HANEIN-1, A NOVEL CONSERVED EUKARYOTIC PROTEIN UBIQUITOUSLY EXPRESSED IN HUMAN TISSUES

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

BY
SERAP ERKEK
AUGUST 2008

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist.Prof. Dr. Cengiz Yakıcıer

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Elif Erson

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist.Prof.Dr. Uygar H. Tazebay

Approved for the Institute of Engineering and Science

Director of Institute of Engineering and Science
Prof. Mehmet Baray

ABSTRACT

HANEIN-1, A NOVEL CONSERVED EUKARYOTIC PROTEIN UBIQUITOUSLY EXPRESSED IN HUMAN TISSUES

Serap Erkek
M.Sc. in Molecular Biology and Genetics
Supervisor: Assist. Prof. Dr. Uygar H. Tazebay
August 2008, 52 Pages

HANEIN-1 was first identified in a study investigating 3' transcriptional regulatory elements of the Na^+/I^- symporter gene. The protein is highly conserved among eukaryotes and bears no domain similarities with any known proteins. In databases, it is described as coiled coil containing-124 hypothetical protein and predicted to encode a 223 amino acid-protein. In this project, our aim was to characterize this highly conserved protein by using several bioinformatics and biochemical methodologies. Sequence similarity search analysis showed that it had around 75% identity in mammals, 50% identity with insects and nematodes. Expressional analysis revealed that *HANEIN-1* was expressed in all tissues ubiquitously with a remarkable expression status in skeletal muscle. Beside providing information about expression status of *HANEIN-1*, northern blotting showed that *HANEIN-1* transcript size was approximately 1000 bp. Regarding protein level expression, western blotting revealed that HANEIN-1 encoded a 33 kDa protein and protein stability was affected in a different way upon labeling with Flag epitope at N-ter and C-ter. Yeast double hybrid screening performed in our laboratory showed that HANEIN-1 interacted with RASGEF1B, which was a guanine nucleotide exchange factor not fully characterized. Expressional analysis displayed that *RASGEF1B* expression profile inversely correlated with *HANEIN-1*. Finally, serine-scanning mutagenesis analysis showed that site-directed mutagenesis of serine at position 194 significantly affected the stability of the protein.

ÖZET

HANEIN-1, TÜM İNSAN DOKULARINDA İFADESİ GÖRÜLEN KORUNMUŞ YENİ BİR ÖKARYOTİK PROTEİN

Serap Erkek
Moleküler Biyoloji ve Genetik Yüksek Lisans Derecesi
Tez Yöneticisi: Yard. Doçent Dr. Uygar H. Tazebay
Agustos 2008, 52 Sayfa

HANEIN-1 Na^+/I^- simporter geninin 3' regulasyon elementleri araştırılırken keşfedilen yeni bir gendir. Bu yeni protein bütün ökaryot canlılarda korunmakla birlikte, bilinen hiçbir proteinle ortak işlevsel bölgelere sahip değildir. Veritabanlarında kıvrım-kıvrım (coiled-coil) bölge içeren-124 hipotetik proteini olarak tanımlanmakta ve 223 amino asit içeren bir protein olduğu tahmin edilmektedir. Bu projededeki amaç HANEIN-1 proteinini çeşitli biyoinformatik ve biyokimyasal analiz yöntemleriyle karakterize etmekti. HANEIN-1 proteinin amino acid dizi benzerliği analizi, bu proteinin memelilerde %75, böcek ve nematodlarda ise % 50 oranında korunduğunu gösterdi. Gen ifade analiz yöntemleri ile *HANEIN-1* ifadesinin en fazla iskelet kasında olduğu tespit edildi. Bunun yanı sıra, northern blot analizi ile *HANEIN-1* transcript büyüklüğü yaklaşık 1000 bp olarak belirlendi. Western blot tekniği ile protein düzeyinde yapılan çalışmalar, *HANEIN-1* geninin 33 kDa büyüklüğünde bir proteini kodladığını ve proteini N-son ucu ve C-son ucu Flag epitopu ile işaretlemenin protein stabilitesini farklı şekilde etkilediğini gösterdi. Laboratuvarımızda yapılan maya ikili-hibrid deneyleriyle, HANEIN-1 proteinin henüz tam olarak karakterize edilmemiş, bir guanin nükleotit değişim faktörü (Guanine Exchange Factor) olan RASGEF1B proteiniyle etkileştiği bulunmuştur. Gen ifade analizleri *RASGEF1B* ve *HANEIN-1* ifade profilleri arasında ters bir bağlantı olduğunu göstermiştir. Son olarak, serin-tarayıcı mutasyon analizleriyle, 194. konumdaki serin mutasyonunun protein stabilitesini önemli ölçüde değiştirdiği saptanmıştır.

DEDICATION PAGE

To my family

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Assist. Prof. Dr. Uygar Tazebay for his wonderful supervision of this project and his support at every respect.

I would like to thank Dr. Hani Alotaibi for his excellent support at every respect especially during the planning of the experiments and for his friendship.

I would like to thank Elif Yaman and Pelin Telkoparan for their technical support and great friendship.

I would like to thank Nilgün Taşdemir for providing us breast carcinoma cell lines cDNA and *GAPDH* expression data, Prof. Dr. Mehmet Öztürk's group for providing us β- actin primers and Assist. Prof. Dr. Elif Erson for providing us *GAPDH* primers to carry out northern blot analysis.

I would like to thank all members of Molecular Biology and Genetics Department for the wonderful atmosphere, especially Aydan Bulut and Hande Koçak for their great friendship.

I would like to thank my family for their invaluable moral support and patience.

Finally, I thank TUBITAK for providing me a scholarship throughout my M.Sc. studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Human Genome Project was completed in 2003 and there are >600 completely sequenced genomes of cellular organisms (Liolios *et al.*, 2006). However, only 50-60 % of genes have been annotated in most completely sequenced genomes. The remaining genes are either homologous of genes of unknown function, or genes which do not have any known homologs and named as "hypothetical" (Sivashankari *et al.*, 2006). Therefore, characterization of genes whose functions are unknown is important in order to understand an organism's biology. In fact, ENCODE (ENCyclopedia Of DNA Elements) project which was started in September 2003 by National Human Genome Research Institute is one of the major projects which aims to characterize human genome sequence and it aims identification of both the function of gene products and non-coding functional elements. However, even with this project only 1 % of human genome was functionally analyzed up to now (ENCODE project consortium, 2007). Therefore, it seems that understanding our genome is sophisticated and the development of novel experimental and computational tools will play a crucial role in characterization of genes.

1. 1. Functional Analysis of Novel Gene Products

Identification of the function of gene products may be performed by using experimental and bioinformatics methodologies. Currently, 20%, 7%, 10%, and 1% of annotated proteins in *Homo sapiens, Mus musculus, Drosophila melanogoster* and *Caernohabditis elegans* genomes, respectively, have been experimentally characterized. Experimental methods' being time-consuming and costly limits their usage and makes computational methods for prediction of gene function more

attractive (Lee *et al.*, 2007). However, experimental characterization of gene products is more reliable since experiments provide direct evidences at the cellular level. Bioinformatics tools may be utilized in order to have general insights about the possible functions of the gene product. And then computational data gives clues about how to plan functional analysis experiments.

1.1.1. Bioinformatics Analysis

One of the major computational strategies to identify gene function is based on sequence similarity. At this point, sequence database searching programs such as FASTA (Pearson *et al.*, 1988) and PSI-BLAST (Altschul *et al.*, 1997) are among most classical examples. In addition to these, there are certain methods based on comparative genomics. Here, it is assumed that proteins that function together either in a metabolic pathway or in structural complex are expected to evolve together (Sivashankari *et al.*, 2006). Functional linkage is identified by "Phylogenetic Profiling". In this technique, phylogenetic profile is a string with one bit and 'n' entries, where n is the number of genomes under consideration. If the n^{th} genome contains a homolog for the protein then the n^{th} entry is represented as unity in the phylogenetic profile. These profiles are clustered to determine which proteins have common profiles (Pellegrini *et al.*, 1999). Also there are some methods based on clustering approaches, which assume that genes of the same cluster carry out similar functions and genome context methods which predict functional associations between proteins such as physical interactions, or co-membership in pathways (Sivashankari *et al.*, 2006). Regarding the genome context methods, one of the most important databases is STRING, which is a database of known and predicted protein-protein interactions (von Mering *et al.*, 2007).

In order to carry out pattern and profile searches, there are also several tools such as PROSCAN (Combet *et al.*, 2003), MotifScan (Pagni *et al.*, 2004), and ELM (Puntervoll *et al.*, 2003), which provides information about protein domain / families, protein motifs recognized by other proteins and some biochemical characteristics

from the input protein sequence. In addition to these, SMART (Schultz *et al.*, 1998) is another valuable tool for identification of protein domains.

In addition to these, it is possible to predict most of the post translational modifications of protein. For instance, NetPhos (Blom *et al.*, 1999), NetNGlyc (Gupta *et al.*, 2004), SUMOPlot (Xue *et al.*, 2006) predict Ser/Thr/Tyr phosphorylation sites, N-glycosylation sites and Sumo attachment sites, respectively.

Expression of novel genes can be investigated by microarray databases such as Oncomine database (Rhodes *et al.*, 2004) and Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002). For instance, Oncomine research platform provides information about the expression status of a gene in cancerous and normal tissue as well as expression profile of the gene in cases with molecular alteration in a signaling pathway.

1.1.2. Sub-cellular Localization Analysis

Identification of sub-cellular localization of novel gene products could provide information about its function to some extent since function of a protein is often correlated with its localization. That means that if the cellular organelle or compartment where the protein localized is known, this information could help the researcher to establish a hypothesis about the function of the protein.

There are several strategies to find out a protein's localization in cell. One of the most-widely used methodologies is tagging proteins with GFP (Tsien *et al.*, 1998) by using fluorescence microscopy. As an alternative to GFP labeling, a protein can be labeled with specific epitopes such as FLAG and fluorescent microscopy analysis may be performed by monoclonal antibodies developed for epitopes and then following detection by secondary antibodies attached to fluorophores. A second useful strategy to identify sub-cellular localization is fractionation of cells. In this technique, cells are fractionated in a sucrose gradient by ultracentrifugation and

protein constituents of cellular compartments are identified by 2D gel electrophoresis and mass spectrometry (Mueller *et al.*, 2006).

1.1.3. Biochemical Analysis

1.1.3.1. Protein Purification

Purification of new gene products may be important in terms of using the purified protein in further biochemical analysis, identification of 3D structure of the protein and production of monoclonal antibodies specific to novel protein. For this purpose, protein affinity tags are used since they have high affinity and selectivity for binding to specific resins to facilitate purification, and fusing the protein of interest to glutathione S-transferase (GST) tag is one of the most popular strategies (Nilsson *et al.*, 1997).

1.1.3.2. Post-translational Modifications

Phosphorylation is a post-translational modification which might have a particularly important role in cellular signaling and enzyme activity. Therefore, determination of phosphorylation status of an unknown protein can provide important clues about its function. Phosphorylated proteins can be identified by methods based on 2D gel electrophoresis (anti-phospho-amino acid antibodies or phosphatase treatment) (Mueller *et al.*, 2006). Besides, phosphorylation sites may be determined via phospo-specific cleavage analysis followed by mass spectrometry (Knight *et al.*, 2003).

It is known that nearly half of all known proteins are glycosylated (Apweiler *et al.*, 1999) and attachment of oligosaccharides to proteins regulate a variety of processes from protein folding to cellular communication. Concerning the glycosylation pathway, nucleotide-sugar donors prepared from monosaccharides enter the endoplasmic reticulum or Golgi lumen through the action of specific antiport transporters and they are attached to proteins by glycosyltransferases. This process is a typical O-linked (serine- or threonine-attached) glycoprotein

biosynthesis. For N-linked (asparagines-attached) glycoproteins, a core oligosaccharide is assembled in the cytosol, then transported to ER and attached to proteins by glycosyltransferases (Yarema *et al.*, 2001). Glycosylation status of the proteins may be studied by western blotting (Tazebay *et al.*, 2000), mass spectrometry (Dell *et al.*, 2001), NMR (Manzi *et al.*, 2000) and liquid chromatography for glycan sequencing (Blixt *et al.*, 2004).

Proteolytic cleavage is one of non-reversible post-translational modifications and it plays an essential role in sub-cellular localization and activity of the protein. For instance, insulin is initially synthesized as a precursor and mature hormone is produced after a series of proteolytic cleavages (Duckworth *et al.*, 1979). Another important example may be the p53 protein whose level is controlled by its degradation, which is mediated by Mdm2 targeted ubiquitination of p53 (Haupt *et al.*, 1997 and Kubbutat *et al.*, 1997). Proteolytic cleavage of the proteins may be identified via isotope-coded affinity tag (ICAT) labeling and 2D gel electrophoresis followed by MS/MS (Gevaert *et al.*, 2003). Also performing a western blot analysis could be useful particularly if antibodies recognizing termini epitopes are available (see below).

1.1.4. Immunological Analysis

In order to analyze the proteins immunologically, first an antibody which specifically recognizes an epitope of the protein is raised. Antibodies generated after immunization in the laboratory are a mixture of different specificities and affinities, so they are polyclonal. At this point monoclonal antibody production may be important with respect to immunological characterization of the protein since it increases specificity. In this technique, spleen cells from an immunized mouse are fused to cells of a mouse myeloma and hybridomas are produced. Since each hybridoma is a clone derived from fusion with a single B cell, all the antibody molecules it produces are identical so they are monoclonal (Köhler *et al.*, 1992). Produced antibody may be used to identify tissue expression pattern via western

blotting, immuno-histochemistry and several other techniques. Furthermore, it may be determined whether the protein is secreted by using the specific antibody in an ELISA assay.

The Atlas of Protein Expression project and the Swedish Human Protein Atlas (Uhlen *et al.*, 2005) try to annotate human proteome by using antibody based proteomics and in these projects, it is aimed to establish a database which provides information about human proteins' expression in normal tissue types and several types of cancer.

1.1.5 Protein-Protein Interaction Analysis

Identification of interacting partners of novel gene products may provide direct information about the cellular pathway in which protein of interest is involved and cellular localization. There are several tools for analysis of protein-protein interaction. One of the most widely used techniques is yeast two-hybrid system (Fields *et al.*, 1989). This technique is based on the principle of activation of a reporter gene by a transcription activator. Protein of interest is fused to DNA-binding domain of transcription activator and introduced into the yeast. Transcription activating domain of the activator is attached to proteins from a certain cDNA library and these constructs are individually introduced into yeast cells containing the protein of interest. If protein interacts with a protein of cDNA library, transcriptional activation and DNA-binding domains are united and a reporter gene is transcribed. Once the interaction data is obtained, interaction partner may be confirmed via carrying out a co-immunoprecipitation or fluorescence resonance energy transfer (FRET). Regarding co-immunoprecipitation, an antibody/protein complex is pelleted using sepharose beads. If there are any interacting proteins with target protein in the complex, interacting partner is also pelleted and it can be identified by carrying out a western blot. In FRET, proteins expected to be interacting are labeled with different fluorochromes such that if proteins interact, energy absorbed by first fluorochrome will be transferred to second fluorochrome and second fluorochrome will emit light

(Miyawaki *et al.*, 2000).

1.1.6. Loss of function Analysis

1.1.6.1. Down-regulation of a gene product by siRNA Strategy

RNA interference refers to the post-transcriptional silencing of gene expression due to the introduction of homologous double stranded RNA (Fire *et al.*, 1998). Nowadays, it is the most widely used technique to silence gene expression and analyze gene function. Use of siRNA as a means of gene silencing depends on several factors such as the degree to which a gene can be silenced, the length of time for which the gene remains silenced, the degree of recovery of gene function, and the effects of silencing process on general cell functions (Zhang *et al.*, 2004). The most effective siRNAs are 21nt dsRNA with 2nt 3' overhangs (Elbashir *et al.*, 2001) and another important point is the sequence specificity of siRNA since even single base pair mismatches between siRNA and its target mRNA reduce silencing to a great extent (Brummelkamp *et al.*, 2002). siRNAs can be generated via several strategies. Chemical syntheses, in vitro transcription, digestion of long dsRNA by an RNase III family enzyme, siRNA expression vectors, gene silencing by PCR product are the most powerful methodologies (Zhang *et al.*, 2004).

1.1.6.2. Knocking-out a gene in mice

Gene function analysis via targeting the gene in mouse embryonic stem cells (Thomas *et al.*, 1987) has been a powerful tool for many years. In this technique, a special vector containing flanking homologous sequences with the targeted locus and an antibiotics resistance gene is constructed and transfected to embryonic stem cells. Homologously recombined stem cells are then injected to a blastocyst and following implantation and selective mating give rise to animals with the desired genetic alteration. However, analysis of gene function by this procedure may be difficult since complete absence of the gene may be embryonically lethal. At this point, using a conditionally knock-out mice model might be more useful. In this strategy, target

gene is flanked by loxP sites, which are required for homologous recombination, and vector construct is transfected to embryonic stem cells as in the complete targeting protocol. Then, mice are crossed with mice expressing Cre recombinase in a tissue specific manner, ligand specific manner, or both. Cre mediates excision of the targeted gene in the desired tissue (Gu *et al.*, 1994), and in response to the ligand, for example, tamoxifen- inducible Cre recombinase (Danielian *et al.*, 1998). By this conditional knock-out strategy, the function of a gene can be analyzed at the organism level, and it can precisely be established.

1.2. Identification and General Characteristics of *HANEIN-1*

HANEIN-1 was firstly identified in an unrelated study in which 3' transcriptional regulatory regions of Na^+/I^- Symporter (*NIS*) gene were being analyzed. In that study 10 conserved putative regulatory regions were identified via using a software tool called VISTA (Bary *et al.*, 2003 and Couronne *et al.*, 2003). Investigation of one these putative regions revealed that this region was not a 3' regulatory element of *NIS* but it was a region controlling expression of *HANEIN-1* (Hani Alotaibi and Uygar Tazebay, personal communication). HANEIN-1 is described as hypothetical protein coiled coil domain in the Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002).

The most important reason why we are interested in characterization of HANEIN-1 is that it is conserved in all eukaryotes from *Saccharomyces pombe* to *Homo sapiens*. It has 75% identity in mammals, 50% identity with insects and nematodes and 30-40% identity with plants and ascomycetes. Since it is conserved we think that it carries out a major function in the cell.

HANEIN-1 has no domain similarities with any known proteins. It contains 4 coding exons and it is predicted to encode a 223 amino acid protein in human.

1.3. Aim

In this study, the aim was the functional characterization of HANEIN-1. By bioinformatics analysis we first aimed to identify the conservation pattern of the protein among eukaryotic species, possible functional sites and to find some clues in order to plan experimental strategies. Experimentally, we mainly concentrated on the biochemical characterization of the protein. Expression of HANEIN-1 was investigated both at the transcript and at the protein level. In addition to these, sub-cellular localization analysis and mutagenesis analysis of some functional sites were performed in order to characterize the function of HANEIN-1.

2. MATERIALS AND METHODS

2.1. Bioinformatics Tools

Genomic DNA sequence, mRNA sequence of the genes and amino acid sequence of the proteins were identified by using Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002). In order to analyze the conservation pattern of the novel gene among eukaryotic species ClustalW2 multiple alignment program (Larkin *et al.*, 2007) was utilized and sequence similarity searches were carried out via FASTA (Pearson *et al.*, 1988). Pattern and profile searches of the protein were investigated via ELM (Puntervoll *et al.*, 2003), SMART (Schultz *et al.*, 1998) and PROSCAN (Combet *et al.*, 2003). Post-translational modifications were predicted by using NetPhos (Blom *et al.*, 1999) and SUMOPlot (Xue *et al.*, 2006) programs.

2.2. Cell Culture

2.2.1. Cell Lines

In this study, Hep3B (Human Negroid hepatocyte carcinoma), HepG2 (Human Caucasian hepatocyte carcinoma) liver carcinoma cell lines and MCF-7 (Caucasian female breast adenocarcinoma) breast cancer cell line were used for expression analysis.

2.2.2. Growth Media

Cell lines were grown in high glucose Dulbecco's modified Eagle's medium (Gibco) with the addition of 10% fetal bovine serum (FBS), 1% penicillin/streptomycin and 1% L-glutamine (Biochrom) at 37˚C in a 5% CO_2 incubator.

2.2.3. Transfection

Transfection was performed by using FuGene-6 reagent (Roche). Optimized transfection reagent (µl) /plasmid DNA (µg) was 3:1. Firstly, transfection reagent was diluted in serum and antibiotics free medium. After 10 minutes of incubation, plasmid DNA was added and FuGene-plasmid DNA mixture was incubated for 30 minutes. During this time interval, medium of the cells which were cultivated at 6-well plates at 90-95% confluency was changed. Finally, FuGene-DNA complex was transfected to cells in a drop-wise manner.

Transfection efficiencies of plasmids were tested by RT-PCR analysis of neomycin resistance gene found in the transfected plasmids. Neomycin primers are Forward: ACAAGATGGATTGCACGCAG and Reverse: TTCGCCCAATAGCAGCCAGT

2.3. Cloning

2.3.1. Site-Directed Mutagenesis

PCR-based mutagenesis was carried out in order to convert 4 possible serine phosphorylation sites of HANEIN-1 to alanine by designing mutagenesis primers. Sequence of the mutagenesis primers are presented in Table 2.1. PCRs were set up in 50 µl reaction volume with 5 µl of $MgCl_2$ containing reaction buffer, 1 µl of 10mM dNTP, 1.25 µl's of (10 pmol/ µl) sense and antisense primers, 1 µl of Pfu polymerase (Fermentas) and 2.5 µl of (20 ng/ µl) plasmid DNA in which *HANEIN-1* was cloned (p3XFlag-Locus). Reaction conditions were as follows: 30 s initial denaturation at 95˚C, 18 cycles of 30 s denaturation at 95˚C, 1 min primer annealing at 55˚C and 7 min extension at 68˚C, and final extension was disabled. After the PCR reaction, 1 µl of DpnI was added to each PCR tube with minimum 1 hour incubation at 37˚C in order to degrade methylated DNA strand. Then 5 µl of mutated product was transformed to competent *E. coli* as described in Molecular Cloning (Maniatis *et al.*, 1989). Single-colonies were picked up from the transformed bacteria and plasmid isolation was carried out with Fermentas GeneJet MiniPrep Plasmid Kit. Finally,

mutagenized products were confirmed by sequencing.

Table 2.1. Site-directed mutagenesis primers

Primer	Sequence
S92A Sense	AGGCGCCGCGGGTGGCCACGGCCAGCAAGGTCACCCGGGCCCA
S92A Anti-sense	TGGGCCCGGGTGACCTTGCTGGCCGTGGCCACCCGCGGCGCCT
S122A Sense	ACACAGCCGAGAAAGCCAAGGCCCATCTGGAGGTGCCGCTGGA
S122A Anti-sense	TCCAGCGGCACCTCCAGATGGGCCTTGGCTTTCTCGGCTGTGT
S194A Sense	AGAACCCCAACATGCGGCTGGCGCAGCTGAAACAGCTGCTCAA
S194A Anti-sense	TTGAGCAGCTGTTTCAGCTGCGCCAGCCGCATGTTGGGGTTCT
S207A Sense	TCAAGAAGGAGTGGCTCCGCGCGCCTGACAACCCCATGAACCA
S207A Anti-sense	TGGTTCATGGGGTTGTCAGGCGCGCGGAGCCACTCCTTCTTGA

2.3.2. Gene Cloning

Two isoforms of *RASGEF1B* and *RASGEF1A* were cloned via using p3XFLAG-CMV14 expression vector. Firstly, gene products (inserts) were amplified via specific cloning primers in PCR; for *RASGEF1B* 1st isoform PCR conditions were as follows: 5 min initial denaturation at 95˚C, 35 cycles of 30 s denaturation at 95˚C, 20 sec primer annealing at 58˚C and 1.5 min extension at 72˚C, and a 5 min final extension at 72˚C, for *RASGEF1B* 2nd isoform: 5 min initial denaturation at 95˚C, 35 cycles of 30 s denaturation at 95˚C, 20 sec primer annealing at 58˚C and 35 sec extension at 72˚C, and a 5 min final extension at 72˚C and for *RASGEF1A*: 5 min initial denaturation at 95˚C, 35 cycles of 30 s denaturation at 95˚C, 20 sec primer annealing at 60.7˚C and 1.5 min extension at 72˚C, and a 5 min final extension at 72˚C . PCR products and vector were digested with NotI and XbaI (Fermentas) restriction enzymes and digested products were ligated with T4 DNA Ligase (Fermentas) and transformed to competent *E. coli.* Single colonies from transformed bacteria were picked up and tested for positive cloning via both PCR and restriction enzyme digestion. Cloning primers are presented in Table 2.2.

Table 2.2. Cloning primers

Gene	Primer	Sequence
RASGEF1B 1st isoform	Forward	AAGCTTGCGGCGGCATGGAACAAAAACTCATCTCAGAAGA GGATCTGATGCCTCAGACTCCTCCCTTTTCAG
	Reverse	GGATCCTCTAGAAACTCTGCCTAAGAGGCTCGACCTT
RASGEF1B 2nd isoform	Forward	AAGCTTGCGGCGGCATGGAACAAAAACTCATCTCAGAAGA GGATCTGATGCCTCAGACTCCTCCCTTTTCAG
	Reverse	GGATCCTCTAGATATCCCTTTGAAGTGGGATGGTATA
RASGEF1A	Forward	AAGCTTGCGGCCGCATGGAACAAAAACTCATCTCAGAAGA GGATCTGATGCCCCAGACGTCCGTTGTCTTCT
	Reverse	GGATCCTCTAGAGGCTCTGTTCAGAAGGGTGGTCCTG

2.4. RNA Isolation

RNA from mouse tissue samples and hepatocellular carcinoma cell lines were isolated via using Macherey-Nagel Nucleospin RNA isolation kit. For RNA isolation from cell lines, isolation was carried out as described in manufacturer's protocol via first lysing cell pellets. In order to isolate RNA from tissue samples, firstly tissue samples were ground by using pestle and mortar in liquid N_2, and tissue powder was homogenized via a teflon-glass homogenizer, then tissue powder was lysed and kit protocol was followed. RNA concentrations were determined by using nanodrop spectrophotometer.

2.5. cDNA synthesis and RT-PCR

cDNA synthesis was performed with approximately 1 µg of RNA via using RevertAid First Strand cDNA Synthesis kit (Fermentas). Synthesized cDNAs were amplified via semi-quantitative RT-PCR. PCRs were set up in 25 µl reaction volume with 2.5 µl of 10X Taq Polymerase reaction buffer, 1.5 µl of 25mM $MgCl_2$, 0.5 µl of 10mM dNTP, 1µl's of (10 pmol/ µl) forward and reverse primers, 0.2µl of Taq polymerase (Fermentas) and required volume of cDNA. PCR conditions for

HANEIN-1 were as follows: 5 min initial denaturation at 95˚C, 35 cycles of 30 sec denaturation at 95˚C, 30 sec primer annealing at 60˚C and 30 sec extension at 72˚C, and a 5 min final extension at 72˚C, for *RASGEF1B*: 5 min initial denaturation at 95˚C, 35 cycles of 30 sec denaturation at 95˚C, 30 sec primer annealing at 58˚C and 30 sec extension at 72˚C, and a 5 min final extension at 72˚C and for *GAPDH*: 5 min initial denaturation at 95˚C, 21 cycles of 30 sec denaturation at 95˚C, 30 sec primer annealing at 60˚C and 30 sec extension at 72˚C, and a 5 min final extension at 72˚C. RT-PCR primers are presented in Table 2.3. PCR products were visualized via agarose gel electrophoresis.

Table 2.3. RT-PCR primers and expected product sizes

Gene	Primer	Sequence	Product Size
HANEIN-1	HALNG-F	GAATTCAAGCTTATGCCCAAGAAGTTCCAG	1060 bp
	HALNG-R	GGATCCTCTAGAGCTCACTTGGGGGCATTG	
	Loc-Seq F	TGGCCACGTCCAGCAAGGTC	210 bp
	Loc-Seq R	TCCGCCACGCTGAGCACTGCA	
RASGEF1B	RASGEF1B RT-F	GAGCACCAGAGACTAAGTGA	325 bp
	RASGEF1B RT-R	CCCTTTGTATAGACTGTGGC	
GAPDH	GAPDH F	GGCTGAGAACGGGAAGCTTGTCAT	150 bp
	GAPDH R	CAGCCTTCTCCATGGTGGTGAAGA	

2.6. Protein Isolation

Proteins from cell lines were isolated for western blotting analysis. Cell pellets were lysed in 50 µl lysis buffer consisting of 50 mM Tris Base, 250 mM NaCl, 1X proteinase inhibitor cocktail and 0.1% NP40. Protein concentrations were determined via using Bradford assay.

2.7. Western Blotting

10 µg of proteins were denaturated in 1X gel loading buffer consisting of 50 mM Tris-Cl (pH 6.8), 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol and 5% ß-mercaptoethanol by boiling for 5 minutes. Denaturated proteins were run in 10% SDS-polyacrylamide gel in Tris-glycine electrophoresis buffer (25 mM Tris, 250 mM glycine (pH 8.3) and 0.1% SDS). Then, proteins were transferred to PVDF membrane (Millipore) and for 1 hour blocked with Blotto consisting of 1X TBS (Tris-buffered saline), 0.5% Tween 20 and 5% milk powder. Primary antibodies were prepared in Blotto with required dilutions (Table 2.4.). Membranes were incubated in primary antibodies for 1 hour and washed 3 x 10 min with 1X TBS-T. Then membranes were incubated in HRP-conjugated secondary antibodies (1:5000, Sigma) for another 1 hour and washed 3 x 10 min with 1X TBS-T. Proteins on the membrane were detected via Super Signal West Dura (Pierce) and exposed to X-ray films (AGFA) for 1min.

For the immunoblotting analysis in which blockage of the primary antibody is required, approximately 10^{-5} molar (M) antibody specific peptide was used in order to block antigen/antibody binding.

Table 2.4. List of primary antibodies used in western blot analysis. Type of the antibody, organism in which antibody produced and working concentrations are presented.

Primary antibody	Description	Concentration in blotto
N-ter specific	Rabbit, polyclonal	2 µg/ml
Anti-Flag (Sigma)	Mouse, monoclonal	1 µg/ml
Anti-Calnexin	Rabbit, polyclonal	0.1 µg/ml
Anti-RARα (Santa Cruz)	Rabbit, polyclonal	1 µg/ml

2.8. Northern Blotting

Northern blotting was performed on FirstChoice Northern Human Blot1 membrane (each lane containing 2 μg poly (A) RNA) (Ambion).

2.8.1. Probe Synthesis

DNA templates for probe preparation were formed via restriction enzyme digestion from PCR product cloned plasmids for *HANEIN-1* and *RASGEF1B* or directly from PCR products for *GAPDH* and β-actin. PCR products for *GAPDH* and β-actin formed via primers which were kindly provided by Ayşe Elif Erson and Mehmet Öztürk's group, respectively. Then, DNA templates were labeled by north2south biotin random prime labeling kit (Pierce). Probes were synthesized as described in manufacturer's protocol. Probes and their sizes are presented in Table 2.5.

Table 2.5. Northern blotting probes. Starting DNA material, used restriction enzymes for probe synthesis and synthesized probe sizes are indicated.

Gene	Starting material	Restriction Enzyme	Probe Size
HANEIN-1	p3XFlagHANEIN-1	Apa I and Xba I	381 bp
RASGEF1B	p3XFlagRASGEF1B	Bgl II and BamHI	284 bp
GAPDH	GAPDH PCR product	-	408 bp
β-actin	β-actin PCR product	-	539 bp

2.8.2. Nucleic Acid Hybridization and Detection

In order to hybridize the nucleic acids and perform the detection by synthesized probes, north2south chemiluminescent hybridization and detection kit (Pierce) was utilized. Kit's protocol was exactly followed and resulting blots were exposed to film for approximately 1 min.

3. RESULTS

3.1. Identification of *HANEIN-1*

HANEIN-1 was firstly identified in a study investigating 3' *cis*-acting transcriptional control elements of the Na^+/I^- Symporter (*NIS*). In that study, a 90 kb genomic DNA fragment including and flanking the *NIS* gene were being analyzed in order to identify at least 50% conserved putative regulatory elements of *NIS* in human, mouse and rat via using VISTA tool (Bary *et al.*, 2003 and Couronne *et al.*, 2003). 10 conserved regions were identified by this analysis and 10^{th} region was not a *NIS* gene transcriptional regulatory element but a region controlling the expression of *HANEIN-1* (Alotaibi *et al.*, unpublished data).

3.2. Bioinformatics Analysis of HANEIN-1

HANEIN-1 is described as coiled-coiled domain containing 124 (CCDC-124) in Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002). It consists of 4 coding exons and it is predicted to encode a 223 amino acid protein.

Biochemical features of HANEIN-1 protein were analyzed via using ProtParam tool (Gasteiger *et al.*, 2005). Molecular weight and isoelectric point of the protein were calculated as 25835.2 Da and 9.54, respectively. Besides, this analysis showed that protein consisted of 43 negatively charged, 51 positively charged residues, in other words, 42% of the protein consisted of charged residues.

HANEIN-1 protein sequence of several species was aligned via using

ClustalW2 multiple alignment program (Larkin *et al.*, 2007). This analysis revealed a remarkable alignment among the sequences analyzed especially through residues between 25-80 and 160-210 (Figure 3.1).

Sequence similarity search was performed via FASTA (Pearson *et al.*, 1988) by referencing *Homo sapiens* HANEIN-1 protein sequence. This analysis revealed that identity and similarity of *Homo sapiens* HANEIN-1 protein sequence with other species are high among the eukaryotes. This analysis is presented in Table 3.1.

Table 3.1. Sequence similarity search of *Homo sapien*'s HANEIN-1 protein. % identities, % similarities and proteins' amino acid length are indicated.

Alignment	**Amino acid Length**	**% Identity**	**% Similarity**
Homo sapiens	223	100	100
Pongo pygmaeus	223	99.1	99.6
Canis familiaris	223	95.5	98.7
Mus musculus	217	88.8	95.5
Xenopus laevis	217	72.8	91.0
Danio rerio	216	70.0	89.1
Caenorhabditis elegans	223	50.4	72.6
Drosophila melanogester	213	43.8	73.7
Oryza sativa	238	38.9	65.0
Aspergillus nidulans	222	35.1	58.2

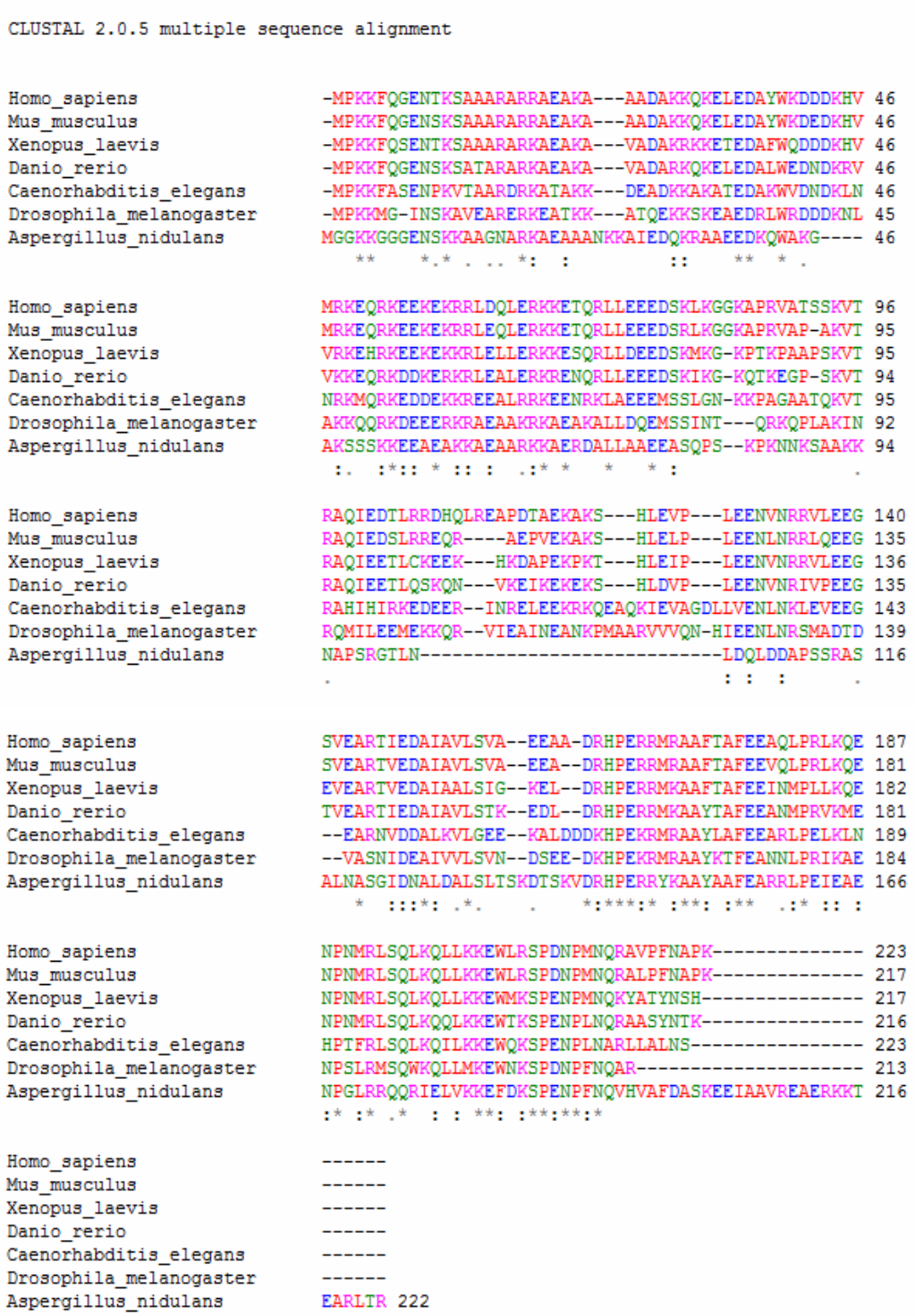

Figure 3.1. ClustalW2 multiple alignment of HANEIN-1 in selected species. Species name, amino acid positions and alignment for each amino acid are shown.

Domains and possible functional sites of HANEIN-1 were investigated via using some of Expasy Proteomics Tools such as SMART (Schultz *et al.*, 1998), ELM (Puntervoll *et al.*, 2003) and PROSCAN (Combet *et al.*, 2003).

SMART confidently predicts that protein has DUF1014 domain between 1-216 amino acids with an E value of 1.10e-118 and Pfam analysis of Sanger Institute shows that DUF1014 domain is a member of clan HMG-box (CL0114) and HMG-box family is described as a clan including DNA-binding HMG-box proteins as well as the YABBY-like transcription factors.

ELM search for identification of functional sites in HANEIN-1 resulted in many ELM descriptions, from cleavage sites to motifs recognized by signaling domains. Some results of this analysis are presented in Table 3.2.

Table 3.2. ELM search for identification of functional sites in HANEIN-1.

ELM description	**Position**	**Matched Sequence**	**Pattern**
Furin (PACE) cleavage site	16-20	RARRA	R.[RK]R.
Nuclear receptor box motif	195-201	QLKQLLK	LXXLL
Motif recognized by those SH3 domains with a non-canonical class I recognition specificity	205-211	LRSPDNP	…[PV]..P
PKA phosphorylation site	191-197	MRLSQLK	.R.([ST])…
Site phosphorylated by the Polo-like-kinase	100-106 138-144	IEDTLRR EEGSVEA	.[DE].[ST] [ILFWVMA]..
Motif recognized for modification by SUMO-1	184-187 200-203	LKQE LKKE	[VILMAFP]K.E

PROSCAN search resulted in several phosphorylation site descriptions and results of this analysis are presented in Table 3.3.

Table 3.3. PROSCAN search results for HANEIN-1.

PROSITE description	Position	Matched Sequence	Pattern
cAMP- and cGMP-dependent protein kinase phosphorylation site	67-70	KKET	[RK](2)-x-[ST]
Protein kinase C phosphorylation site	70-72 92-94 103-105	TQR SSK TLR	[ST]-x-[RK]
Casein kinase II phosphorylation site	122-125 146-149 155-158 174-177	SHLE TIED SVAE TAFE	[ST]-x(2)-[DE]

Bioinformatics cellular location and function analysis of HANEIN-1 was performed via Gene Ontology Annotation (GOA) database. It was predicted that with 83.87% probability HANEIN-1 localized in the nucleus and with 61.55% probability in the myosin complex. In addition to cellular location, it was predicted to be involved in DNA-dependent regulation of transcription with 79.21% probability and in cellular transport with 37.47 % probability in the context of its biological function.

Phosphorylation sites were predicted via NetPhos (Blom *et al.*, 1999) and sumoylation sites were predicted via SUMOplot prediction tool (Xue *et al.*, 2006). NetPhos predicted 5 serine, 7 threonine and 1 tyrosine phosphorylation sites and this analysis is presented graphically in Figure 3.2 and predictions are tabulated in Table 3.5. Among these predicted phosphorylation sites serine residue at position 92 and threonine at position 103 are presented as being phosphorylated in UNiProtKB/Swiss Prot entry database based on the experimental evidence provided by the study "Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry" (Molina *et al.* , 2007)

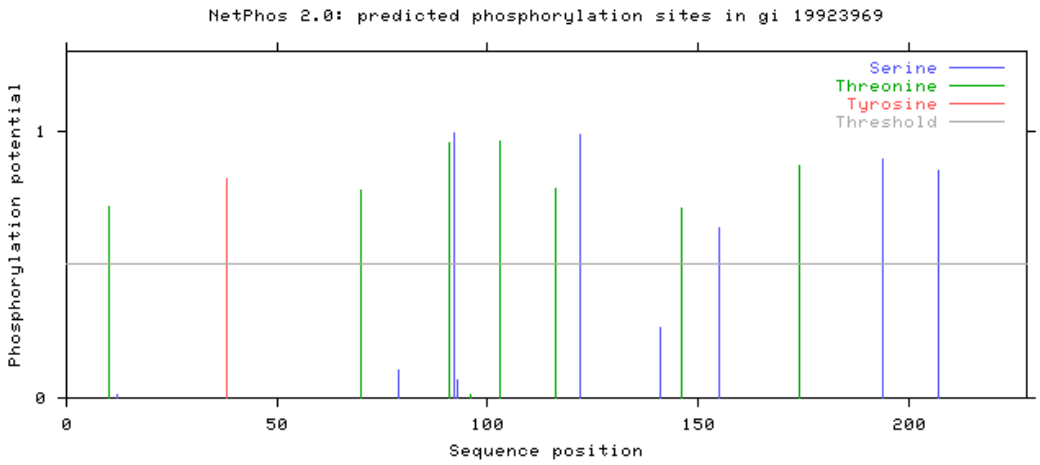

Figure 3.2. Graphical representation of phosphorylation sites of HANEIN-1

SUMOPlot prediction tool predicted 3 lysine sumoylation motifs with high probability and 4 lysine sumoylation motifs with low probability. Results of this analysis are presented in Table 3.4.

Table 3.4. SUMOPlot analysis for HANEIN-1 protein.

Position	Group	Score
201	QLKQL **LKKE** WLRSP	0.91
185	AQLPR **LKQE** NPNMR	0.91
82	EEDSK **LKGG** KAPRV	0.73
40	LEDAY **WKDD** DKHVM	0.64
85	LEDAY **WKDD** DKHVM	0.57
67	LDQLE **RKKE** TQRLL	0.44
53	MRKEQ **RKEE** KEKRR	0.44

Table 3.5. Phosphorylation predictions of HANEIN-1

Serine Predictions			
Position	Context	Score	Prediction
12	ENTKSAAAR	0.014	-
79	EEEDSKLKG	0.102	-
92	RVATSSKVT	0.995	Phosphorylation
93	VATSSKVTR	0.065	-
122	EKAKSHLEV	0.985	Phosphorylation
141	LEEGSVEAR	0.264	-
155	IAVLSVAEE	0.640	Phosphorylation
194	NMRLSQLKQ	0.893	Phosphorylation
207	EWLRSPDNP	0.851	Phosphorylation
Threonine Predictions			
Position	Context	Score	Prediction
10	QGENTKSAA	0.717	Phosphorylation
70	RKKETQRLL	0.780	Phosphorylation
91	PRVATSSKV	0.957	Phosphorylation
96	SSKVTRAQI	0.014	-
103	QIEDTLRRD	0.964	Phosphorylation
116	EAPDTAEKA	0.785	Phosphorylation
146	VEARTIEDA	0.709	Phosphorylation
174	VEARTIEDA	0.871	Phosphorylation
Tyrosine Predictions			
Position	Context	Score	Prediction
38	LEDAYWKDD	0.820	Phosphorylation

Gene expression analysis of *HANEIN-1* was carried out via Oncomine Research Platform (Rhodes *et al.*, 2004). Normal tissue expression analysis revealed

a remarkable expression in skeletal muscle (Figure 3.3). In addition to these, it was identified that EGFR amplification positive glioma expressed *HANEIN-1* at a higher level as compared to EGFR amplification negative glioma (Figure 3.4A). Expressional analysis of T cell acute lymphoblastic leukemia cell line also showed that *HANEIN-1* expression was lower in cases with mutant PTEN than cases with wild type PTEN (Figure 3.4B).

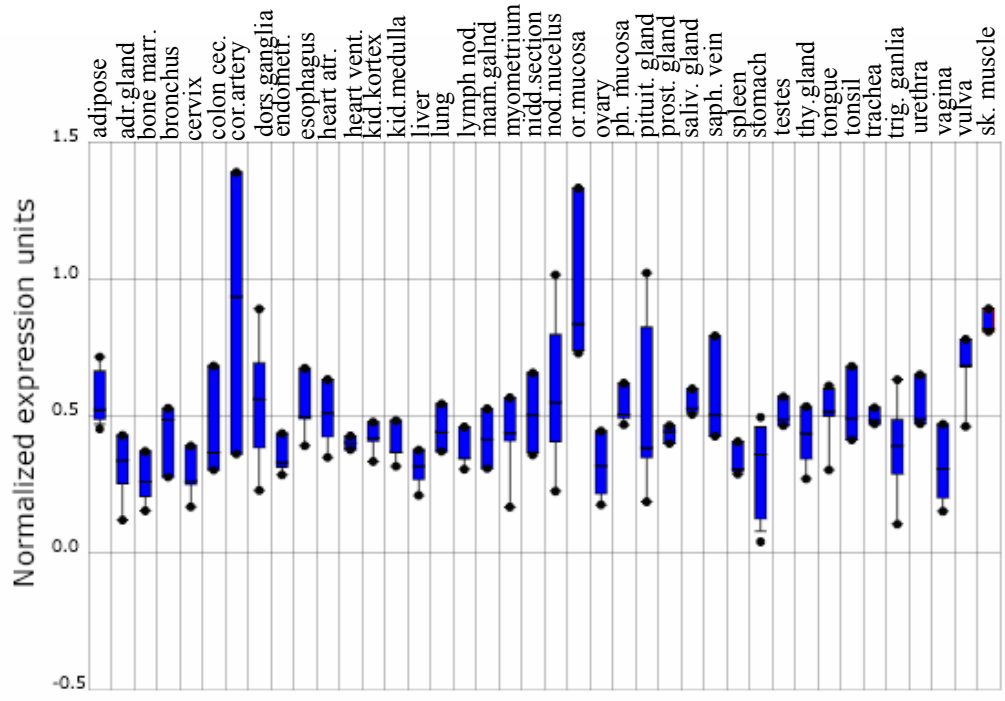

Figure 3.3. Oncomine expression profile of *HANEIN-1* in normal tissues. Bars represent expression status for adipose, adrenal gland, bone marrow, bronchus, cervix, colon cecum, coronary artery, dorsal root ganglia, endometrium, esophagus, heart atrium, heart ventricle, kidney cortex, kidney medulla, liver, lung, lymph nodes, mammary gland, myometrium, nipple cross-section, nodose nucleus, oral mucosa, ovary, pharyngeal mucosa, pituitary gland, prostate gland, salivary gland, saphenous vein, spleen, stomach, testes, thyroid gland, tongue, tonsil, trachea, trigeminal ganglia, urethra, vagina, vulva and skeletal muscle, respectively. Asterisks show the minimum and maximum values obtained for each tissue.

A) B)

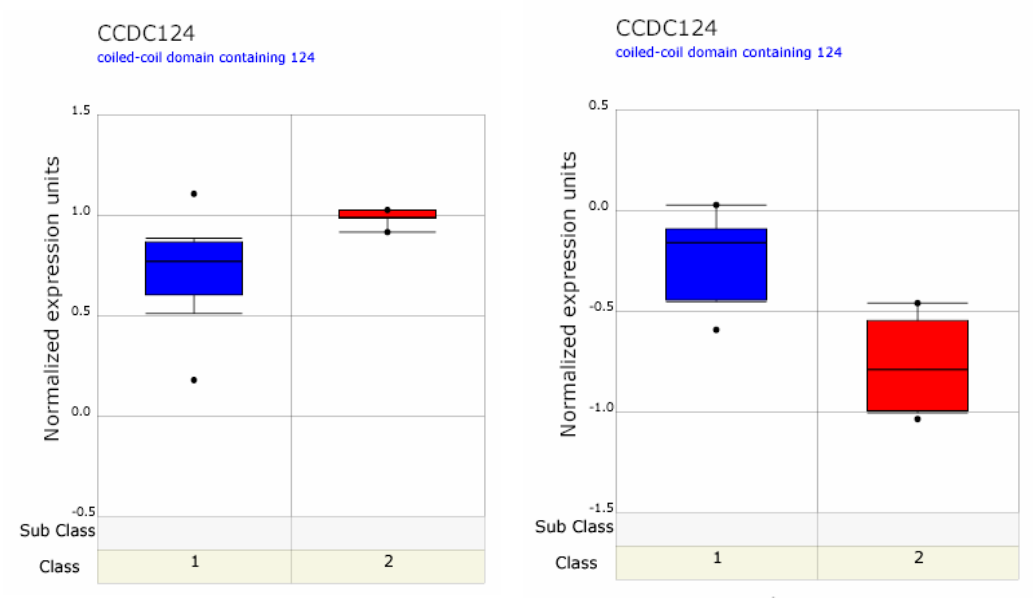

Figure 3.4. Oncomine expressional analysis of *HANEIN-1* in cases with molecular alteration. A) Gliomas with EGFR amplification negative (class 1) and positive (class 2) status. T-test: -5.451, P-value: 1.9E-5. B) T cell acute lymphoblastic leukemia cell line with PTEN wild type (class 1) and mutant (class 2) status. T-test: 5.547, P-value: 3.2E-5

3.3. Expressional Analysis of *HANEIN-1*

3.3.1. Cell Line Expression Analysis by RT-PCR

Expression of *HANEIN-1* was studied in several breast carcinoma cell lines. Breast carcinoma cell lines' cDNA and their *GAPDH* expression profiles were kindly provided by Nilgün Taşdemir. RT-PCR analysis showed that *HANEIN-1* was expressed in all breast carcinoma cell lines studied and there was not significant expression variation among different cell lines (Figure 3.5).

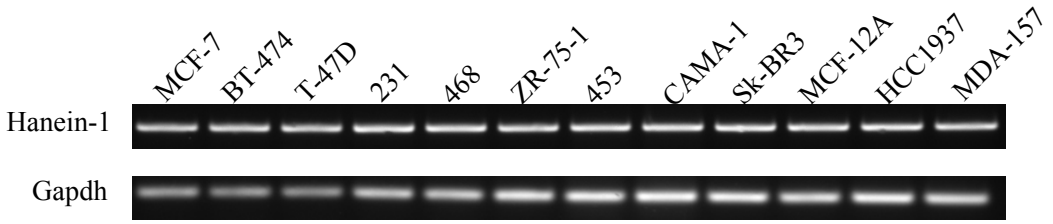

Figure 3.5. *HANEIN-1* expression in breast carcinoma cell lines. cDNA from different breast carcinoma cell lines were amplified with specific primers (see Materials and Methods) and resulting PCR products were analyzed via agarose gel electrophoresis.

3.3.2. Mouse Tissue Expression Analysis by RT-PCR

Mouse tissue expression analysis was performed with female mouse tissues which were kindly provided by Dr. Hani Alotaibi. Reverse transcription of mouse RNAs revealed that *HANEIN-1* was expressed in all mouse tissues and it showed a high level expression pattern especially in thyroid, liver, stomach and lactating mammary gland (LMG) (Figure 3.6).

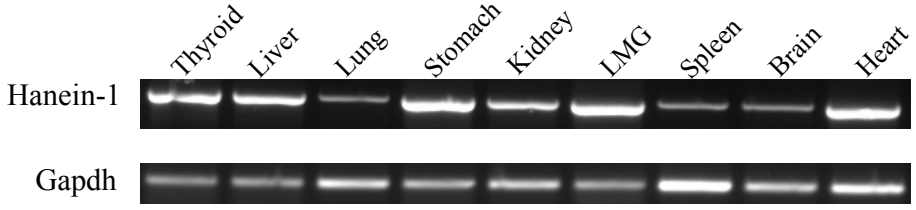

Figure 3.6. Expressional analysis of *HANEIN-1* in mouse tissues. Tissue samples were ground in liquid N_2 and homogenized via teflon-glass homogenizer. RNA isolated from homogenized tissue powder was reverse transcribed and cDNA was amplified with specific primers for *HANEIN-1*. Resulting PCR products were analyzed by agarose gel electrophoresis. *GAPDH* was used as a loading control.

3.3.3. Human Tissue Expression Analysis by Northern Blotting

HANEIN-1 expression in human tissues was analyzed via northern blotting, which was carried out on a human tissue RNA transferred membrane (Human Blot 1, Ambion) and biotin labeled probes. In Figure 3.7 probes and their specific targets are presented.

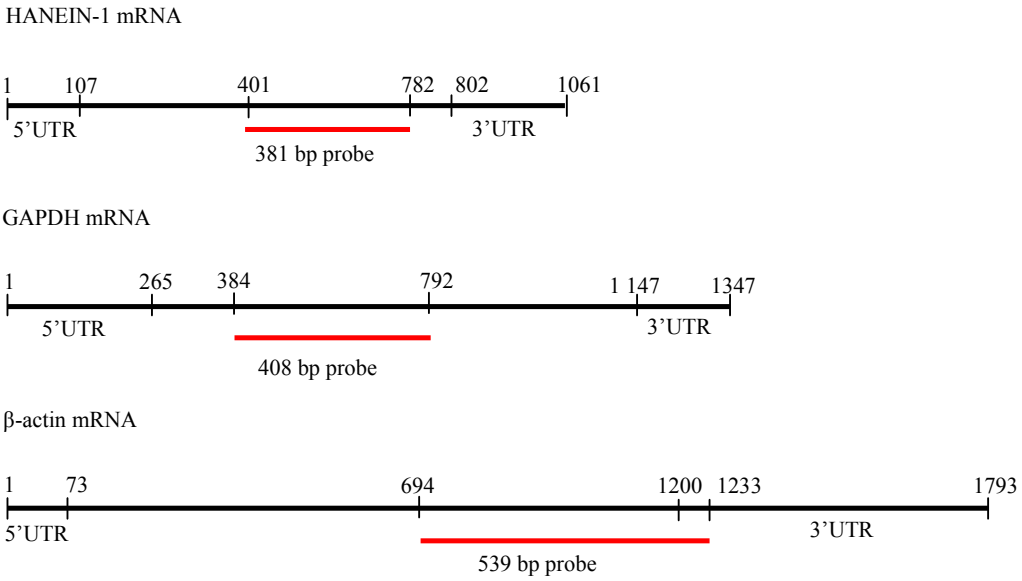

Figure 3.7. Northern blotting probes for *HANEIN-1*, *GAPDH* and β-actin. mRNA size, 5'UTR, 3'UTR and probe size & location are indicated for each gene.

Detection with specific probes revealed that *HANEIN-1* was ubiquitously expressed in human tissues (Figure 3.8). Highest expression level was observed in skeletal muscle. In addition to these, *HANEIN-1* transcript size was determined as approximately 1000 bp. In Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002), *HANEIN-1* transcript is predicted to be 1061 bp, consisting of 5 exons (4 coding and 1 non-coding). Our results show that *HANEIN-1*

is ubiquitously expressed as full transcript. Besides, northern blot analysis carried out by Dr. Uygar Tazebay on a different membrane revealed that *HANEIN-1* was expressed as two isoforms in placenta, one 1061 bp form and an additional 950 bp form (data not shown).

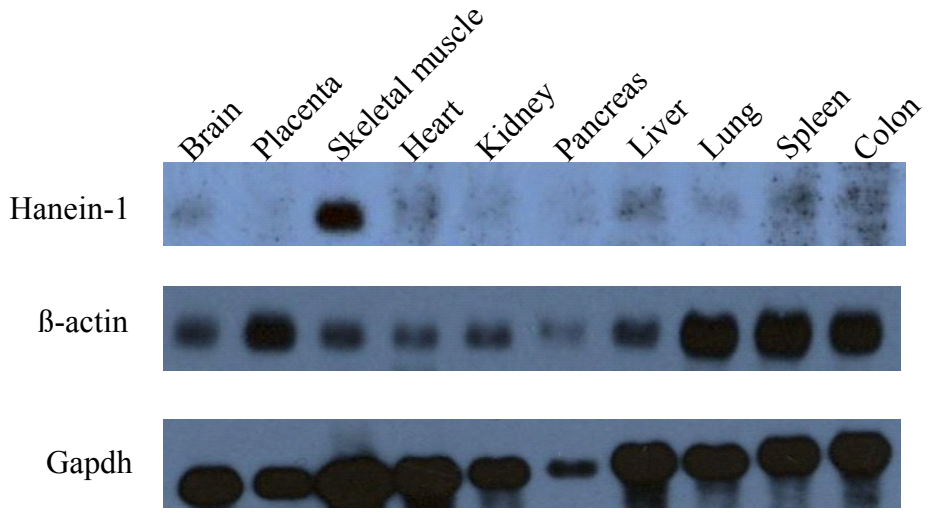

Figure 3.8. Human tissue expression analysis of *HANEIN-1* by northern blotting. Probes were labeled via biotin and hybridized to Human Blot 1 membrane (see Materials and Methods). Detected blots were exposed to X-ray films. β-actin and *GAPDH* were used as internal controls.

3.4. Immunological Analysis

3.4.1. Identification of HANEIN-1 Protein via Western Blotting

HANEIN-1 cDNA was previously cloned into pcDNA3 expression vector by Dr. Hani Alotaibi. Hep3B liver carcinoma cell line was transfected with HANEIN-1 expressing plasmid and an immunoblotting experiment was performed with the proteins extracted from Hep3B cell line. In order to detect HANEIN-1, an antibody previously generated against N-terminal 24 amino acid of the protein was utilized. This analysis showed that HANEIN-1 encoded approximately a 33 kDa protein since

intensity of this 33 kDa band increased in transfected cell line (Figure 3.9). In addition to a 33 kDa protein band, several upper and lower bands were also detected with N-ter specific antibody. Therefore, in order to identify whether these bands are specific to HANEIN-1, immunoblotting analysis was repeated with N-ter specific antibody blocked with its specific peptide (described in Materials and Methods). It was identified that all the bands present in the immunoblot carried out with N-ter specific antibody were lost upon blocking the antibody with its specific peptide. Based on these results, it may be concluded that antibody recognized all the bands with its Fab region, i.e, all the bands may be specific to HANEIN-1 or there can be other proteins with the same epitope that can be recognized by the used antibody.

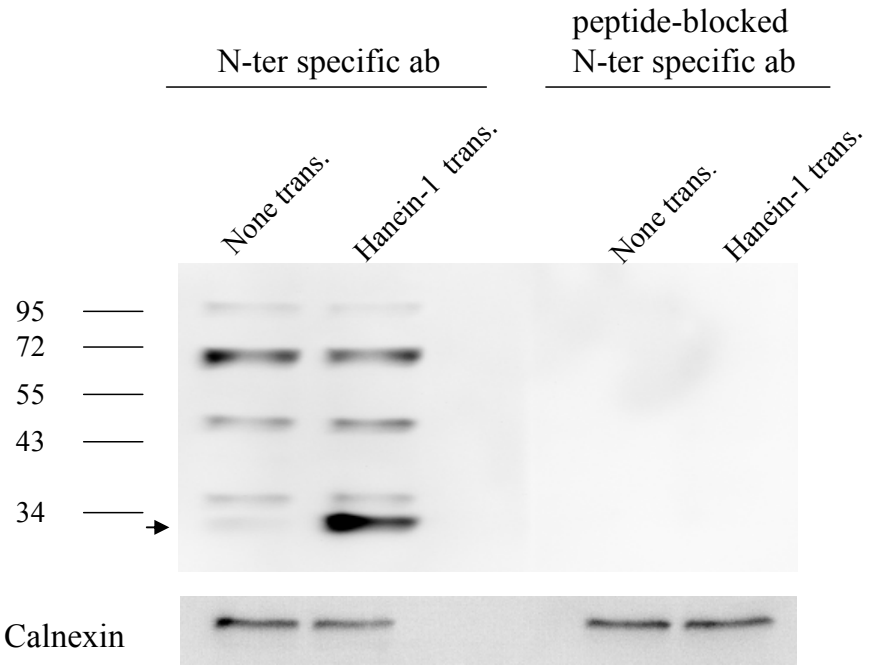

Figure 3.9. HANEIN-1 encodes a 33 kDa protein. Proteins isolated from none transfected and HANEIN-1 transfected Hep3B cell line were denatured via SDS-PAGE and then transferred to membranes for immunoblotting analysis. Membranes were incubated with N-ter specific antibody and peptide-blocked N-ter specific antibody to analyze protein expression. On the left marker sizes were presented (in kDa) and calnexin was used as a loading control.→ displays the 33 kDa HANEIN-1 band.

3.4.2. Immunoblotting Analysis of HANEIN-1 Labeled with Flag epitope

Flag epitope fused HANEIN-1 protein was analyzed via western blotting by using plasmids p3XFlag-Locus and pLocus-3XFlag, which were kindly provided by Dr. Hani Alotaibi. In these plasmids, HANEIN-1 was fused with Flag epitope at N-terminus and at C-terminus. Immunoblotting analysis carried out with anti-Flag antibody in Flag epitope labeled HANEIN-1 transfected cell lines displayed that labeling of the protein with Flag epitope resulted in a different fragmentation pattern of the protein depending on the position of the Flag tag (Figure 3.10). Another important result of this analysis was that HANEIN-1 protein expression pattern was different for different cell lines, each cell line exhibiting a different intensity for the same band.

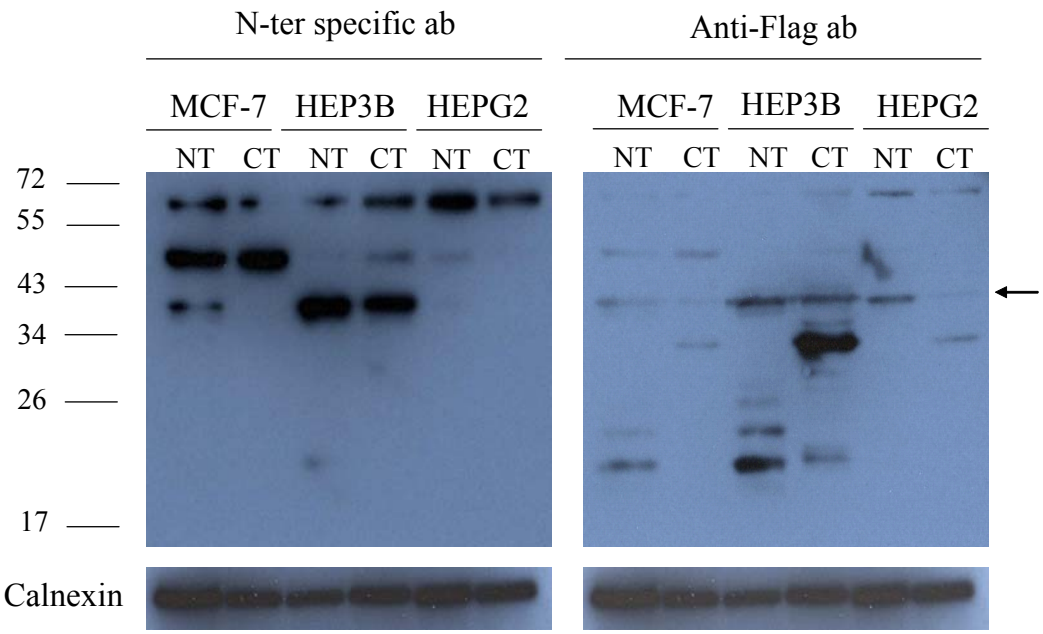

Figure 3.10. Flag epitope labeled HANEIN-1 was analyzed by western blotting. MCF-7, Hep3B and HepG2 cell lines were transfected with plasmids expressing Flag fused HANEIN-1. NT denotes N-ter Flag labeling and CT denotes C-ter Flag labeling. Proteins isolated from transfected cell lines were first denatured via SDS-PAGE, then protein transferred membranes were analyzed by immunoblotting with N-ter specific antibody and anti-flag antibody. Calnexin was used as a loading control. ← displays the Flag tagged HANEIN-1 band, approximately 36 kDa.

3.4.3. Subcellular Localization Analysis via Western Blotting

In order to identify HANEIN-1 protein level expression in nuclear and cytoplasmic parts of the cell, nuclear extract, cytoplasmic extract and total extract obtained from MCF-7 cell line, which were kindly provided by Dr. Hani Alotaibi, were used. Immunoblotting performed with N-ter specific antibody showed that HANEIN-1 might be present in both nuclear extract and cytoplasmic extract (Figure 3.11). In order to be sure about the contents of nuclear and cytoplasmic compartments, membrane was also detected with anti-RAR-α antibody and anti-calnexin antibody. Here, it is expected that retinoic acid receptor-α (RAR-α) localizes in the nucleus and calnexin localizes in the endoplasmic reticulum (ER). Detection with anti-RAR-α antibody gave results as expected, RAR- α was present in nuclear extract and total extract. However, detection with anti-calnexin revealed that calnexin was present in all the compartments analyzed. Based on these results, we thought that calnexin might not be a good marker in order to discriminate nuclear extract since calnexin is an ER protein and nuclear membrane is continuous with ER or another explanation may be that nuclear extract might be somehow contaminated with cytoplasmic extract. Therefore use of a clear cytoplasmic marker would be essential in the future studies.

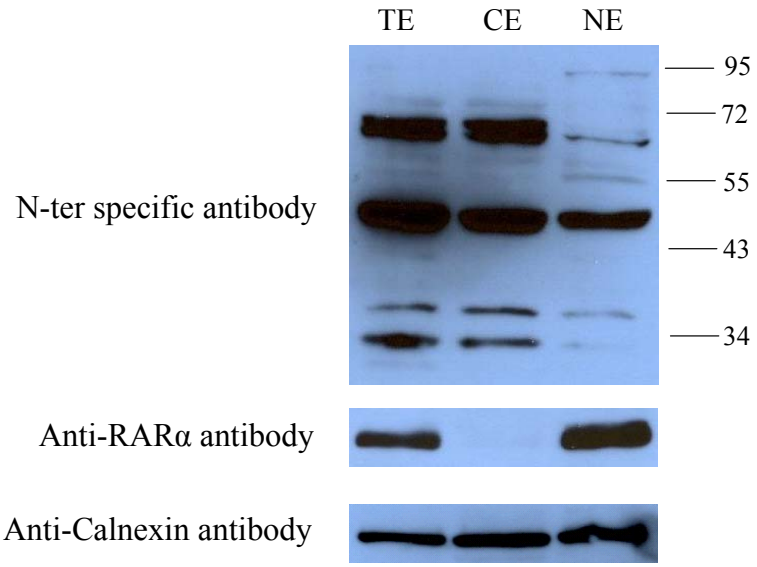

Figure 3.11. Sub-cellular localization analysis of HANEIN-1 via western blotting. Total extract (TE), cytoplasmic extract (CE) and nuclear extract (NE) proteins from MCF-7 cell line were denatured by SDS-PAGE and transferred to PVDF membrane. Membranes were incubated with N-ter specific antibody (for HANEIN-1), anti-RARα antibody and anti-calnexin antibody, respectively. Equal loading was not taken into consideration for this experiment.

3.5. Interaction Partners of HANEIN-1

Yeast double hybrid screening performed in Dr. Uygar Tazebay's lab displayed that HANEIN-1 interacted with RASGEF1B. RASGEF1B is described as a guanine nucleotide exchange factor for Ras-like small GTPases. However, it is not known whether this protein has a similar function with the Son of sevenless (SOS), best characterized example of guanine exchange factors catalyzing the exchange of GDP to GTP and activating Ras (Bonfini *et al.*, 1992) or other known GEFs since GEF domain of RASGEF proteins display characteristic differences when compared to the GEF domains SOS and related Ras activating proteins (Epting *et al.*, 2006).

3.5.1. Bioinformatics Analysis of RASGEF1B

According to Genome Browser database at the University of California Santa Cruz, RASGEF1B has two splice isoforms. First isoform (long form) encodes a 473 amino acid protein and second isoform (short form) encodes a 210 amino acid protein. SMART analysis of this protein showed that first isoform contained both RASGEFN and RASGEF domain, on the other hand, second isoform contained only RASGEFN domain (Figure 3.12). RASGEFN domain is described as alpha helical and playing a purely structural role according to recent crystal structure of Sos (Boriack-Sjodin *et al.*, 1998). RASGEF domain is described as the molecular switch mediating the loss of bound GDP and uptake of GTP in SMART.

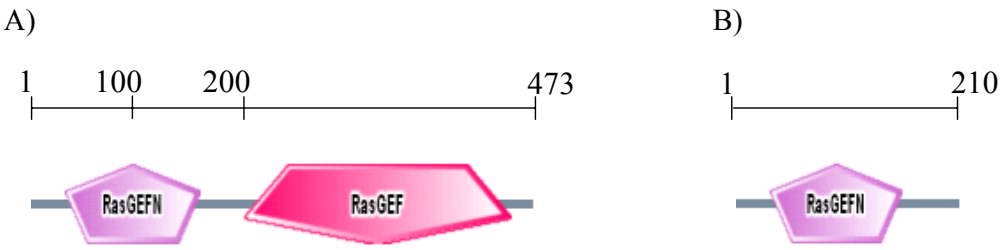

Figure 3.12. Analysis of RASGEF1B domains via SMART. A) First isoform has RASGEFN domain through 33-161 residues, RASGEF domain through 201-454 residues. B) Second isoform contains RASGEFN domain through 33-157 residues.

RASGEF1B is also a conserved protein like its interacting partner, HANEIN-1. Sequence similarity search via FASTA (Pearson *et al.*, 1988) reveals that there are putative uncharacterized proteins having RASGEFN and RASGEF domain even in simple eukaryotes. For instance, human RASGEF1B protein sequence has 40 % and 25 % identities with proteins from *C. elegans* (Uniprot ID: Q21758) and *Aspergillus nidulans* (Uniprot ID: Q5B8H1), respectively.

Expressional analysis of *RASGEF1B* (first isoform) in Oncomine (Rhodes *et al.*, 2004) showed that *RASGEF1B* expression inversely correlated with expression of *HANEIN-1* (Figure 3.13), this observation was extremely remarkable for skeletal muscle since *HANEIN-1* is highly expressed in skeletal muscle (Figure 3.3); on the other hand *RASGEF1B* is expressed at very low levels in skeletal muscle.

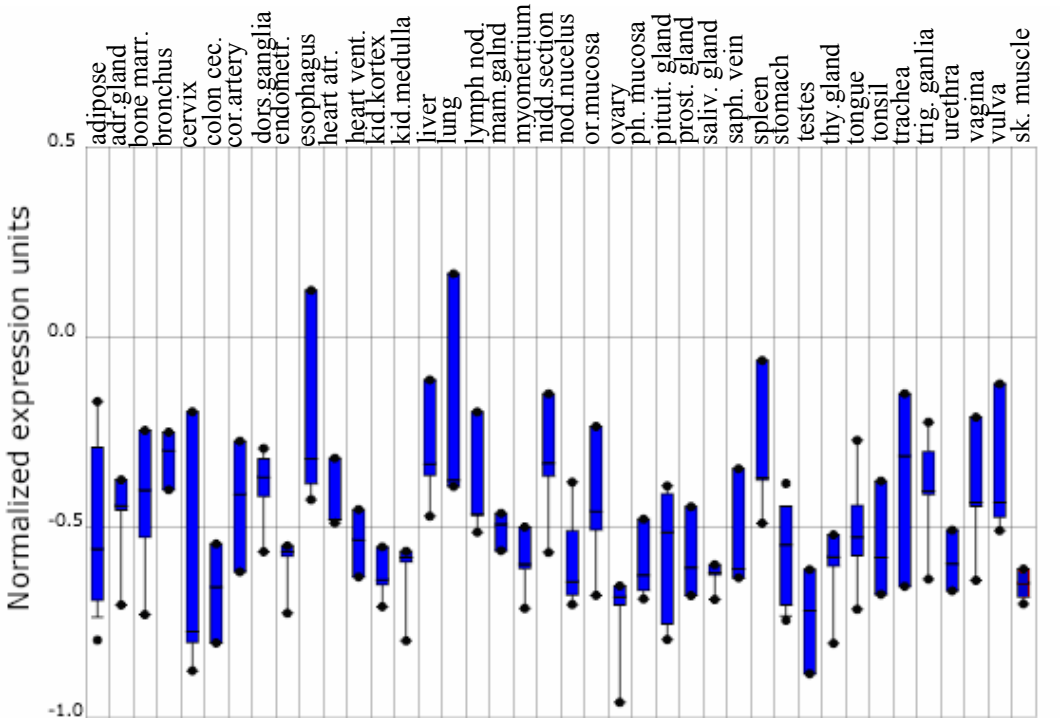

Figure 3.13. Oncomine expression profile of *RASGEF1B* in normal tissues. Bars represent expression status for adipose, adrenal gland, bone marrow, bronchus, cervix, colon cecum, coronary artery, dorsal root ganglia, endometrium, esophagus, heart atrium, heart ventricle, kidney cortex, kidney medulla, liver, lung, lymph nodes, mammary gland, myometrium, nipple cross-section, nodose nucleus, oral mucosa, ovary, pharyngeal mucosa, pituitary gland, prostate gland, salivary gland, saphenous vein, spleen, stomach, testes, thyroid gland, tongue, tonsil, trachea, trigeminal ganglia, urethra, vagina, vulva and skeletal muscle, respectively. Asterisks show the minimum and maximum values obtained for each tissue.

3.5.2. Expressional Analysis of *RASGEF1B*

RASGEF1B (first isoform) expressional analysis was performed in mouse tissues, which were used for *HANEIN-1* expression also. This analysis showed that *RASGEF1B* was expressed ubiquitously expressed in mouse tissues. Another important finding was that in the tissues where *HANEIN-1* expression was high (also presented in Figure 3.6) *RASGEF1B* expression was low (Figure 3.14), which was a quite similar observation obtained with Oncomine *RASGEF1B* expression analysis.

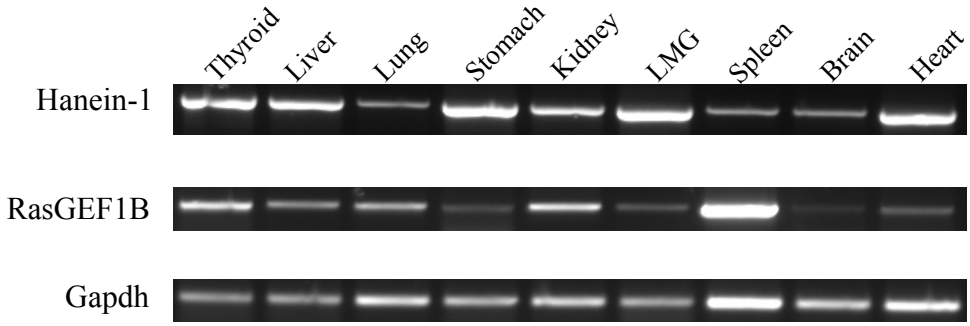

Figure 3.14 Expressional analysis of *RASGEF1B* in mouse tissues. Tissue samples were ground in liquid N_2 and homogenized via teflon-glass homogenizer. RNA isolated from homogenized tissue powder was reverse transcribed and cDNA was amplified with specific primers for *RASGEF1B*. Resulting PCR products were analyzed by agarose gel electrophoresis. *GAPDH* was used as a loading control.

In addition to RT-PCR analysis of *RASGEF1B* expression in mouse tissues, *RASGEF1B* expression was also analyzed by northern blotting (same membrane used for *HANEIN-1* expression). However, *RASGEF1B* expression could not be observed with *RASGEF1B* specific probe on the membrane.

3.5.3. Cloning of *RASGEF1B* and Validation of Interaction via Immunoprecipiation

After obtaining interaction data in yeast double hybrid screening, we decided to validate this interaction by immunoprecipitation. In order to carry out this analysis, *RASGEF1B* two isoforms and its very close homolog *RASGEF1A* were cloned into p3XFLAG-CMV14 expression vector. *RASGEF1B* and *RASGEF1A* gene products were amplified with suitable primers so that resulting products have Not I site and Myc epitope at the 5', and Xba I site at the 3'. Since p3XFLAG-CMV14 plasmid contains Flag epitope just after the Xba I site (Figure 3.15), cloned genes also contain a Flag epitope at the C-terminus (Figure 3.16). The idea behind this cloning strategy is that RASGEF1B interacting proteins can be immunoprecipitated via using Flag epitope and immunoprecipitates can be detected via anti-myc antibody.

Yeast double hybrid screening interaction data was validated by Elif Yaman in MCF- and Huh-7 cell lines via using prepared vectors.

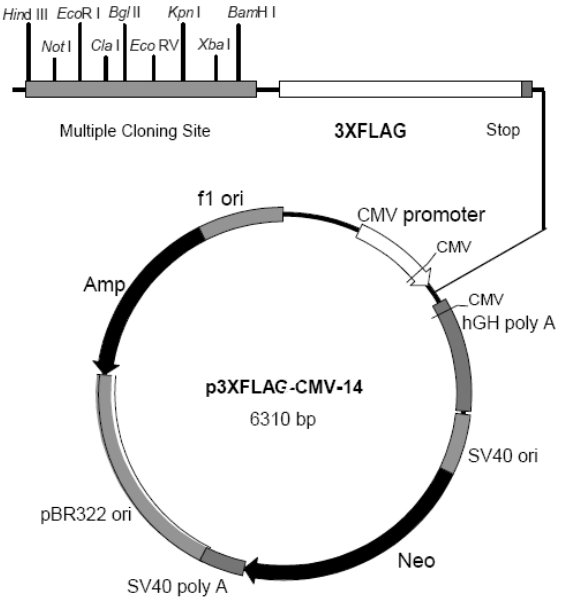

Figure 3.15. Schematic representation of p3XFLAG-CMV14 expression vector

Figure 3.16. Schematic representation of cloned products. Gene Coding DNA Sequence refers to RASGEF1B 1st and 2nd isoform and RASGEF1A. Cloned gene products contain Myc epitope at N-ter and Flag epitope at C-ter.

3.6. Post-translational Modification Analysis of HANEIN-1

Bioinformatics analysis revealed that HANEIN-1 had several highly expected serine, threonine and tyrosine phosphorylation sites. Serine scanning mutagenesis was performed for four serine sites with high probability of phosphorylation at positions 92, 122, 194 and 207 (Table 3.5). Serine sites were site-directed mutagenized via converting serines to alanines with PCR-based experimental strategy (see Materials and Methods). Then, site-directed mutagenized plasmids were transfected to Hep3B cell line. Outcomes of this experiment were analyzed by carrying out a western blot in order to identify the effect of each serine mutagenesis on the stability of the protein.

Western blot analysis carried out with anti-Flag antibody showed that site-directed mutagenesis of the serine site at position 194 resulted in a significant decrease in exogenous HANEIN-1 protein level compared to wild type, in other words, it affected the stability of the protein to a great extent (Figure 3.17). Site-directed mutagenesis of other serine sites also led to a slight decrease in exogenous HANEIN-1 level but these are not as significant as the one at position 194.

In order to exclude the possibility that results of western blot analysis are because of different transfection efficiencies of plasmids, neomycin resistance gene (present in the site-directed mutagenized plasmids (p3XFlag-Locus)) expression, which is expected to be equal for all transfected plasmids, was analyzed by RT-PCR. This analysis revealed that transfection efficiencies were approximately same for each case (Figure 3.17).

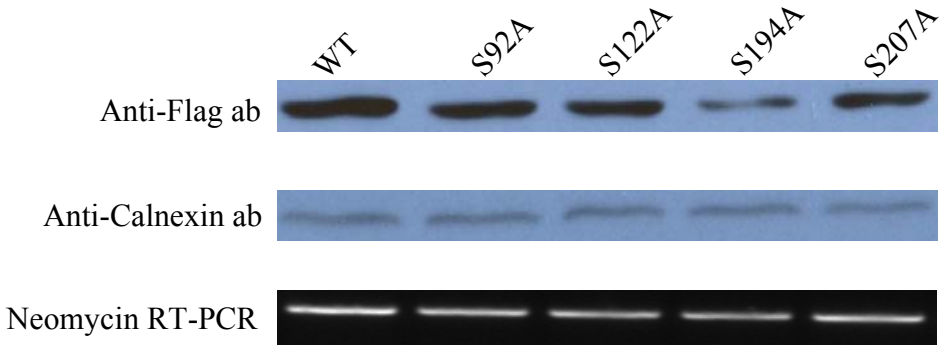

Figure 3.17. Western blot analysis of serine scanning mutagenesis. Serines at position 92, 122, 194 and 207 were site-directed mutagenized and site-directed mutagenized plasmids were transfected to Hep3B liver carcinoma cell line. Proteins isolated from transfected cell lines were analyzed by western blotting with anti-flag antibody. Calnexin was used as an internal control. Neomycin RT-PCR analysis was carried out in order to test transfection efficiencies.

4. DISCUSSION

4.1. Expressional Analysis

Expressional studies carried out in breast carcinoma cell lines, mouse and human tissues showed that *HANEIN-1* was ubiquitously expressed. Northern blot analysis revealed that ubiquitously expressed *HANEIN-1* transcript size was 1061 bp (Figure 3.8). In addition to these, there was an approximately 950 bp form in placenta. At this point, this additional transcript may be formed via alternative splicing or alternative transcription start sites. *HANEIN-1* contains 1 non-coding and 4 coding exons. First exon (non-coding) and fourth exon are relatively short (100 bp) (Figure 4.1). It may be possible that fourth exon is not included in the additional transcript or there may be an additional transcriptional start site in the intron just after the first non-coding exon. Actually, it seems that intronic region between first and second exon may contain a TATA-box like region (Dr. Uygar Tazebay and Dr. Hani Alotaibi, personal communication).

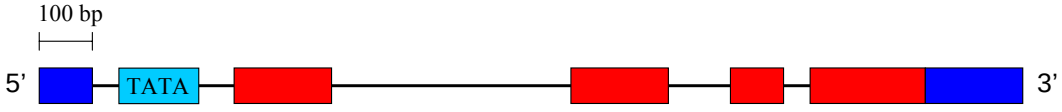

Figure 4.1. Exon and intron structure of *HANEIN-1*. Coding-exons are shown as red boxes and non-coding (5' UTR or 3' UTR) exons are shown as blue boxes. Box sizes are proportional to exon sizes and a representative scale is shown at the top of the first exon. Introns are presented as a line and a putative TATA-box region in the first intron is shown as a light blue box.

Besides providing information about the expression status of *HANEIN-1* in human tissues, northern blotting revealed that both *GAPDH* and β-actin were not good choices as internal controls. We had different *GAPDH* and β-actin expressionprofiles for each tissue although we had equal (2 μg) polyA mRNA on the membrane. In fact, choice of the housekeeping gene for normalizing expression data depends on many factors such as total RNA used and type of the tissues analyzed so it is not expected that selected house-keeping gene is expressed in a similar pattern in every tissue (Herrera *et al.*, 2005).

Both northern blot analysis and Oncomine expression analysis resulted in a remarkable expression of *HANEIN-1* in skeletal muscle. Identification of biological significance of this finding requires more experimental data. However, the first idea that comes to mind is the probability of its involvement in muscle movement, Ca^{++} signaling or glucose utilization since skeletal muscle is in question. At this point, it may be argued that heart is another tissue which has muscle movement, but *HANEIN-1* expression in heart is not as remarkable as in skeletal muscle. Since heart muscle does not depend on motor neurons to be stimulated, HANEIN-1 may be also involved in transmittance of neural impulse to skeletal muscle fiber.

4.2. Immunological Analysis

HANEIN-1 protein size was determined as 33 kDa by performing a western blot with an antibody specific to N-ter of this novel protein. Beside this 33 kDa band, detection with N-ter specific antibody reveals several upper bands also (Figure 3.9). Furthermore, all the bands become lost after the blockage of the antibody with its specific antibody. If all these bands belonged to HANEIN-1, one of the most challenging questions would be why intensity of upper bands did not change after transfection. At this point, upper bands might be the regulated forms of HANEIN-1 so that protein dosage does not change even if cells are transfected with vectors expressing HANEIN-1. SUMOPlot analysis of HANEIN-1 resulted in several possible sumoylation sites (Table 3.4). In this context, upper size bands might be

sumoylated forms of HANEIN-1 and this hypothesis can be investigated via carrying out immunoprecipitation experiments with anti-sumo antibody. In addition to these, HANEIN-1 can be present in the cell as dimers and this question can be addressed via carrying out a native PAGE.

Immunoblotting analysis performed with Flag fused HANEIN-1 showed that addition of a Flag tag affected the stability of the protein. Furthermore, depending on the position of the Flag tag fragmentation pattern of the protein changed (Figure 3.10). Based on these observations a model of proteolytic cleavage can be put forward such that when Flag epitope is at N-terminus, Flag fusion protein (approximately 36 kDa) is cleaved to 20 kDa (detected with anti-Flag) and 16 kDa (not detected with anti-Flag) fragments (Figure 4.1A), on the other hand when Flag epitope is at the C-terminus, a 2 kDa fragment is cleaved from the protein and the remaining 34 kDa protein can be detected with anti-Flag (Figure 4.1B). Upon inspecting Figure 3.9, it can be realized that there are additional cleavage fragments other than the ones included in this model. However, the ones in this model are only the major fragments which can be detected at higher intensities and at all the cell lines studied. Another important question may be the reason why 20 kDa fragment obtained when Flag epitope is at N-ter can not be observed with N-terminus specific antibody. At this point, it can be argued that this observation may be related to film exposure time.

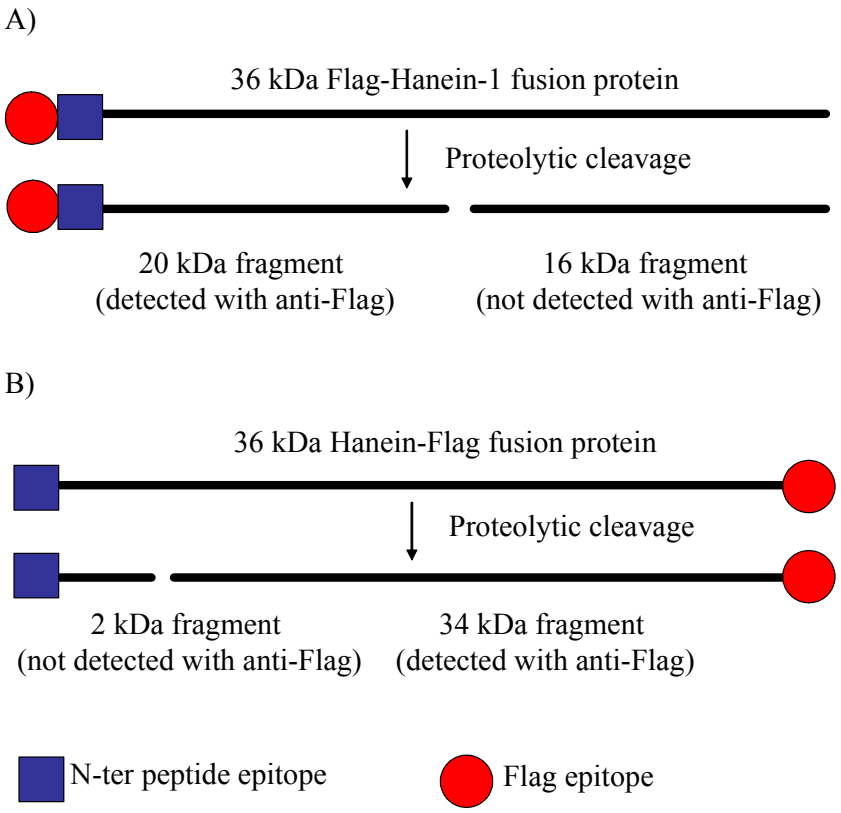

Figure 4.2. Model for proteolytic cleavage pattern of HANEIN-1. A) Flag tag is at N-terminus. B) Flag tag is at C-terminus.

4.3. Subcellular Localization Analysis

In this study, subcellular localization analysis of HANEIN-1 was carried out via immunoblotting. Nuclear, cytoplasmic and total extract proteins were detected with N-ter specific antibody. Results showed that HANEIN-1 definitely localize in cytoplasm. However, we are not quite sure about nuclear localization since we have problems in discrimination of cytoplasmic and nuclear compartments.

Bioinformatics analysis reveals that this protein likely to localize in nucleus with 84% probability. Besides, SMART predicts that HANEIN-1 has DUF1014 domain which is a member of clan HMG-box and HMG-box family contains various DNA binding proteins including transcription factors and subunits of chromatin-

remodeling complexes (Stros *et al.*, 2007). On the other hand, immunofluorescence microscopy analysis performed in our laboratory shows that HANEIN-1 mostly localizes in the cytoplasm. According to microscopy analysis, nuclear localization of HANEIN-1 in the nucleus is very rare, 2 % (Dr. Hani Alotaibi, personal communication. Also confocal microscopy analysis reveals that HANEIN-1 mostly localizes in the perinuclear region (Elif Yaman, personal communication).

Although bioinformatics analysis predicts HANEIN-1 as a nuclear protein, our experimental analysis remarkably shows that HANEIN-1 localizes in the cytoplasm. However, localization of a protein in the cytoplasm does not necessarily mean that cellular function of the protein is in the cytoplasm. For instance, β-catenin localize in the cytoplasm without Wnt signal, on the other hand after the cell receives the signal, it translocates to nucleus and forms a complex with LEF-1 transcription factor (Huber *et al.*, 1996).

4.4. Interacting Partners of HANEIN-1

It was identified that HANEIN-1 interacted with RASGEF1B by yeast two-hybrid screening. Besides, immunoprecipitation experiments performed in our laboratory confirmed this interaction. Although it has a RASGEF domain, mediating the exchange of GDP to GTP, there is not much known about RASGEF1B and it is not known whether it carries out a similar function with Sos. There are only a few studies on RASGEF1B. In one study, RASGEF1B embryonic expression in zebrafish was investigated and it was shown that RASGEF1B expression varied depending on the developmental stage and zygotic expression was regulated by Nodal and FGF signals (Epting *et al.*, 2006). In another study, it was shown that E2F1 upregulated RASGEF1B expression and this RASGEF1B upregulation resulted in Ras activation (Korotayev *et al.*, 2008). However, although this paper provides direct evidence concerning RASGEF1B upregulation by E2F1, direct evidence which shows that Ras activation is in fact carried out by RASGEF1B is not presented. Finally, there is a paper investigating its quite close homolog RASGEF1A

enhanced expression in intrahepatic cholangiocarcinoma and displaying that it activates K-RAS, H-RAS, and N-RAS in vitro (Ura *et al.*, 2006).

In several databases, RASGEF1B is also named with a synonym- GPI gamma-4. GPI gamma-4 was first identified in a study investigating transcripts abundant in macrophages exposed to *Trypanosome cruzi* parasites in mouse (Ferreira *et al.*, 2002). It was a novel single-copy gene conserved in mouse and human genomes, encoding a protein with homology to guanine exchange factors and bearing a Ras-GEFN and Ras-GEF domain. Expression of GPI gamma-4 was also confirmed in mouse and human tissues. It was identified that expression of GPI gamma-4 gene was enhanced during activation of lymphoid and myeloid cells. Furthermore, it was claimed that induction of GPI gamma-4 most probably occurs through NF-κB since promoter region of GPI gamma-4 contained 3 NF-κB binding sites. On the whole, this study may be important in terms of providing some hints about function of HANEIN-1 since our protein interacts with RASGEF1B.

Mouse tissue expression analysis revealed that *RASGEF1B* was expressed ubiquitously like *HANEIN-1*. On the other hand, there was an inverse correlation between expression profiles of *HANEIN-1* and *RASGEF1B*, which was also observed by Oncomine analysis. There may be multiple explanations for the observed transcriptional level expression. However, regarding the cellular functions of RASGEF1B and HANEIN-1, it may be argued that these two proteins interact because one functions to inhibit the functioning of the other based on this inverse gene expression profiles of *HANEIN-1* and *RASGEF1B*.

4.5. Post-translational Modification Analysis

Serine residues of HANEIN-1 with high probability of phosphorylation were site-directed mutagenized and possible involvement of these modifications in the exogenous protein stability was investigated by performing a western blot. This

analysis showed that site-directed mutagenesis of the serine residue at position 194 resulted in a significant decrease in exogenous HANEIN-1 protein level. At this point, it should be noted that site-directed mutagenized plasmids have Flag epitope before HANEIN-1 ORF therefore, exogenous HANEIN-1 was detected with anti-flag antibody at N-terminus. On the other hand, we can not exclude the possibility of obtaining different exogenous HANEIN-1 protein stability profile upon carrying out the same site-directed mutagenesis on HANEIN-1 having the Flag epitope at the C-terminus.

Among the site-directed mutagenized serine residues, serine at position 92 has been shown to be phosphorylated by mass spectrometry (Molina *et al.*, 2007). In this context, our results show that site-directed mutagenesis of serine at position 92 does not seem to affect the protein stability at a significant level although it is phosphorylated.

4.6. Conclusions and Future Perspectives

In conclusion, HANEIN-1 is a highly conserved protein, it is ubiquitously expressed and most probably subject to post-translational modifications. Although a certain conclusion about its function can not be achieved, its interaction with RASGEF1B point out to a possible involvement in Ras pathway.

One of the major goals of our future studies is to knock-down HANEIN-1 by a siRNA strategy and to analyze effects of this disruption via observing cellular morphology, lethality, etc. and investigating possible interference with RASGEF1B interaction and Ras pathway.

Another important point would be to identify whether RASGEF1B activates Ras. This can be easily identified by designing an in vitro Ras dissociation assay (Hall *et al.*, 2001). Once we identify that RASGEF1B activates Ras, we may perform experiments investigating whether HANEIN-1 functions downstream or upstream of Ras. At this point, siRNA knock-down strategy may be very effective. For instance,

if HANEIN-1 functions upstream of Ras, knock-down of HANEIN-1 may affect the activation of Ras at a certain degree after the Ras activating signal (for example EGF) is received. On the other hand, if HANEIN-1 functions downstream of Ras, down-regulation of HANEIN-1 may affect the activation of Ras effectors and this can in turn be reflected by expression profiles of genes activated by Ras.

Finally, one of the long term goals is to conditionally knock-out HANEIN-1 in mouse. With this study, we aim to characterize HANEIN-1 function at the organism level and to analyze the effect of HANEIN-1 function loss in a tissue specific manner.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25,** 3389-3402.

Apweiler, R., Hermjakob, H. & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473,** 4-8.

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C. H. & Paulson, J. C. (2004). Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A* **101,** 17033-17038.

Blom, N., Gammeltoft, S. & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294,** 1351-1362.

Bonfini, L., Karlovich, C. A., Dasgupta, C. & Banerjee, U. (1992). The Son of sevenless gene product: a putative activator of Ras. *Science* **255,** 603-606.

Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. (1998). The structural basis of the activation of Ras by Sos. *Nature* **394,** 337-343.

Brummelkamp, T. R., Bernards, R. & Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296,** 550-553.

Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L. & Dubchak, I. (2003). Strategies and tools for whole-genome alignments. *Genome Res* **13,** 73-80.

Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K. & McMahon, A. P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr Biol* **8,** 1323-1326.

Dell, A. & Morris, H. R. (2001). Glycoprotein structure determination by mass

spectrometry. *Science* **291,** 2351-2356.

Duckworth, W. C., Stentz, F. B., Heinemann, M. & Kitabchi, A. E. (1979). Initial site of insulin cleavage by insulin protease. *Proc Natl Acad Sci U S A* **76,** 635-639.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494-498.

Epting, D., Vorwerk, S., Hageman, A. & Meyer, D. (2007). Expression of RASGEF1B in zebrafish. *Gene Expr Patterns* **7,** 389-395.

Ferreira, L. R., Abrantes, E. F., Rodrigues, C. V., Caetano, B., Cerqueira, G. C., Salim, A. C., Reis, L. F. & Gazzinelli, R. T. (2002). Identification and characterization of a novel mouse gene encoding a Ras-associated guanine nucleotide exchange factor: expression in macrophages and myocarditis elicited by Trypanosoma cruzi parasites. *J Leukoc Biol* **72,** 1215-1227.

Fields, S. & Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245-246.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806-811.

Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. & Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **21,** 566-569.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. & Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **265,** 103-106.

Hall, B. E., Yang, S. S., Boriack-Sjodin, P. A., Kuriyan, J. & Bar-Sagi, D. (2001). Structure-based mutagenesis reveals distinct functions for Ras switch 1 and switch 2 in Sos-catalyzed guanine nucleotide exchange. *J Biol Chem* **276,** 27629-27637.

Haupt, Y., Maya, R., Kazaz, A. & Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature* **387,** 296-299.

Herrera, F., Martin, V., Antolin, I., Garcia-Santos, G., Rodriguez-Blanco, J. & Rodriguez, C. (2005). Standard curve for housekeeping and target genes: specific criteria for selection of loading control in Northern blot analysis. *J Biotechnol* **117,** 337-341.

Huber, O., Korn, R., McLaughlin, J., Ohsugi, M., Herrmann, B. G. & Kemler, R. (1996). Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. *Mech Dev* **59,** 3-10.

Huynen, M., Snel, B., Lathe, W., 3rd & Bork, P. (2000). Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* **10,** 1204-1210.

Julenius, K., Molgaard, A., Gupta, R. & Brunak, S. (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* **15,** 153-164.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* **12,** 996-1006.

Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W. & Shokat, K. M. (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* **21,** 1047-1054.

Kohler, G. & Milstein, C. (1992). Continuous cultures of fused cells secreting antibody of predefined specificity. 1975. *Biotechnology* **24,** 524-526.

Korotayev, K., Chaussepied, M. & Ginsberg, D. (2008). ERK activation is regulated by E2F1 and is essential for E2F1-induced S phase entry. *Cell Signal* **20,** 1221-1226.

Kubbutat, M. H., Jones, S. N. & Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature* **387,** 299-303.

Kubbutat, M. H. & Vousden, K. H. (1997). Proteolytic cleavage of human p53 by calpain: a potential regulator of protein stability. *Mol Cell Biol* **17,** 460-468.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947-2948.

Lee, D., Redfern, O. & Orengo, C. (2007). Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol* **8,** 995-1005.

Liolios, K., Tavernarakis, N., Hugenholtz, P. & Kyrpides, N. C. (2006). The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res* **34,** D332-334.

Manzi, A. E., Norgard-Sumnicht, K., Argade, S., Marth, J. D., van Halbeek, H. & Varki, A. (2000). Exploring the glycan repertoire of genetically modified

mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. *Glycobiology* **10,** 669-689.

Miyawaki, A. & Tsien, R. Y. (2000). Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol* **327,** 472-500.

Molina, H., Horn, D. M., Tang, N., Mathivanan, S. & Pandey, A. (2007). Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* **104,** 2199-2204.

Mueller, M., Martens, L. & Apweiler, R. (2007). Annotating the human proteome: beyond establishing a parts list. *Biochim Biophys Acta* **1774,** 175-191.

Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M. & Nygren, P. A. (1997). Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr Purif* **11,** 1-16.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V. & Falquet, L. (2004). MyHits: a new interactive resource for protein annotation and domain identification. *Nucleic Acids Res* **32,** W332-335.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85,** 2444-2448.

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* **96,** 4285-4288.

Perriere, G., Combet, C., Penel, S., Blanchet, C., Thioulouse, J., Geourjon, C., Grassot, J., Charavay, C., Gouy, M., Duret, L. & Deleage, G. (2003). Integrated databanks access and sequence/structure analysis services at the PBIL. *Nucleic Acids Res* **31,** 3393-3399.

Puntervoll, P., Linding, R., Gemund, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D. M., Ausiello, G., Brannetti, B., Costantini, A., Ferre, F., Maselli, V., Via, A., Cesareni, G., Diella, F., Superti-Furga, G., Wyrwicz, L., Ramu, C., McGuigan, C., Gudavalli, R., Letunic, I., Bork, P., Rychlewski, L., Kuster, B., Helmer-Citterich, M., Hunter, W. N., Aasland, R. & Gibson, T. J. (2003). ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res* **31,** 3625-3630.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. & Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6,** 1-6.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular

architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95,** 5857-5864.

Sivashankari, S. & Shanmughavel, P. (2006). Functional annotation of hypothetical proteins - A review. *Bioinformation* **1,** 335-338.

Stros, M., Launholt, D. & Grasser, K. D. (2007). The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci* **64,** 2590-2606.

Tazebay, U. H., Wapnir, I. L., Levy, O., Dohan, O., Zuckier, L. S., Zhao, Q. H., Deng, H. F., Amenta, P. S., Fineberg, S., Pestell, R. G. & Carrasco, N. (2000). The mammary gland iodide transporter is expressed during lactation and in breast cancer. *Nat Med* **6,** 871-878.

Thomas, K. R. & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51,** 503-512.

Tsien, R. Y. (1998). The green fluorescent protein. *Annu Rev Biochem* **67,** 509-544.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skollermo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S. & Ponten, F. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **4,** 1920-1932.

Ura, K., Obama, K., Satoh, S., Sakai, Y., Nakamura, Y. & Furukawa, Y. (2006). Enhanced RASGEF1A expression is involved in the growth and migration of intrahepatic cholangiocarcinoma. *Clin Cancer Res* **12,** 6611-6616.

von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. & Bork, P. (2007). STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35,** D358-362.

Yarema, K. J. & Bertozzi, C. R. (2001). Characterizing glycosylation pathways. *Genome Biol* **2,** REVIEWS0004.

Yurchenko, V., Xue, Z. & Sadofsky, M. J. (2006). SUMO modification of human

XRCC4 regulates its localization and function in DNA double-strand break repair. *Mol Cell Biol* **26,** 1786-1794.

Zhang, J. & Hua, Z. C. (2004). Targeted gene silencing by small interfering RNA-based knock-down technology. *Curr Pharm Biotechnol* **5,** 1-7.