. M., Mohamed, O. A., & Alsayed, G. H. A Comprehensive Study by Using Different Alignment Algorithms to Demonstrate the Genetic Evolution of Heat Shock Factor 1 (HSF1) in Different Eukaryotic Organisms. progressive. (2013); 3((2)).

[4] Babaie Y, Herwig R, Greber B, Brink TC, Wruck W, Groth D, et al. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells. 2007; 25(2): 500-10.

[5] Grifone R, Laclef C, Spitz F, Lopez S, Demignon J, Guidotti JE, et al. Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. Mol Cell Biol. 2004; 24(14): 6253-67.

[6] Xu PX, Woo I, Her H, Beier DR, Maas RL. Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. Development. 1997; 124(1): 219-31.

[7] Xu PX, Cheng J, Epstein JA, Maas RL. Mouse Eya genes are expressed during limb tendon development and encode a transcriptional activation function. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94(22): 11974-9.

[8] Ford HL, Kabingu EN, Bump EA, Mutter GL, Pardee AB. Abrogation of the G2 cell cycle checkpoint associated with overexpression of HSIX1: a possible mechanism of breast carcinogenesis. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(21): 12608-13.

[9] Coletta RD, Christensen K, Reichenberger KJ, Lamb J, Micomonaco D, Huang L, et al. The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(17): 6478-83.

[10] Jemc J, Rebay I. The eyes absent family of phosphotyrosine phosphatases: properties and roles in developmental regulation of transcription. Annual review of biochemistry. 2007; 76: 513-38.

[11] Li X, Oghi KA, Zhang J, Krones A, Bush KT, Glass CK, et al. Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. Nature. 2003; 426(6964): 247-54.

[12] Eliasz S, Liang S, Chen Y, De Marco MA, Machek O, Skucha S, et al. Notch-1 stimulates survival of lung adenocarcinoma cells during hypoxia by activating the IGF-1R pathway. Oncogene. 2010; 29(17): 2488-98.

[13] Chen Y, Li D, Liu H, Xu H, Zheng H, Qian F, et al. Notch-1 signaling facilitates survivin expression in human non-small cell lung cancer cells. Cancer biology & therapy. 2011; 11(1): 14-21.

[14] Laclef C, Hamard G, Demignon J, Souil E, Houbron C, Maire P. Altered myogenesis in Six1-deficient mice. Development. 2003; 130(10): 2239-52.

[15] Fujimoto Y, Tanaka SS, Yamaguchi YL, Kobayashi H, Kuroki S, Tachibana M, et al. Homeoproteins Six1 and Six4 regulate male sex determination and mouse gonadal development. Developmental cell. 2013; 26(4): 416-30.

[16] Ahmed M, Xu J, Xu PX. EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. Development. 2012; 139(11): 1965-77.

[17] Rockich BE, Hrycaj SM, Shih HP, Nagy MS, Ferguson MA, Kopp JL, et al. Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(47): E4456-64.

[18] Turcatel G, Rubin N, Menke DB, Martin G, Shi W, Warburton D. Lung mesenchymal expression of Sox9 plays a critical role in tracheal development. BMC biology. 2013; 11: 117.

[19] Matheu A, Collado M, Wise C, Manterola L, Cekaite L, Tye AJ, et al. Oncogenicity of the developmental transcription factor Sox9. Cancer research. 2012; 72(5): 1301-15.

[20] Wang L, He S, Yuan J, Mao X, Cao Y, Zong J, et al. Oncogenic role of SOX9 expression in human malignant glioma. Med Oncol. 2012; 29(5): 3484-90.

[21] Jiang SS, Fang WT, Hou YH, Huang SF, Yen BL, Chang JL, et al. Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. Clinical cancer research: an official journal of the American Association for Cancer Research. 2010; 16(17): 4363-73.

[22] Perl AK, Wert SE, Loudy DE, Shan Z, Blair PA, Whitsett JA. Conditional recombination reveals distinct subsets of epithelial cells in trachea, bronchi, and alveoli. American journal of respiratory cell and molecular biology. 2005; 33(5): 455-62.

[23] Liu Y, Hogan BL. Differential gene expression in the distal tip endoderm of the embryonic mouse lung. Gene expression patterns: GEP. 2002; 2(3-4): 229-33.

[24] Kataoka F, Umeyama, N., Nishijima, T., Tonegawa, H., Susaki, S., Okamoto, H., ... & Tanaka, K. . U.S. Patent Application No. 29/559,489. (2017).

[25] Ahmed M. Alzohairy GG, Mohamed F. Ramadan, Sherif Edris, J.S.M. Sabir, Robert K. Jansen, Ahmed Bahieldin. . Retrotransposon-based molecular markers for assessment of genomic diversity. Functional Plant Biology. (2014); 41(8) 781-789 ((IF: 2.471).).

[26] Mansour A JATdS, Gábor Gyulai . Assessment of molecular (dis)similarity: The role of multiple sequence alignments (MSA) programs in biological research. Genes, genomes and genomics 2009; (Special Issue 1): 23-30 ((Print ISSN 1749-0383) (Bioinformatics SI).).

[27] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30(12): 2725-9.

[28] Mansour.M.A. Phylip and Phylogenetics (Research note) Genes, genomes and genomics. (Bioinformatics SI). 2009; (Volume 3, 2009) ((Special Issue 1): 46-49): (Print ISSN 1749-0383).

[29] Chothia C. Proteins. One thousand families for the molecular biologist. Nature. 1992; 357(6379): 543-4.

[30] Ghelis C, Yon JM. [Conformational coupling between structural units. A decisive step in the functional structure formation]. Comptes rendus des seances de l'Academie des sciences Serie D, Sciences naturelles. 1979; 289(2): 197-9.

[31] Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry. 2003; 72: 291-336.

[32] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics. 2008; 40(12): 1413-5.

[33] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic acids research. 2013; 41(D1): D808-D15.

[34] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research. 2014:gku1003.

[35] Boucher CA, Carey N, Edwards YH, Siciliano MJ, Johnson KJ. Cloning of the human SIX1 gene and its assignment to chromosome 14. Genomics. 1996; 33(1): 140-2.

[36] Grifone R, Demignon J, Houbron C, Souil E, Niro C, Seller MJ, et al. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development. 2005; 132(9): 2235-49.

[37] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4(4): 406-25.

[38] Pauling ZEa. Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by Baryon and H.Q. Vogel. Academic Press, New York. (1965);, PP. 97-166.

[39] Brohee S, Van Helden J. Evaluation of clustering algorithms for protein-protein interaction networks. BMC bioinformatics. 2006; 7(1): 1.