Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 11—22

http://www.ripublication.com/jcib.htm

In-Silico Probing of Molecular Genetic Factors of Cerebellum Rhabdoid Tumour in Human.

Aditi Bajpayee[1] and Saradindu Ghosh[2]

[1]Department of Biotechnology, Integral University,
Lucknow, Uttar Pradesh-226026, India
E-mail: abajpayee85@gmail.com
[2]Department of Biotechnology,
Majhighariani Institute of Technology and Science,
Rayagada, Orissa-765017, India
E-mail: ghosh.saradindu@gmail.com

Abstract

The similarities of INI1 gene of Homo sapiens with that of model organisms from NCBI have evolutionary relationship between different organisms and properties of the gene respectively. Malignant rhabdoid tumour (MRT) is caused due to the mutation in the INI1 gene in human. From sequence analysis we found our protein is closely related to Bos taurus and Rattus norvegicus while distantly related to Cryptosporidium parvum. From structure analysis we found the query protein is unstable with molecular weight: 43158.40 which is hydrophilic in nature and contains more numbers of random coils. SMARCB1 may be a tumour suppressor gene for malignant rhabdoid tumour. Results demonstrate that deletions and mutations of the INI1 gene can occur in rare composite rhabdoid tumours of adulthood. All conserved domains of INI1/hSNF5/BAF47 are needed for CSF1 transcription and INI1/ hSNF5/BAF47 is recruited to the region of the CSF1 promoter. The tumour suppressor gene hSNF5 was lacking in the malignant rhabdoid tumour of the liver. Chromatin remodelling factor encoded by INI1 induces G1 arrest and apoptosis in INI1-deficient cells. As Germline hSNF5 mutation is associated with rhabdoid predisposition syndrome, it strongly suggests that the SNF5 homology domain presents species-specific functions.

Keywords: MRT, SMARCB1, INI1, Germline hSNF5, SCOP, SMART, PROSITE.

Introduction

Malignant rhabdoid tumor (MRT) is one of the most aggressive and lethal malignancies in pediatric oncology. Malignant rhabdoid tumor was initially described in 1978 as a rhabdomyosarcomatoid variant of a Wilms tumor because of its occurrence in the kidney and because of the resemblance of its cells to rhabdomyoblasts. The absence of muscular differentiation led Haas and colleagues to coin the term rhabdoid tumor of the kidney in 1981[1].

Although renal malignant rhabdoid tumor was historically included in treatment protocols of the National Wilms Tumor Study (NWTS) Group, this tumor is now recognized as an entity separate from a Wilms tumor. In contrast to a Wilms tumor, a malignant rhabdoid tumor of the kidney is characterized by the early onset of local and distant metastases and resistance to chemotherapy. Whereas the overall survival rate for Wilms tumors exceeds 85%, the survival rate for renal malignant rhabdoid tumors is only 20-25%.

Because rhabdoid tumor of the kidney was originally described, malignant rhabdoid tumors have been reported in practically every location in the body, including the brain, liver, soft tissues, lung, skin, and heart. Cytogenetic, fluorescence in situ hybridization (FISH), and loss-of-heterozygosity (LOH) studies have revealed that malignant rhabdoid tumors frequently contain deletions at chromosome locus 22q11.1. Positional cloning efforts revealed that this locus contains the *SWI/SNF* related, matrix-associated, actin-dependent regulator of chromatin, subfamily B, member 1 (*SMARCB1*) gene, also known as human sucrose nonfermenting gene number 5 (*hSNF5*), integrase interactor 1 (*INI1*) , or 47-Kd *Brg1/Bam–* associated factor (*BAF47*)[2]. *SMARCB1* encodes a member of the human *SWI/SNF* complex. Combined analyses including FISH, coding sequence analysis, high-density single nucleotide polymorphism–based oligonucleotide arrays, and multiplex ligation-dependent probe amplification enable the identification of biallelic, inactivating perturbations of *SMARCB1* in nearly all malignant rhabdoid tumors, consistent with the 2-hit model of tumor formation [3]. Thus, *SMARCB1* is presumed to function as a classic tumor suppressor and the primary gene responsible for malignant rhabdoid tumor development.

Homozygous inactivation of *SMARCB1* in mice demonstrates embryonic lethality, whereas heterozygous *SMARCB1* mice demonstrate a normal phenotype at birth, with 20% developing sarcomas at a median age of 1 year. Similar to human malignant rhabdoid tumors, murine tumors in these mice acquire a second hit to the *SMARCB1* locus. All mice harboring a conditional biallelic inactivation of *SMARCB1* develop cancer with a median onset of 11 weeks, revealing one of the most aggressive cancer predisposition genotype-phenotype correlations known. Unexpectedly, despite an aggressive clinical pattern of behavior, malignant rhabdoid tumors are generally diploid and genomically stable, without recurrent gene amplifications or deletions. The mechanism by which *SMARCB1* perturbation leads to aggressive neoplasia therefore likely relates to its role in epigenetic modification. The *SWI/SNF* complex acts in an adenosine triphosphate (ATP)–dependent manner to remodel chromatin, which regulates gene transcription and DNA repair. Reports to date have demonstrated that *SMARCB1* loss can promote cell cycle progression resulting from

upregulation of targets of the p16INK4a-Rb-E2F pathway. Rb family loss has been shown to increase malignant rhabdoid tumor tumorigenesis and progression, whereas ablation of *CyclinD1* abrogates malignant rhabdoid tumor evolution in mouse models. Similarly, tumour development in *SMARCB1* -deficient mice is greatly accelerated in the absence of functional p53 protein. These findings suggest a cooperative effect between *SMARCB1* and the *pRB*, *CyclinD1,* and *p53* pathways [4-5].

The present study has been carried out to investigate the molecular genetics of Rhabdoid tumour occurring within the cerebellum and to detect the gene responsible for the same. The objective of this study is to investigate the role of genetic factors using different in-silico tools. The role and relationship of different genes responsible for the rhabdoid tumour has to be identified to get the interaction of them [6].The sequence analysis, phylogenetic analysis and structural analysis is necessary to establish the similarity comparison and to establish evolutionary relationship of sequence in question to the model organisms that has already been well characterized [7]. It is important to know about the mutations in the gene responsible [8]. The pathway for INI1 gene processing should be understood to get the better insight of the rhabdoid tumour.

Materials and Methods

The sequence similarity is being searched by using Basic Local Alignment Search Tool [9-10]. Evolutionary relationship established in between the sequence in question and stored database sequences by using ClustalW [11]. Texshade and Boxshade are being performed to get the conserved residue of the query [12]. The structure of the query is being studied by using Protparam, Hierarchial neural network (HNN), CPH model, HH pred & 3D-pssm to determine primary to tertiary structures [13-15]. Understood the structures using RasMol and Swiss PDB viewer. [16-17]. Performed protein domain and family classification by Prodom and pfam [18, 19]. Structure of the protein is being understood by using SCOP and CATH [20-21]. Signal peptide cleavage site and sub cellular location of protein understood by Signal p and Target P respectively [22-23]. Protein domain architecture was revealed by SMART [24] while Psort revealed the protein sorting signal [25]. SOSUI [26] has been used to get the protein solubility when Gene Card revealed the expressional level of the gene in different tissues [27].

Results and Discussion

INI1 GENE (GI Number 121946732) was obtained from NCBI, of sequence length 376 amino acids.From phylogenetic analysis we found the evolutionary relationship of the model organisms to the INI1 gene by the help of dendogram (Fig 1) on the basis of function. We observed the conserved similar, identical, non-conserved regions by the help of Texshade (Fig 2) & Boxshade (Fig 3). On the basis of molecular clock we found the evolutionary relationship between the model organisms and INI1 gene by using the tool Clustal Distance Matrix.

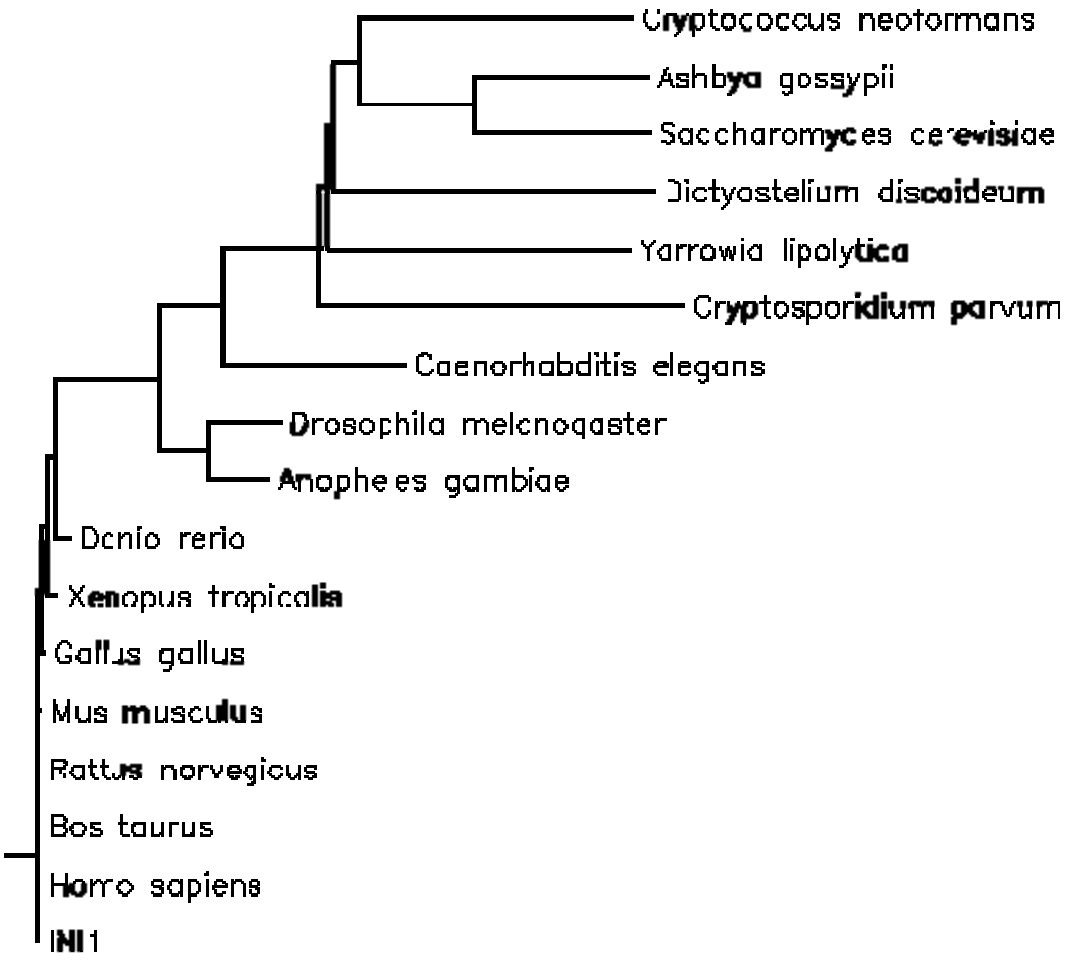


Figure 1: Relationship between Human INI1 gene and the other BLAST significant sequences.

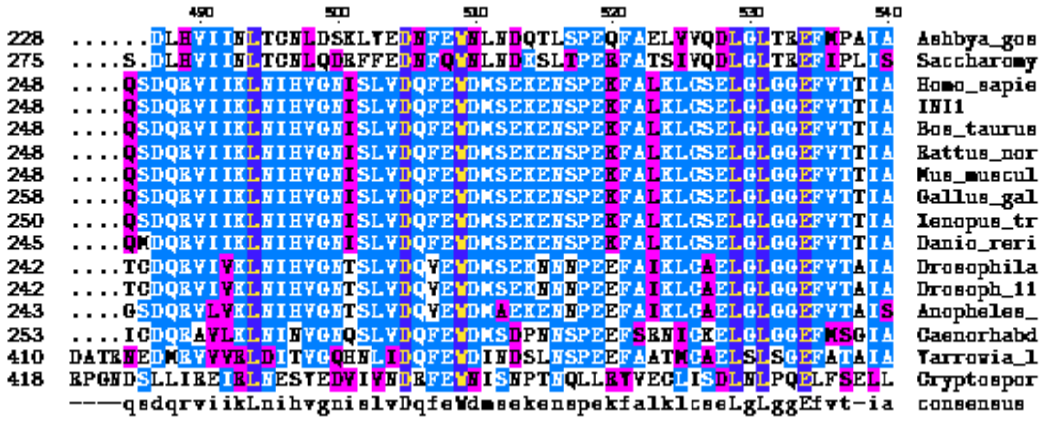


Figure 2: Finding conserved sequence in query with database (TEXSHADE RESULT: Violet: identical, Blue: conserved, Pink: similar, White: non-conserved).

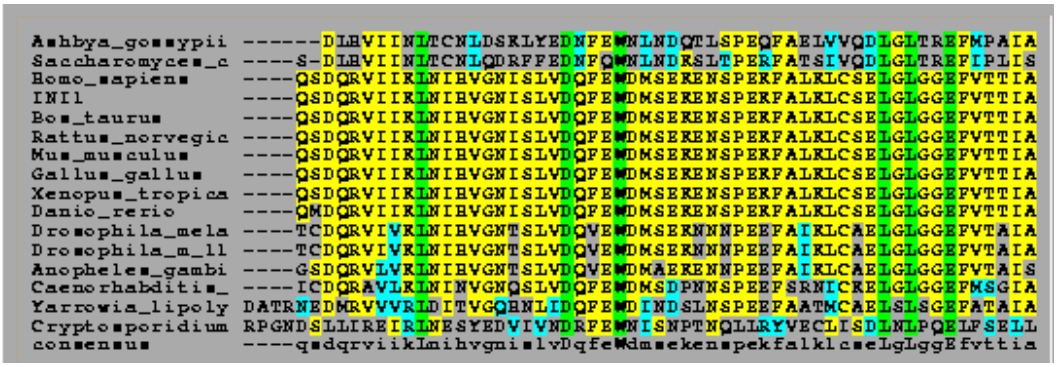
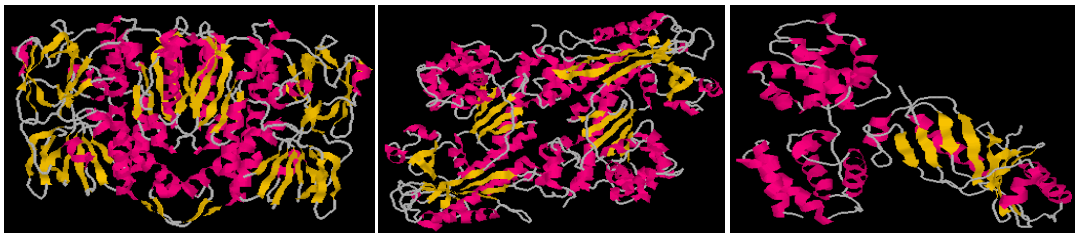


Figure 3: Conserved amino acid sequences among the significant sequences (Boxshade, Green: fully conserved, Yellow: identical, Cyan: similar, Black: non-conserved).

In structural analysis we have done primary, secondary and tertiary analysis. From primary structural analysis we found the instability index as 50.26 which classify our protein as unstable and the GRAVY value as -0.530 which implies protein is Hydrophilic and the theoretical PI value at 5.44 by using the tool Protparam. From secondary structure analysis we got the information about the number of alpha-helix, beta-sheet and random coils having, 130, 70 , 176 residues respectively. The numbers of random coils found are more. From tertiary structure analysis we got different PDB IDs by using tools such as CPH model, HH-Pred &3D pssm (Fig 4). By using these IDs we have predicted the structure of protein.



a b c

Figure 4: TERTIARY STRUCTURE ANALYSIS a- CPH model (2A8X), b- HHPred model (2R8J), c- 3DPSSM model of query protein (1VOL).

With the SCOP result, it was found that the query protein belongs to Cyclin superfamily with duplication in two fold, consisting all alpha helix (5) as a core of protein. The targetted protein may act as transcription factor II D (TFIID) in human (TaxId: 9606). While CATH depicted it belongs to mainly alpha helix (1), orthogonal Bundle (1.10) of Cyclin A domain1 (1.10.472) with Homologous Superfamily (1.10.472.10).

SOSUI is a free online tool that predicts a part of the secondary structure of proteins from a given amino acid sequence (AAS). The main objective is to determine

whether the protein in question is a soluble or a trans-membrane protein. SOSUI predicts that the query in question is a soluble protein with hydrophobicity -0.0530053.

PFAM is a collection of protein families and domains. Pfam contains multiple protein alignments and profiles-HMMs of these families. Pfam is a semi-automatic protein family database, which aims to be comprehensive as well as accurate. Protein sorting signal is a computer program for the prediction of protein localization site in the cell Here we got our target is very much belongs to SNF5 family and it may be a nuclear protein.

From the Prodom results protein domains are found. This ProDom domain is producing High-scoring Segment Pairs. Query protein is closest with PD017359 protein domain with actin member dependent subfamily, having E value- 7e-97 and score 903. Target P 1.1 server predicts the sub-cellular location of eukaryotic protein. The location assignment based on the predicted presence of any of the N-terminal pre-sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide. The protein belongs to mTP. Signal P 3.0 server predicts the presence and location of signal peptide cleavage site in amino acid sequence from different organism: Gram positive prokaryotes and eukaryotes. The method incorporates a prediction of cleavage site and signal peptide/ non-signal peptide prediction based on a combination of several artificial neural networks in Hidden –Markov models (Fig 5)

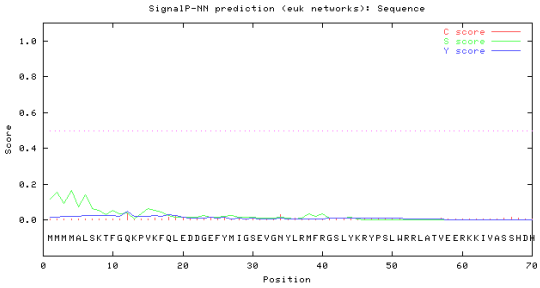


Figure 5: a- Signal P NN output value.

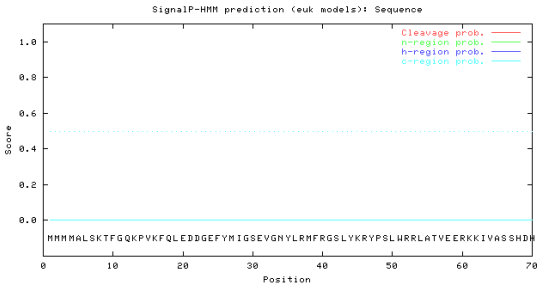


Figure 5: b-Signal P HMM output value, Sequence Prediction: Non-secretory protein, Signal peptide probability: 0.000, Signal anchor probability: 0.000, Max cleavage site probability: 0.000 between pos. 18 and 19.

SMART can be used to explore domain architectures and find exact domains counts in various genomes. Domains within *Homo sapiens* protein Q17S11_HUMAN (Q17S11)(SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1). This domain starts with 138, end at 364 number amino acid, showing E value 3.00e-153 at 95 percent confidence level to produce an overlap. PROSITE is the database of protein families, domains and functional sites as well as associated patterns and profile to identify them found: 19 hits in 1 sequence (Table 1).

Table 1: PROSITE output showing the grove responsible for potential mutational position.

Id	Site name	Residue No	Pattern
PS00008	MYRISTYL	29-34	GSevGN
PS00006	CK2_PHOSPHO_SITE	56-59	TveE
		66 - 69	SshD
		85 - 88	SevE
		172 - 175	SqpE
		243 - 246	SilE
		265 - 268	SlvD
		275 - 278	SekE
		346 - 349	TltD
		348 - 351	TdaE
PS00004	CAMP_PHOSPHO_SITE	117 - 120	KRnS
		151 - 154	KKrT
PS00001	ASN_GLYCOSYLATION	128 - 131	NSSH
		170 - 173	NASQ
		263 - 266	NISL
PS00005	PKC_PHOSPHO_SITE	275 - 277	SeK
		305 - 307	SiR
		363 - 365	TrR

Gene Cards

The protein yields similarities with INI1 gene (SNF5 sucrose non fermenting, yeast, homolog)-like 1) with Entrez Gene id 6598(Fig 6). While validating with Swiss Prot we found it validates with SNF5_Human, Q12848. Swiss Prot Recommended Name: SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1 Size: 385 amino acids; 44141 Dalton. Sub cellular location: Nucleus, Alternative splicing: 2 isoforms: Q12824-1 Q12824-2. Again database search of OMIM gives the gene symbol SMARCB1 (Table 2).

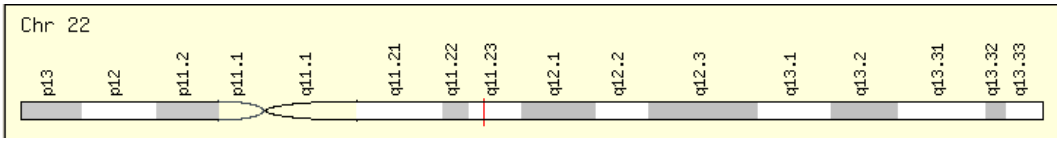


Figure 6: GENE CARD data *Entrez Gene cytogenetic band:* 22q11.23|22q11 *Ensembl cytogenetic band:* **22q11.23** *HGNC cytogenetic band:* 22q11.2.

Table 2: Search Results for GC22P022453: SMARCB1

GeneCards ID	GC22P022453
Description	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1
Gene Symbol	SMARCB1
Ensembl ID	ENSG00000099956
LocusLink ID	6598
Chromosomal Location	Chromosome: 22, Orientation: +, Start: 22459150, End: 22506705

Defects in SMARCB1 are a cause of rhabdoid tumour (RDT) [MIM:609322]; also called malignant rhabdoid tumour (MRT). Tumour suppressor. Inactivated in rhabdoid tumours. Rhabdoid tumours are a highly malignant group of neoplasms that usually occur in early childhood. SMARCB1/INI1 is also frequently inactivated in epithelioid sarcomas. It belongs to the SNF5 family. INI gene is highly expressed in lung, liver & bone marrow in comparison with other tissues (Fig7).

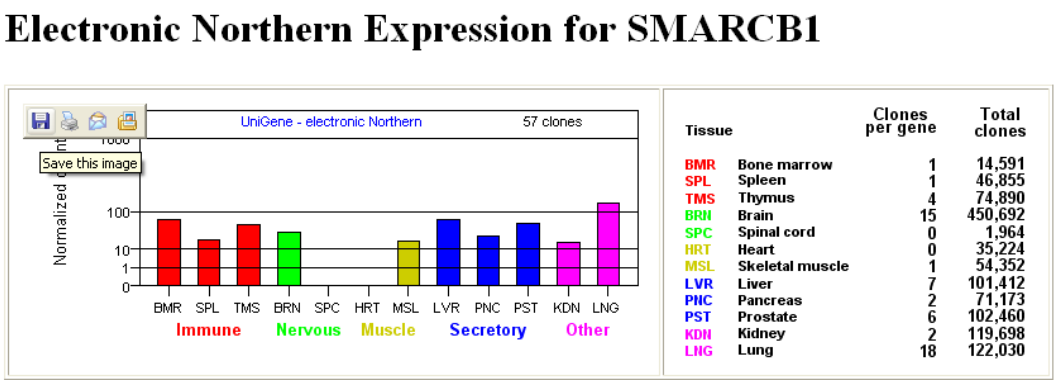


Figure 7: e Northern expression system for SMARCB1 at tissues.

The results show a similar mechanism of tumorigenesis due to loss of one INI1 allele and mutation of the remaining allele were followed by the development of the rhabdoid tumour component. The rhabdoid components in these tumours are, in

contrast to paediatric rhabdoid tumours, typically immune reactive with the BAF47 antibody. Deletions of 22q were detected in 10/16 rhabdoid meningiomas but deletions or losses of 22q are characteristic for meningiomas regardless of their histologic appearance. Deletion of 22q was detected in a case of rhabdoid carcinoma of the small intestine, but this tumour remained immunoreactive for the INI1 protein. Composite rhabdoid tumours of endometrium are rare and clinically aggressive. Only four cases have been reported.

Recent studies have shown that cyclin D1 is over expressed in rhabdoid tumours with inactivated INI1 .Over expression of cyclin D1 in the composite endometrial rhabdoid tumour with an inactivated INI1 gene, in contrast to low expression of cyclin D1 in the tumour with expressed INI1, was also consistent with the molecular etiologic of this tumour. Cyclin D1 plays a key role in cell cycle regulation during the G1 to S phase transition by binding to cyclin-dependent kinases 4 and 6, resulting in phosphorylation and inactivation of the tumour suppressing retinoblastoma protein RB and entrance of cells to S phase . The role of cyclin D1 as a key mediator in the genesis of rhabtumoursumors was supported by a study of mice with disrupted Ini1 and cyclin D1 genes. Ini1+/- mice with cyclin D1 deficiency did not develop rhabdoid tumours, in contrast to the parental Ini1+/- mice, supporting the hypothesis that cyclin D1 is necessary for tumorigenesis. Therapeutic targeting of cyclin D1 induced G1 arrest and apoptosis in rhabdoid cell lines and resulted in growth inhibition of xenografted rhabdoid tumours. Although specific targeting of cyclin D1 in human tumour cells is not yet feasible, this approach may offer a new strategy for treatment of paediatric rhabdoid tumours and for composite rhabdoid tumours with inactivated INI1 genes.

We hypothesized that if this cerebellar tumor is a PNET with multilineal differentiation its cells might show a cytogenetic abnormality more consistent with this diagnosis than with a diagnosis of germ cell tumor. An isochromosome 17q (i17q) is present in about one third of medulloblastomas but is not characteristic of intracranial germ cell tumors.

Conclusion

In this study, we have done the similarity search of the INI1 gene of Homo sapiens with that of model organisms from NCBI by using BLAST. Through the phylogenetic analysis & structural analysis we found the evolutionary relationship between different organisms and properties of the gene respectively. By using some other tools we found the class, architecture, topology, homology, family, super-family, domain etc. From the whole work we found that the main cause is the mutation in the INI1 gene. From sequence analysis we found our protein is closely related to Boss Taurus and Rates norvegicus. Our protein is distantly related to Cryptosporidium parvum. From structure analysis we found our protein unstable with molecular weight: 43158.40 which is hydrophilic in nature and contains more numbers of random coils. Aggressive therapy has prolonged the natural history in a subset of children. SMARCB1 may be a tumour suppressor gene for malignant rhabdoid tumour. Transmission of a germline INI1-mutation in a rhabdoid tumour predisposition

syndrome family via nonpenetrant males. Inactivation of the SMARCB1/INI1 gene is associated with rhabdoid tumour. Loss of INI1 expression may rarely be encountered in tumours undergoing malignant transformation, but this is accompanied by mutation in the INI1 gene. Aberrations of the hSNF5/INI1 gene are restricted to malignant rhabdoid tumours. Results demonstrate that deletions and mutations of the INI1 gene can occur in rare composite rhabdoid tumours of adulthood. By interacting with IN, SNF5/Ini1 interferes with early steps of HIV-1 infection.INI1 immunohistochemistry is a relatively simple, sensitive, and specific technique for distinguishing malignant rhabdoid tumour. All conserved domains of INI1/hSNF5/BAF47 are needed for CSF1 transcription and INI1/ hSNF5/BAF47 is recruited to the region of the CSF1 promoter. The tumour suppressor gene hSNF5 was lacking in the malignant rhabdoid tumour of the liver. Chromatin remodelling factor encoded by ini1 induces G1 arrest and apoptosis in ini1-deficient cells. INI1 is dispensable for retrovirus-induced cytoplasmic accumulation of PML and does not interfere with virus integration. Germline hSNF5 mutation is associated with rhabdoid predisposition syndrome. This strongly suggests that the SNF5 homology domain presents species-specific functions.

Acknowledgement

We are thankful to Chairman, Integral University (Lucknow) and Majhighariani Institute of Technology & Science to provide us the lab for computing facility. We are really thankful to our friend Mr. Ashish Gupta for being with us throughout the endeavour as friend and well wisher.

References

[1] Haas JE, Palmer NF, Weinberg AG, Beckwith JB. Ultrastructure of malignant rhabdoid tumor of the kidney. A distinctive renal tumor of children. Hum Pathol. Jul 1981;12(7):646-57.

[2] Roberts CW, Biegel JA. The role of SMARCB1/INI1 in development of rhabdoid tumor. Cancer Biol Ther. Mar 2009;8:412-6.

[3] Jackson EM, Sievert AJ, Gai X, et al. Genomic analysis using high-density single nucleotide polymorphism-based oligonucleotide arrays and multiplex ligation-dependent probe amplification provides a comprehensive analysis of INI1/SMARCB1 in malignant rhabdoid tumors. Clin Cancer Res. Mar 2009;15:1923-30. .

[4] James I Geller, Nancy D. Leslie, Hong Yin Malignant Rhabdoid Tumor, emedicine Dec 18, 2009.

[5] Sasaki A, Kurihara H, Ishiuchi S, Hirato J, Saito N, Nakazato Y. Pediatric embryonal tumour of the cerebellum with rhabdoid cells and novel intracytoplasmic inclusions: distinction from atypical teratoid/rhabdoid tumour. Acta Neuropathol. 2005 Jul;110(1):69-76.

[6] Charles W. M. Roberts & Stuart H. Orkin, The SWI/SNF complex — chromatin and cancer, Nature Reviews Cancer 4, 133-142 (February 2004) | doi:10.1038/nrc1273

[7] Song N, Joseph JM, Davis GB, Durand D, 2008 Sequence Similarity Network Reveals Common Ancestry of Multidomain Proteins. PLoS Comput Biol 4(5): e1000063. doi:10.1371/journal.pcbi.1000063

[8] Lévy G, Levi-Acobas F, Blanchard S, Gerber S, Larget-Piet D, Chenal V, Liu XZ, Newton V, Steel KP, Brown SD, Munnich A, Kaplan J, Petit C, Weil D, Myosin VIIA gene: heterogeneity of the mutations responsible for Usher syndrome type IB. Hum Mol Genet. 1997 Jan;6(1):111-6.

[9] NCBI (Homepage on the internet).USA: National Centre for Biotechnology Information. Available from http://www.ncbi.nlm.nih.gov/

[10] NCBI (Homepage on the internet).USA: National Centre for Biotechnology Information. Available from http://www.ncbi.nlm.nih.gov/BLAST.

[11] EBI (Homepage on the internet).UK: European Bioinformatics Institute. Available from http://www.ebi.ac.uk/clustalw/index.html.

[12] Texshade (Homepage on the internet).USA: TeXshade is alignment shading software completely written in TeX/LaTeX; it can process multiple sequence alignments in the. Available from http:// www.ctan.org/tex-archive/

[13] ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in .Available at expasy.org › Tools › Primary structure analysis

[14] Hierarchical Neural Network Actually, even some basic neural nets also have hierarchical structure. For example, the MP model - Perceptron Multi-layer Perceptron BP network. Available at www.cs.iastate.edu/~baojie/acad/current/hnn/hnn.htm

[15] CPHmodels-3.0 Server (Denmark) CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. O. Lund, M. Nielsen, C. Lundegaard, P. Worning . Available at www.cbs.dtu.dk › CBS Prediction Servers.

[16] Rasmol, Version 2.6 Beta - UCB enhanced The new version of RasMol v2.6-ucb is now available for MAC, PC, and UNIX platforms. Available at chemconnections.org/Rasmol/index.html

[17] Swiss Model (Homepage on the internet).Switzerland: Swiss Data Model Database. Available from http://swissmodel.expasy.org//SWISS-MODEL.html.

[18] ProDom (France). ProDom is a comprehensive set of protein domain families automatically generated from the UniProt Knowledge Database. Available at prodom.prabi.fr/

[19] The Pfam protein families database. (USA) Pfam is a large collection of protein multiple sequence alignments and profile hidden Markov models. Pfam is available on the www.ncbi.nlm.nih.gov

[20] SCOP: Structural Classification of Proteins. Structural Classification of Proteins. Structural alignment of SCOP sequences (database + server). Available at scop.berkeley.edu/

[21] CATH: Protein Structure Classification Database, CATH is a manually curetted classification of protein domain structures. Each protein has been chopped into structural domains and assigned into homologous. Available at www.cathdb.

[22] SignalP 3.0 Server (Denmark). SignalP 3.0 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. Available at www.cbs.dtu.dk.

[23] TargetP 1.1 Server (Denmark). TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of Available at www.cbs.dtu.dk

[24] SMART (Germany).You can use SMART in two different modes: normal or genomic. Available at http://smart.embl-heidelberg.de

[25] PSORT WWW Server (Japan). PSORT is mirrored at Tokyo, Okazaki, and Peking. December 1, 1998, Official release of the PSORT II package; June 1, 1999, K. Nakai. Available at psort.hgc.jp

[26] SOSUI WWW Server (Japan). Secondary Structure Prediction of Membrane Proteins. Mitaku Group Department of Applied Physics Nagoya University. Available at bp.nuap.nagoya-u.ac.jp/sosui/

[27] GeneCards V3 - Gene Database. A searchable, integrated database of human genes that provides concise genomic, proteomic, transcriptomic, genetic and functional information on all known. Available at www.genecards.org/.