Structural insights and characterization of Human Npas4 protein

Ammad Fahim 1, Zaira Rehman 1, Muhammad Faraz Bhatti Corresp., 1, Amjad Ali 1, Nasar Virk 1, Amir Rashid 2, Rehan Zafar Paracha Corresp. 3

1 Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan
2 Army Medical College, National University of Medical Sciences, Rawalpindi, Pakistan
3 Research Center for Modeling and Simulation, National University of Sciences & Technology (NUST), Islamabad, Pakistan

Corresponding Authors: Muhammad Faraz Bhatti, Rehan Zafar Paracha
Email address: mfbhatti@asab.nust.edu.pk, rehanzfr@gmail.com

Npas4 is an activity dependent transcription factor which is responsible for gearing the expression of target genes involved in neuro-transmission. Despite the importance of Npas4 in many neuronal diseases, the tertiary structure of Npas4 protein along with its physico-chemical properties is limited. In the current study first we performed the phylogenetic analysis of Npas4 and determined the content of hydrophobic, flexible and order-disorder promoting amino acids. The protein binding regions, posttranslational modifications and crystallization propensity of Npas4 was predicted through different in-silico methods. The three dimensional model of Npas4 was predicted through LOMET, SPARSKS-X, I-Tasser, RaptorX, MUSTER and Pyhre and the best model was selected on the basis of Ramachandran plot, PROSA, Qmean scores. The best model was then subjected to further refinement though MODREFINER. Finally the interacting partners of Npas4 was identified through STRING database. The phylogenetic analysis showed the human Npas4 gene to be closely related to other primates such as chimpanzees, monkey, gibbon. The physiochemical properties of Npas4 showed that it is an intrinsically disorder protein with N-terminal ordered region. The posttranslational modification analyses indicated absence of acetylation and mannosylation sites. Three potential phosphorylation sites (S108, T130 and T136) were found in PAS A domain whilst a single phosphorylation site (S273) was present in PAS B domain. The predicted tertiary structure of Npas4 showed that bHLH domain and PAS domain possess tertiary structures while the rest of the protein exhibited disorder property. Protein-protein interaction analysis revealed NPas4 interaction with various proteins which are mainly involved in nuclear trafficking of proteins to cytoplasm, activity regulated gene transcription and neurodevelopmental disorders. Moreover the analysis also highlighted direct relation to proteins involved in promoting neuronal survival, plasticity and cAMP responsive element binding protein proteins. The current study helps in understanding the physicochemical properties and reveals the neuro-
modulatory role of Npas4 in crucial pathways involved in neuronal survival and neural signalling hemostasis.
1 Structural insights and characterization of Human Npas4 protein

Ammad Fahim, Zaira Rehman, Muhammad Faraz Bhatti, Amjad Ali, NasarVirk, Amir Rashid, Rehan Zafar Paracha

1 Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Sector H-12, Kashmir Highway, Islamabad 44000, Pakistan
2 Research Center for Modeling and Simulation (RCMS), National University of Sciences and Technology (NUST), Sector H-12, Kashmir Highway, Islamabad 44000, Pakistan
3 Department of Biochemistry, National University of Medical Sciences, Rawalpindi, Pakistan

Authors to whom correspondence should be addressed:
Muhammad Faraz Bhatti* Tel. Office: +92 51 9085 6136, Fax # +92 51 9085 6102
1* Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Sector H-12, Kashmir Highway, Islamabad 44000, Pakistan
mfbbatti@asab.nust.edu.pk
Rehan Zafar Paracha*
2* Research Center for Modeling and Simulation (RCMS), National University of Sciences and Technology (NUST), Sector H-12, Kashmir Highway, Islamabad 44000, Pakistan, rehanzfr@gmail.com

1. Abstract:

Npas4 is an activity dependent transcription factor which is responsible for gearing the expression of target genes involved in neuro-transmission. Despite the importance of Npas4 in many neuronal diseases, the tertiary structure of Npas4 protein along with its physico-chemical properties is limited. In the current study first we performed the phylogenetic analysis of Npas4 and determined the content of hydrophobic, flexible and order-disorder promoting amino acids. The protein binding regions, posttranslational modifications and crystallization propensity of Npas4 was
predicted through different in-silico methods. The three dimensional model of Npas4 was predicted through LOMET, SPARSKS-X, I-Tasser, RaptorX, MUSTER and Pyhre and the best model was selected on the basis of Ramachandran plot, PROSA, Qmean scores. The best model was then subjected to further refinement though MODREFINER. Finally the interacting partners of Npas4 was identified through STRING database. The phylogenetic analysis showed the human Npas4 gene to be closely related to other primates such as chimpanzees, monkey, gibbon. The physiochemical properties of Npas4 showed that it is an intrinsically disorder protein with N-terminal ordered region. The posttranslational modification analyses indicated absence of acetylation and mannosylation sites. Three potential phosphorylation sites (S108, T130 and T136) were found in PAS A domain whilst a single phosphorylation site (S273) was present in PAS B domain. The predicted tertiary structure of Npas4 showed that bHLH domain and PAS domain possess tertiary structures while the rest of the protein exhibited disorder property. Protein-protein interaction analysis revealed Npas4 interaction with various proteins which are mainly involved in nuclear trafficking of proteins to cytoplasm, activity regulated gene transcription and neurodevelopmental disorders. Moreover the analysis also highlighted direct relation to proteins involved in promoting neuronal survival, plasticity and cAMP responsive element binding protein proteins. The current study helps in understanding the physicochemical properties and reveals the neuro-modulatory role of Npas4 in crucial pathways involved in neuronal survival and neural signalling hemostasis.

2. **Introduction**

NPas4 belong to the basic helix loop helix/per-arnt-sim (BHLH/PAS) transcription factor family (Shamloo et al. 2006). This family is responsible for gearing the expression of target genes on either positive or negative side (Spiegel et al. 2014). Classically, these transcription factors harbors a DNA binding motif and a dimerisation motif called PAS homology domain (Gu et al. 2000). The word ‘PAS’ pertains to first three proteins of the domain; period (per), aryl hydrocarbon receptor translocator (arnt) and single minded (SIM)(Ooe et al. 2009). Functionally, Npas4 is categorized as an immediate early gene (IEG), responsible for direct control of large number of activity dependent genes that are capable of altering their expression input on the basis of sensory stimuli they receive (Madabhushi et al. 2015; Sun & Lin 2016). Npas4 is also known for development of glutaminergic and GABAergic synapses in neurons suggesting its crucial role in neuro-circuitary
homeostasis and memory formation (Spiegel et al. 2014). Deletion of Npas4 may lead to development of several neuronal plasticity disorders and disorganised management of sensory input (Maya-Vetencourt 2013). Previous studies consolidated its functional role in cerebral ischemia, memory developmental disorders and related pathologies like autism spectrum disorders, and neuropsychiatric disorders (Choy et al. 2015a; Choy et al. 2015b; Ebert & Greenberg 2013; Shepard et al. 2017). Limited studies are available which clearly delineate Npas4 role in regulating its discovered function of neuro-protection and hemostasis in synaptic inter connectivity. The tertiary structure of Npas4 protein along with its physico chemical properties is scarcely available. The current study was mainly aimed to provide the structural insight of Npas4 protein and its possible interaction with other proteins.

3. Materials and Methods:

Flow chart of methodology shown in Figure 1.

3.1 Sequence retrieval and homology search:

The Npas4 protein sequence was retrieved from UniProt database in FASTA format and was subsequently utilized for homology search using protein BLAST program at NCBI (Blastp). The amino acid sequence of Npas4 was used for homology search using BLASTP at NCBI. One hundred orthologous sequences of Npas4 were retrieved. Out of these 100 sequences 24 full length sequences with homology >97% were selected for phylogenetic analysis.

3.2 Multiple sequence alignment and phylogenetic analysis:

Multiple sequence alignment of different orthologous sequences were done using ClustalX (version 2.0). All the default paramenters were used for alignment. For Phylogenetic analysis, the tree was constructed by Mega 7 (Kumar et al. 2016) using NJ (Neighbour Joining) method along with1000 bootstraps. The evolutionary distances were computed using the Poisson correction method. All the other parametres were set as default.

3.3 Primary sequence analysis

3.3.1 Protein order- disorder prediction:

In order to check whether Npas4 is intrinsically disorder protein or not, predictions were computed using MobiDB web tool and Meta Disorder Web server (Kozlowski & Bujnicki 2012; Piovesan et al. 2018). MobiDB predicted the disorder region of proteins through the use of six different tools (DisEMBL, ESPritz, GlobPlot, IUPred, Jtronn, VSL2b(Piovesan et al. 2018).
Metadisorder predicts the disorder regions of proteins through the use of 13 different tools and provides the consensus of all these tools (Kozlowski & Bujnicki 2012). In order to predict the type of disorder in Npas4, charge hydropathy plot and Cumulative Distribution Function analysis was performed through PONDR server (Xue et al. 2010).

### 3.3.2 Physicochemical analysis of Npas4:

The Npas4 primary sequence was used to determine the polarity, mutability, accessibility, bulkiness and refractivity by using Protscale Server on ExPasy platform (http://web.expasy.org/protscale/). The proteins with disorder regions have different composition of flexible and hydrophilic amino acids. The composition of amino acids contribute to flexibility and hydrophilicity was determined through Composition Profiler (http://www.cprofiiler.org/). The signal peptide and cleavage sites were predicted on the basis of artificial neural network through SignalP-4.1 (http://www.cbs.dtu.dk/services/SignalP/) and SecretomeP (http://www.cbs.dtu.dk/services/SecretomeP/).

Prediction of acetylation and mannosylation sites of Npas4 were tallied through NetAcet web server (http://www.cbs.dtu.dk/services/NetAcet) and NetCGlyc web server (http://www.cbs.dtu.dk/services/NetCGlyc/) respectively, while glycosylation sites were determined through NetOglyc and NetNglyc web server (http://www.cbs.dtu.dk/services). Netphos 3.1 server (http://www.cbs.dtu.dk/services/NetPhos) determined the phosphorylation sites for each Thr, Ser and Tyr residues with 0.5 cut off threshold value. NetPhosK 1.0 server (http://www.cbs.dtu.dk/services/NetPhosK) predicted kinase-specific phosphorylation sites in human NPAS4. For surface accessibility of predicted phosphorylated sites, NetSurfP program (http://www.cbs.dtu.dk/services/NetSurfP) was utilized. Protparam Server (http://web.expasy.org/protparam/) estimated half life, molecular weight, and amino acid composition of Npas4 protein.

### 3.3.3 Crystallization propensity of NPAS4:

Either fully or partially disorder proteins have little tendency to crystallization. The probability of crystallization of Npas4 was reckoned through PPCpred (http://biomine.cs.vcu.edu/servers/PPCpred/), ParCrys (http://www.compbio.dundee.ac.uk/xtal/cgi-bin/input.pl) and CRYSTALP2.
ParCrys is based on Parzen Window approach that calculate the probability of crystallization based on hydrophobicity, isoelectric point, and frequencies of Ser, Cys, Gly, Phe, Tyr and Met residues (Overton et al. 2008). PPCpred predict the probability of protein crystallization on the basis of four steps of crystallization (production of protein material, purification, crystallization, and diffraction-quality crystallization) and for each step it calculates the probability and then by combining all the probabilities it gives the final score of propensity of protein crystallization (Mizianty & Kurgan 2011).

### 3.3.4 Prediction of protein binding regions in disorder npas4:

It is the property of disorder proteins that they get an ordered conformation upon binding with some globular protein partner. Hence in an attempt to identify these regions ANCHOR web server (http://anchor.enzim.hu/) was used. It identify the regions within disorder segment that cannot fold and form proper stable structure. ANCHOR utilises the IUPred method for prediction (Meszaros et al. 2009).

### 3.4 Secondary structure prediction:

The secondary structure of Npas4 (Accession No: Q8IUM7) was predicted through GOR4 (Garnier et al. 1996), PsiPred (Jones 1999) and JPred3 (Cole et al. 2008).

### 3.5 Tertiary structure prediction:

Npas4 (Uniprot ID: Q8IUM7) was subjected to BLAST search and less than 25% homology with crystal structure of NPAS3-ARNT complex (5SY7) and NPAS1-ARNT complex (5SY5) was found (figure S1). Due to absence of suitable structural homologue, homology modeling cannot be performed. In the absence of structural homologue, threading protein structure prediction approach was used. Threading is a fold recognition method to predict 3D structure of proteins. I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER), LOMETS (https://zhanglab.ccmb.med.umich.edu/LOMETS/), MUSTER (https://zhanglab.ccmb.med.umich.edu/MUSTER/), SPARSKS-X (http://sparks-lab.org/yueyang/server/SPARKS-X/) RaptorX (http://raptorx.uchicago.edu/) and Phyre 2.0 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) was used for three dimensional structure prediction of human Npas4. LOMETS uses local meta threading approach to predict the tertiary structure of protein. It has nine locally installed threading programs and generates an output on the basis of high scoring target-template alignment. I-TASEER is iterative threading program
that build 3D structure by using hierarchical method (Roy et al. 2010). MUSTER is multisource threading program that works by identifying structural templates from protein data bank and then perform the profile-profile alignment and generates 3D structure (Wu & Zhang 2008). SPARSKS-X is fold recognition method to predict 3D structure (Yang et al. 2011). Phyre works by searching of the structural template for the query sequence followed by multiple sequence alignment and then search the hidden markov model of structures to find the best structure (Kelley et al. 2015).

3.6 Structure validation of Npas4:
Validation of three dimensional structure of NPas4 was performed using PROSA (https://prosa.services.came.sbg.ac.at/prosa.php), Qmean (http://swissmodel.expasy.org/qmean/cgi/index.cgi) and Ramachandran plot (http://mordred.bioc.cam.ac.uk/_rapper/rampage.php). Ramachandran plot is the visualization of \( \psi \) and \( \phi \) torsional angles of protein residues and calculate the residues present in favoured, allowed and outlier region of protein (Lovell et al. 2003). PROSA was used to validate the protein models obtained through different approaches like NMR, X-ray and theoretical calculations (Wiederstein & Sippl 2007). Qmean provided the estimation of quality of model through Q-score (Benkert et al. 2008).

3.7 Structure refinement:
The validated models were then subjected to refinement using Modrefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner/). Modrefiner is an atomic, high resolution protein refinement algorithm (Xu & Zhang 2011).

3.8 Protein-protein interaction analysis:
String database (https://string-db.org/) was used to predict the interaction of Npas4 with other cellular peoteins. STRING is a database of known and predicted protein-protein interactions. In order to get the interacting partners of Npas4 the medium confidence interval value of 0.4 was used and 20 maximum interactions for first shell and 10 maximum interactions for the second shell were used. The interactions may be direct physical interactions or indirect functional associations (Szklarczyk et al. 2017).

4. Results:
4.1 Multiple sequence alignment and phylogenetic analysis:

Multiple sequence alignment of human Npas4 sequence with others species is shown in Figure S2. The phylogenetic analysis of the aligned sequences of Npas suggests Npas4’s evolutionary conservativeness. The analysis of taxonomic classification of all organisms (Supplementary Table S1) revealed human Npas4 to be closely related to other primates including chimpanzees (T.N), monkey, gibbon). The next closest neighbors were rodents (marmart, elephant-shrew, mole, guinea pig). Other related organisms to rodents were artiodactyla like camel and alpaca. The farthest neighbors of Npas4 sequences were from rodents include rabbit and pika (Figure 2).

4.2 Disorder prediction of Npas4:

The disorder prediction was carried out using Metadisorder and Mobidb. According to Metadisorder and Mobidb results indicated Npas4 is an intrinsically disordered protein. According to Metadisorder plot C-terminal half of the protein showed strong disorder while N-terminal of the protein is ordered (position 15-350 a.a). CDF (cumulative distribution function) analysis showed that NPAS4 is mixture of ordered and disordered regions as it intersects the boundary (Figure 3). Primary sequence analysis showed that NPAS4 is enriched in disorder promoting amino acids (Pro, Ser, Glu, Gln, Ala and Gly; Table 1). These amino acids have the tendency to prevent the folding of protein [67–70]. Pro and Ser has the highest probability Amino acid sequence analysis of Npas4 protein showed that almost 50% of the protein composed of disorder region. Flexibility analysis of Npas4 was performed using composition profiler and DynaMine server and both servers showed that Npas4 is enriched with flexible amino acids like Pro, Ser and Gln (Figure 4A). Low hydrophobicity and high hydrophilicity promotes disorder in the proteins(Dyson & Wright 2016).The analysis of according to composition profiler server for suggest NPAS4 to be enriched with hydrophilic amino acids like Gln, Ser, Pro and Thr (Figure 4B).

4.3 Physico-chemical properties of Npas4:

The physico-chemical properties were determined by Protscale server (Figure 5). The higher score suggested a higher probability of that particular property of Npas4. The ratio of the side chain volume to the length of an amino acid suggested protein bulkiness and may affect the local structure of a protein. The bulkiness values of Npas4 (Figure 5A) ranges from 9.713 (position 731aa) to 18.729 (position 295 aa). Hydrophobicity Hopp and wood score (Figure. 5B) and the values were between -0.711 (position 630aa) and 0.422 (position 44 aa). The dipole-dipole intermolecular interactions between the positively and negatively charged particles depicted
polarity as predicted through Zimmerman score (Figure. 5C). The predicted score lies between 0.272 (position 81 and 82aa) and 34.216 (position 630 aa). This data showed that Npas4 possesses polarity. Mutability which is the probability that amino acid will bring a change over a particular evolutionary period of time, was determined through the relative mutability score (Figure. 5D) and the values lies between 46.000 (position 295aa) and 102.889 (position 18 aa) which showed that Npas4 have mutability potential.

The signal peptide of Npas4 was computed using SignalP Server. The presence of signal peptide was measured through C- (raw cleavage site score), S- (signal peptide score) and Y-score (combined cleavage site score). S-score is the estimation of possible signal peptide while D-score is the average of mean S and the max Y-score and it discriminate signal peptide from non-signal peptides. In Npas4, the D-score was 0.450 which was less than the cut off value of 0.5 which showed absence of signal peptide (Figure 6). The secretory nature of Npas4 was predicted through SecretomeP and the NN score which was 0.5.

4.4 Crystallization propensity of Npas4:

Crystallization propensity was predicted through multiple resources in order to get a validated prediction. The ParCrys server predicted Npas4 to be ‘Recalcitrant to Crystallisation’. According to CRYSTALP2, Npas4 is non-crystallizable with 0.447 confidence. The probability of crystallization was also studied through FDETECT and it gives score of 0.64 which mean this protein is difficult to crystallize. PPCpred provided the crystallization probability score of 0.122 for Npas4 while the score above 0.4 means that protein has the ability to crystallize (Mizianty & Kurgan 2011), suggesting similar difficulty in crystallization.

4.5 Protein binding regions in Npas4:

Many disorder proteins bind to some other proteins and transformed from disorder to order and thus perform their function (Dyson & Wright 2016). ANCHOR predicts ten binding sites in the disorder region of Npas4. Out of these ten regions three of them (position 585-599, 662-720, 746-792) were present in transactivation domain of Npas4 (Table 2).

4.6 Post translation modification sites:

Attempts were made to predict the protein modification sites that could be occurring in Npas4. The results indicated absence of acetylation and mannosylation sites. However, two potential N-
glycosylation sites were found in disorder region of Npas4 at 556 (NPTK) and 671 (NLSL) amino acid positions (Figure 7).

A total of 80 O’ linked glycosylation sites were present in Npas4. However, only 56 sites were found available for glycosylation according to the GlyCamserver.

According to Netphos 3.1 server predictions, 34 threonine phosphorylation sites, 53 serine phosphorylation sites and 4 tyrosine specific phosphorylation sites were present in Npas4 (Figure 8). NetsurfP prediction revealed 43 serine, 28 threonine and 3 tyrosine residues were exposed for phosphorylation. Supplementary table S2 showed the 24 Serine, 13 Threonine and 1 Tyrosine kinase specific sites in Npas4. Two phosphorylation sites present (S38, S44) in bHLH domain and three sites (S98, S100 and T136) were observed in PASA domain while 1 site (S273) seen in PAS B domain. Six Ser and four Thr phosphorylation sites present in disordered transactivation domain (Table 3).

Different physiological parameters of Npas4 were predicted through ProtParam server (Table 4).

4.7 Secondary structure of Npas4:
The secondary structure of Npas4 showed that the protein consists of alpha helices, coils and beta sheet (Figure S3).

4.8 Tertiary structure of Npas4:
The three dimensional structure of NPAS4 protein has not been determined to date. Due to the absence of suitable structural template homology modeling can not be used (Figure S1). In order to get the high quality structure of Npas4 ab-initio and threading approaches were used. By using these approaches five models were generated through I-Tasser, ten models, one model from RaptorX, ten models from MUSTER, one model from Phyre and ten models from SPARSK-X (Table 5). All the 37 models were then subjected to validation through PROSA, Qmean and Ramachandran plot. The models with 96% residues in favoured region were further selected for refinement (Table 6 Figure 9). All models suggested that Npas4 is a disordered protein with ordered bHLH and PAS domain. Based on Qmean, PROSA Z-score and ramachandran scores, model 9 generated LOMET was the best predicted three dimensional structure of Npas4. The possible phosphorylation sites in PASA and PASB domain of Npas4 is shown in Figure 10.

4.9 Protein-protein interaction analysis of Npas4:
To study interaction of Npas4 with other proteins, STRING database v10 (Szklarczyk et al. 2015) was used (Figure 11). The results suggested Npas4 interaction with ARNT (Aryl hydrocarbon
receptor nuclear translocator); NXF2B (nuclear RNA export factor 2B); NXF2 (nuclear export factor 2); NXF3 (nuclear export factor 3); NDNL2 (nechin-like 2); MAGEF1 (melanoma antigen family F, 1); C3orf58 (chromosome 3 open reading frame 58); NXF1 (nuclear RNA export factor 1); NUP107 (nucleoporin 107KD); CSAG1 (Chondrosarcoma associated gene 1); MAGEF1 (Melanoma antigen family F, 1); SLC9A6 (Solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6); MAGED2 (Melanoma antigen family D, 2); SLC9A9 (Solute carrier family 9, subfamily A (NHE9, cation proton antiporter 9), member 9); MAGEL2 (MAGE-like 2); SSX2B (Synovial sarcoma, X breakpoint 2B); SLC9A1 (Solute carrier family 9, subfamily A (NHE1, cation proton antiporter 1), member 1); CREB1 (cAMP responsive element binding protein 1); MAGEA1 (Melanoma antigen family A, 1); & EPAS1 (Endothelial PAS domain-containing protein 1 (also known as hypoxia-inducible factor-2alpha (HIF-2alpha)).

5. Discussion:

The Npas4 gene has been in debate for more than a decade, however its structural information is scarcely available. Npas4 expression is primarily traced to neural cells, where it has been reported to be an important contributor of dendritic growth in phases of neuronal development and for modulating limbic patterning and function (Moser et al. 2004). Recently, it has been reported to be also expressed in pancreatic cells (Speckmann et al. 2016). Functionally, Npas4 is an immediate early gene, which under activation, triggers the activation of battery of genes involved in regulating brain plasticity and cognition, attributed mainly to its interaction with number of transcription factors. Npas4 belongs to bHLH-PAS family of transcription factors and the other Npas members (1 and 3) are also linked to numerous psychiatric disorders namely autism, bipolar disorders, schizophrenia and depressive disorders (Adachi 2014; Kamnasaran et al. 2003). A simulated dimerised structure of Npas4 with ARNT was reported before to discuss the potential implication of gene variants (Bersten et al. 2014). Current findings addresses physicochemical properties on human Npas4 protein, with a plausible 3D model, providing useful information about amino acid characteristics and possible identification of interactable proteins.

Human Npas4 is located on human chromosome 11 reference genomic contig NC000011.10, mapping to the chromosomal position 11q13.2. It possess 11 exons that encodes 2406 bp mRNA which translates into 802 amino acid long protein with 87.1 KDa molecular weight (https://www.ncbi.nlm.nih.gov/gene/266743).
In the current study the functional characterization of Npas4 has been done using *in silico* approaches. The MSA and phylogentic analysis showed that Npas4 is evolutionry conserved, with reflecting highest homology to primates as Chimpanzee and Monkey. Npas4 does not have any signal peptide hence it may not being classically secreted extracellularly. The NN score of SecretomeP also verifies Npas4 to be a non-classically secreted protein. The physicochemical and functional properties of proteins are affected by posttranslational modifications. Phosphorylation, acetylation and glycosylation are the most common type of posttranslational modifications of proteins (Koh et al. 2012). Acetylation is one of the major post-translational modifications (PTMs) and is important in determining the cellular localization of proteins (Qin et al. 2011). Theres is no acetylation site present in Npas4 protein. Glycosylation, the addition of glycosyl moiety to protein, is the most common post-translational modification in eukaryotes. N-linked and O-linked glycosylations are common while C-linked (mannosylation) is rare (Koh et al. 2012). No C-linked glycosylation sites were present in Npas4. In Npas4, 2 N-linked glycosylation sites were present but these sites were not available for glycosylation due to the absence of signal peptide. Phosphorylation is another important feature contributing in PTM. Phosphorylation changes the conformation of protein making it active, deactive or modifying its function (Raghava et al. 2014). In Npas4 there were 90 potential phosphorylation sites which include 34 Threonine, 53 Serine and 4 Tyrosine phosphorylation sites. But not all the sites were available for phosphorylation. There were 24 Serine, 13 Theronine and 1 Tyrosine site available for phosphorylation. The important phosphorylation sites present in PAS A and PAS B domain were found at position 130, 136 and 273 respectively. Further experimental studies are needed to explore the role of these sites in Npas4 function.

Prediction analysis in the current study suggested presence of intrinsically disordered protein residues (IDRs) in Npas4 sequence at the C terminal, that can have profound effect on its functional versatility. Conventionally, proteins at physiological temperature exhibit a particular conformational ensemble based on optimal thermal accessibility of the ensemble to various molecular crowders and interacting proteins (Wright & Dyson 1999). However, intrinsically disordered proteins or intrinsically disordered protein residues (IDRs) in a particular protein
enables the protein to interconvert among various topological conformations (Dyson & Wright 2005). These IDRs therefore, can provide Npas4, ease of flexibility to engage multiple targets. Moreover, IDRs can also aid in interactions by efficient utilization of less residues. This will enhance spontaneous disassociation or displacement in neuronal pathways. Previously it has been reported that structural homogeneity exhibited by IDRs is deficient which results from varied composition of Gly and Pro charged residues (Wright & Dyson 2015), as is the case of Npas4, which may subsequently resulted in decreased protein folding. This can further be implicated to enhance conformational plasticity of Npas4 which may help in behaving differently to various transcription factors and enhancing its functional repertoire. The presence of IDRs C-terminal regulatory domain have also been reported in other Bhlh proteins (Fribourgh & Partch 2017)

Moreover, previous reported behavior of IDRs suggest that the presence of more charged residues enhances electrostatic interactions which augments the propensity of post translational modifications (Hofmann et al. 2012; Mao et al. 2010). This subsequently resulted in altered binding affinities and increased range of structural heterogeneity derived from disorder to order transitions altering protein compactness. The simultaneous consideration of Npas4 as an ‘immediate early gene’, alongwith all the predicted and implied consequences of harboring IDRs in its sequence. The PAS domains carrying multi-ligand binding properties can aid Npas4 to impart its role as a ‘hub protein’ in forming various complexes and key mediator in neuronal signaling.

The presence of IDR within Npas4 sequence can pose difficulty in sculpting protein structure through X-ray crystallography which may be attributable to hindrances caused by the crystal packing forces from promiscuous conformations of disordered regions (Wells et al. 2008). This could be one of the reasons that determination of Npas4 crystallization propensity was reported to be on very low scores from all servers used.

The function of protein can be best predicted if we have insight about the protein tertiary structure. The Npas4 sequence have less than 25% homology with crystal structure of PAS domain of Npas3 an Npas1. In the absence of any structural template, ab-initio modeling approach for Npas4 structure prediction was opted. The modeled structure was composed of alpha helices and B sheets
in the bHLH domain and PAS A and B domains, while the transactivation domain does not possess any tertiary structure due to the presence of IDRs. Due to the presence of proper tertiary structure of PAS domain it involved in heterodimerization with its partner (ARNT). Current study of amino acid properties suggested preponderance of hydrophobic amino acids in both PAS A and B domains. These findings are in accordance with other bHLH proteins like Npas1, Npas3, BMAL1 and CLOCK where PAS domains are harboring internal hydrophobic cavities (Wu et al. 2016).

The Npas4 was also subjected to prediction of interacting proteins and it reflected its interaction with ARNT, which is its main partner for dimerization. The STRING analysis also showed some important interactions. The analysis showed that Npas4 functional interaction can be divided into 4 clusters or groups. Cluster 1 involves proteins which are responsible for forming large protein complex (the SMC5-SMC6 complex). This protein complex is responsible for inducing double stranded DNA breaks (DSB) under neuronal stimulation, which is crucial for the expression downstream promoters such as BDNF (Madabhushi et al. 2015). Through its interaction with CDK5, which has a key role in BDNF induced dendrite development, it implicates the crucial role of Npas4 in memory plasticity through regulating the balance of inhibition in neuronal synapses. This interaction may also be related to Rett Syndrome, a disorder in which activity dependent BDNF transcription is hampered (Liang et al. 2015). Cluster 2 (Figure 8) suggested Npas4 interaction to Nuclear RNA export factor (NXF) proteins, which are RNA binding proteins, with various paralogs reported to be found in dendritic granules (Mamon, Ginanova et. al. 2017), suggesting Npas4’s active involvement in the nuclear export of different mRNA and translational control over different proteins. Cluster 3 proteins reported to have altered expression in autism spectrum disorders (ASD), that may consequently result from knockdown of activity regulated gene transcription mediated by Npas4 (Morrow et al. 2008). Cluster 4 represented interaction of Npas4 interaction with cyclic AMP responsive element binding (CREB) proteins. These proteins under phosphorylated state can alter chromatin structure by histones acetylation which speeds up the RNA polymerase II recruitment, trigerring the gene transcriptional programs involved in synapse development (Cohen & Greenberg 2008; Greer & Greenberg 2008). The CREB alongwith Npas4 is also involved in regulating inhibitory synapses of excitatoty neurons by activating BDNF transcription (Hong et al. 2008; Lin et al. 2008). Interstingly, we found three binding sites in the transactivation domain which falls in the region of IDR. This information provides a plausible rationale that CREB may be interacting with any of these residues from the transactvation domain.
through interaction with transcriptional coactivator, CBP which is known to cause such interaction through IDRs (Dyson & Wright 2016).

6. Conclusion:

Npas4 is an intrinsically disorder protein with ordered bHLH and PAS domain. The amino acid sequence analysis showed that Npas4 have high proportion of flexible and hydrophobic amino acids that promote the disorder properties of proteins. The model 9 predicted through LOMET is the best structure of Npas4 and can be used for further analysis. As this protein is difficult to crystallize so inorder to determine its tertiary structure we can use NMR and other related techniques. It has strong interaction with NXF 2B and NXF 3 proteins which implicates its potential role in cytoplasmic export of proteins from nucleus, thus influencing protein translation. Current study also elucidates Npas4 activation with CREB proteins, suggesting Npas4 reliance on Ca²⁺ dependent kinases (CDKs). The information compiled in this research can serve as useful information for identifying new drug targets, which can modulate synaptic hemostasis in neuropshychiatric and neuro-developmental disorders. Moreover, possible interaction with Nuclear RNA export factor (NXF) protein family role identification needs further elaboration. The results reported by our study are more of structural and theoretical in nature, however, they may help in biophysical studies, NMR and crystallographic studies directed towards Npas4. Moreover, the considerations of results may help in future studies designed to understand binding interactions of Npas4.

Abbreviations:

ARNT (Aryl hydrocarbon receptor nuclear translocator);
NXF2B (nuclear RNA export factor 2B);
NXF2 (nuclear export factor 2);
NXF3 (nuclear export factor 3);
NDNL2 (necdin-like 2);
MAGEF1 (melanoma antigen family F, 1);
C3orf58 (chromosome 3 open reading frame 58);
NXF1 (nuclear RNA export factor 1);
NUP107 (nucleoporin 107KDa);
CSAG1 (Chondrosarcoma associated gene 1);
MAGEF1 (Melanoma antigen family F, 1);
SLC9A6 (Solute carrier family 9, subfamily A NHE6, cation proton antiporter 6), member 6);
MAGED2 (Melanoma antigen family D, 2);
SLC9A9 (Solute carrier family 9, subfamily A NHE9, cation proton antiporter 9), member 9);
MAGEL2 (MAGE-like 2);
SSX2B (Synovial sarcoma, X breakpoint 2B)
SLC9A1 (Solute carrier family 9, subfamily A NHE1, cation proton antiporter 1), member 1);
CREB1 (cAMP responsive element binding protein 1);
MAGEA1 (Melanoma antigen family A, 1) & EPAS1 (Endothelial PAS domain-containing protein 1 (also known as hypoxia-inducible factor-2alpha (HIF-2alpha)).

References:

Adachi N. 2014. New insight in expression, transport, and secretion of brain-derived neurotrophic factor: Implications in brain-related diseases. World Journal of Biological Chemistry 5:409. 10.4331/wjbc.v5.i4.409
Benkert P, Tosatto SC, and Schomburg D. 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins 71:261-277. 10.1002/prot.21715
Bersten DC, Bruning JB, Peet DJ, and Whitelaw ML. 2014. Human variants in the neuronal basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) transcription factor complex NPAS4/ARNT2 disrupt function. PLoS ONE 9:e85768. 10.1371/journal.pone.0085768
Cohen S, and Greenberg ME. 2008. Communication Between the Synapse and the Nucleus in Neuronal Development, Plasticity, and Disease. Annual Review of Cell and Developmental Biology 24:183-209. 10.1146/annurev.cellbio.24.110707.175235
Cole C, Barber JD, and Barton GJ. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Research 36:W197-W201. 10.1093/nar/gkn238
Dyson HJ, and Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197-208. nrm1589 [pii] 10.1038/nrm1589

Dyson HJ, and Wright PE. 2016. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J Biol Chem* 291:6714-6722. 10.1074/jbc.R115.692020 R115.692020 [pii]

Ebert DH, and Greenberg ME. 2013. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* 493:327-337. 10.1038/nature11860

Fribourgh JL, and Partch CL. 2017. Assembly and function of bHLH-PAS complexes. *Proc Natl Acad Sci U S A* 114:5330-5332. 10.1073/pnas.1705408114 1705408114 [pii]

Garnier J, Gibrat JF, and Robson B. 1996. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* 266:540-553. S0076-6879(96)66034-0 [pii]

Greer PL, and Greenberg ME. 2008. From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron* 59:846-860. 10.1016/j.neuron.2008.09.002

Gu YZ, Hogenesch JB, and Bradfield CA. 2000. The PAS superfamily: sensors of environmental and developmental signals. *Annu Rev Pharmacol Toxicol* 40:519-561. 10.1146/annurev.pharmtox.40.1.519

Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, and Schuler B. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* 109:16155-16160. 10.1073/pnas.1207719109 1207719109 [pii]

Hong EJ, McCord AE, and Greenberg ME. 2008. A Biological Function for the Neuronal Activity-Dependent Component of Bdnf Transcription in the Development of Cortical Inhibition. *Neuron* 60:610-624. 10.1016/j.neuron.2008.09.024

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195-202. 10.1006/jmbi.1999.3091

Kamnasaran D, Muir WJ, Ferguson-Smith MA, and Cox DW. 2003. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. *J Med Genet* 40:325-332.

Kelley LA, Mezulis S, Yates CM, Wass MN, and Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845-858. 10.1038/nprot.2015.053 nprot.2015.053 [pii]

Koh GCKW, Porras P, Aranda B, Hermjakob H, and Orchard SE. 2012. Analyzing Protein–Protein Interaction Networks. *Journal of Proteome Research* 11:2014-2031. 10.1021/pr201211w

Kozlowski LP, and Bujnicki JM. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* 13:111. 10.1186/1471-2105-13-111

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33:1870-1874. 10.1093/molbev/msw054
Liang Z, Ye T, Zhou X, Lai KO, Fu AKY, and Ip NY. 2015. Cdk5 Regulates Activity-Dependent Gene Expression and Dendrite Development. *Journal of Neuroscience* 35:15127-15134. 10.1523/jneurosci.1443-15.2015

Lin Y, Bloodgood BL, Hauser JL, Lapan AD, Koon AC, Kim T-K, Hu LS, Malik AN, and Greenberg ME. 2008. Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* 455:1198-1204. 10.1038/nature07319

Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, and Richardson DC. 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50:437-450. 10.1002/prot.10286

Madabhushi R, Gao F, Pfenning Andreas R, Pan L, Yamakawa S, Seo J, Rueda R, Phan TX, Yamakawa H, Pao P-C, Stott Ryan T, Gjoneska E, Nott A, Cho S, Kellis M, and Tsai L-H. 2015. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* 161:1592-1605. 10.1016/j.cell.2015.05.032

Mao AH, Crick SL, Vitalis A, Chicoine CL, and Pappu RV. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* 107:8183-8188. 10.1073/pnas.0911107107

Maya-Vetencourt JF. 2013. Activity-Dependent NPAS4 Expression and the Regulation of Gene Programs Underlying Plasticity in the Central Nervous System. *Neural Plasticity* 2013:1-12. 10.1155/2013/683909

Meszaros B, Simon I, and Dosztanyi Z. 2009. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 5:e1000376. 10.1371/journal.pcbi.1000376

Mizianty MJ, and Kurgan L. 2011. Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics* 27:i24-33. 10.1093/bioinformatics/btr229

Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, Mukaddes NM, Balkhy S, Gascon G, Hashmi A, Al-Saad S, Ware J, Joseph RM, Greenblatt R, Gleason D, Ertelt JA, Apse KA, Bodell A, Partlow JN, Barry B, Yao H, Markianos K, Ferland RJ, Greenberg ME, and Walsh CA. 2008. Identifying Autism Loci and Genes by Tracing Recent Shared Ancestry. *Science* 321:218-223. 10.1126/science.1157657

Moser M, Knoth R, Bode C, and Patterson C. 2004. LE-PAS, a novel Arnt-dependent HLH-PAS protein, is expressed in limbic tissues and transactivates the CNS midline enhancer element. *Molecular Brain Research* 128:141-149. 10.1016/j.molbrainres.2004.06.023

Ooe N, Saito K, and Kaneko H. 2009. Characterization of functional heterodimers partners in brain for a bHLH-PAS factor NXF. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1789:192-197. 10.1016/j.bbagrm.2009.01.003

Overton IM, Padovani G, Girolami MA, and Barton GJ. 2008. ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction. *Bioinformatics* 24:901-907. 10.1093/bioinformatics/btn055

Piovesan D, Tabaro F, Paladini L, Necci M, Micetic I, Camilloni C, Davey N, Dosztanyi Z, Meszaros B, Monzon AM, Parisi G, Schad E, Sormanni P, Tompa P, Vendruscolo M, Vranken WF, and Tosatto SCE. 2018. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res* 46:D471-D476. 10.1093/nar/gkx1071
Qin S, Pang X, and Zhou H-X. 2011. Automated Prediction of Protein Association Rate Constants. *Structure* 19:1744-1751. 10.1016/j.str.2011.10.015

Raghava GPS, Yu C-S, Cheng C-W, Su W-C, Chang K-C, Huang S-W, Hwang J-K, and Lu C-H. 2014. CELLO2GO: A Web Server for Protein subCELlular LOcalization Prediction with Functional Gene Ontology Annotation. *PLoS ONE* 9:e99368. 10.1371/journal.pone.0099368

Roy A, Kucukural A, and Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725-738. 10.1038/nprot.2010.5

Shamloo M, Soriano L, von Schack D, Rickhag M, Chin DJ, Gonzalez-Zulueta M, Gido G, Urf er R, Wieloch T, and Nikolic K. 2006. Npas4, a novel helix-loop-helix PAS domain protein, is regulated in response to cerebral ischemia. *European Journal of Neuroscience* 24:2705-2720. 10.1111/j.1460-9568.2006.05172.x

Shepard R, Heslin K, and Coutellier L. 2017. The transcription factor Npas4 contributes to adolescent development of prefrontal inhibitory circuits, and to cognitive and emotional functions: Implications for neuropsychiatric disorders. *Neurobiology of Disease* 99:36-46. 10.1016/j.nbd.2016.12.012

Speckmann T, Sabatini PV, Nian C, Smith RG, and Lynn FC. 2016. Npas4 Transcription Factor Expression Is Regulated by Calcium Signaling Pathways and Prevents Tacrolimus-induced Cytotoxicity in Pancreatic Beta Cells. *J Biol Chem* 291:2682-2695. 10.1074/jbc.M115.704098

Spiegel I, Mardinly Alan R, Gabel Harrison W, Bazinet Jeremy E, Couch Cameron H, Tzeng Christopher P, Harmin David A, and Greenberg Michael E. 2014. Npas4 Regulates Excitatory-Inhibitory Balance within Neural Circuits through Cell-Type-Specific Gene Programs. *Cell* 157:1216-1229. 10.1016/j.cell.2014.03.058

Sun X, and Lin Y. 2016. Npas4: Linking Neuronal Activity to Memory. *Trends in Neurosciences* 39:264-275. 10.1016/j.tins.2016.02.003

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, and von Mering C. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43:D447-D452. 10.1093/nar/gku1003

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, and von Mering C. 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* 45:D362-D368. 10.1093/nar/gkw937

gkw937 [pii]

Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, and Fersht AR. 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci U S A* 105:5762-5767. 10.1073/pnas.0801353105

0801353105 [pii]

Wiederstein M, and Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407-410. gkm290 [pii]

10.1093/nar/gkm290

Wright PE, and Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321-331. 10.1006/jmbi.1999.3110

10.1006/jmbi.1999.3110
Wright PE, and Dyson HJ. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16:18-29. 10.1038/nrm3920

Wu D, Su X, Potluri N, Kim Y, and Rastinejad F. 2016. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. *Elife* 5. 10.7554/eLife.18790

Wu S, and Zhang Y. 2008. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* 72:547-556. 10.1002/prot.21945

Xu D, and Zhang Y. 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* 101:2525-2534. 10.1016/j.bpj.2011.10.024

Xue B, Dunbrack RL, Williams RW, Dunker AK, and Uversky VN. 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:996-1010. 10.1016/j.bbapap.2010.01.011

Yang Y, Faraggi E, Zhao H, and Zhou Y. 2011. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 27:2076-2082. 10.1093/bioinformatics/btr350

btr350 [pii]
Figure 1

Flowchart used for functional annotation of Npas4.

Sequence retrieval of Npas4 followed by homology search and multiple sequence alignment and phylogenetic analysis. Assessment of physiological properties, posttranslational modifications of Npas4 were predicted using primary sequence. Subsequently, secondary and tertiary structure were predicted through Ab-initio modeling.
Figure 2

The phylogenetic tree of human Npas4 generated by NJ method with 1000 bootstrap value.

The tree showed that human Npas4 closest homology with other primates like Chimpanzee (Pan troglodytes), Monkey (Rhinopithecus bieti) and Gibbons (Nomascus leucogenys) while the farthest orthologs were Rabbit (Oryctolagus cuniculus) and Pika (Ochotona princeps).
Figure 3

Disorder prediction of Npas4.

(A) Metadisorder server predicts the disorder regions in Npas4 primary sequence. The plot shows the prediction of three versions of Metadisorder. The value above 0.5 shows disorder region from position 350 to 802. (B) CDF plot of Npas4 calculated through PONDR server. It predicts either order or disorder proteins. Black line is the boundary, the proteins above the boundary are predicted to be ordered, below the boundary are disordered while the proteins that intersect the boundary is the mixture of order and disorder. According to CDF, Npas4 is mixture of order and disorder.
Figure 4

Amino acid composition of Npas4

(A) Flexibility pattern of amino acids in Npas4 calculated through Composition Profiler. Rigid amino acids are shown in green while flexible amino acids are shown in red. (B) Hydrophobicity of amino acids in Npas4 calculated through Composition Profiler. Hydrophilic amino acids are shown in green color while hydrophobic amino acids are shown in black color.
Figure 5

Prediction of physico-chemical properties of Npas4

X-axis shows amino acid sequence from N- to C-terminal while Y-axis show scores computed by each algorithm. (A) bulkiness; (B) hydrophobicity; (D) polarity; (E) relative mutability.
Figure 6

Prediction of signal peptide of NPAS4 by SignalP server.

X-axis represents amino acid sequence from N- to C- terminal while Y-axis represents scores. Purple line shows the threshold value of 0.5. The proteins having C, S and D scores above threshold showing the probability of having the signal peptide.
Figure 7

N-linked glycosylation sites of NPAS4 predicted through NetNGlyc.

X axis represents amino acid sequence from N- to C- terminal while Y axis represents scores. 0.5 is the threshold, the amino acid positions above threshold level means having potential of glycosylation.
Figure 8

Potential phosphorylation sites of NPAS4 predicted through NetPhos.

X axis represents amino acid sequence from N- to C- terminal while Y axis represents scores. 0.5 is the threshold, the amino acid positions above threshold level means having potential of phosphorylation.
Figure 9

Three dimensional structure of Npas4 predicted through MUSTER and LOMET.

(A) MUSTER-4, (B) MUSTER-5, (C) LOMET-1, (D) LOMET-3, (E) LOMET-4, and (F) LOMET-9. All predicted models show that Npas4 is mixture of order and disorder regions.
Figure 10

Posttranslational modifications in bHLH and PAS domains.

bHLH domain is shown in blue (position 10-52 a.a), PAS A domain is shown in yellow (position 72-135 a.a) and PAS B domain is shown in red (position 216-273 a.a). Residues responsible for phosphorylation are highlighted in green.
Figure 11

Interactions of Npas4 with other proteins.

ARNT (Aryl hydrocarbon receptor nuclear translocator); NXF2B (nuclear RNA export factor 2B); NXF2 (nuclear export factor 2); NXF3 (nuclear export factor 3); NDNL2 (necdin-like 2); MAGEF1 (melanoma antigen family F, 1); C3orf58 (chromosome 3 open reading frame 58); NXF1 (nuclear RNA export factor 1); NUP88 (nucleoporin 88KDa); CSAG1 (Chondrosarcoma associated gene 1); MAGEF1 (Melanoma antigen family F, 1); SLC9A6 (Solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6); MAGED2 (Melanoma antigen family D, 2); SLC9A9 (Solute carrier family 9, subfamily A (NHE9, cation proton antiporter 9), member 9); MAGEL2 (MAGE-like 2); SSX2B (Synovial sarcoma, X breakpoint 2B) SLC9A1 (Solute carrier family 9, subfamily A (NHE1, cation proton antiporter 1), member 1); CREB1 (cAMP responsive element binding protein 1); MAGEA1 (Melanoma antigen family A, 1) & EPAS1 (Endothelial PAS domain-containing protein 1 (also known as hypoxia-inducible factor-2alpha (HIF-2alpha)))
**Table 1** (on next page)

Disorder promoting amino acids of Npas4
| Disorder promoting amino acids | Order promoting amino acids |
|-------------------------------|-------------------------------|
| **Polar/charged amino acids** | **Small amino acids**         | **Hydrophobic or bulky** |
| Pro                           | Ala                           | Leu                          | 13.0% |
| 10.3%                         | 7.8%                          | 4.9%                         |
| Ser                           | Gly                           | Val                          | 3.2% |
| 10.2%                         | 6.6%                          | 2.6%                         |
| Glu                           |                               | Tyr                          | 2.6% |
| 7.2%                          |                               | 2.5%                         |
| Gln                           |                               | Ile                          | 2.5% |
| 5.2%                          |                               | 2.1%                         |
| Arg                           |                               | Asn                          | 1.6% |
| 3.4%                          |                               | Cys                          | 1.6% |
| Lys                           |                               | Trp                          | 0.87% |
| 2.4%                          |                               |                              |      |
Table 2 (on next page)

Protein binding regions in NPAS4
|   | From | To  | Length |
|---|------|-----|--------|
| 1 | 362  | 367 | 6      |
| 2 | 397  | 404 | 8      |
| 3 | 416  | 422 | 7      |
| 4 | 432  | 444 | 13     |
| 5 | 464  | 521 | 58     |
| 6 | 535  | 543 | 9      |
| 7 | 560  | 573 | 14     |
| 8 | 585  | 599 | 15     |
| 9 | 662  | 720 | 59     |
|10 | 746  | 792 | 47     |
Table 3 (on next page)

Phosphorylation sites in disorder region of Npas4.
| S.No | Phosphorylation site | Kinases          |
|------|----------------------|------------------|
| 1    | Ser 611              | P38MAPK, CDK5    |
| 2    | Thr 620              | PKC              |
| 3    | Thr 640              | CDC2             |
| 4    | Ser 656              | CKII             |
| 5    | Thr 673              | CKI              |
| 6    | Ser 719              | P38MAPK, CKII    |
| 7    | Ser 728              | CKII             |
| 8    | Ser 759              | CKII             |
| 9    | Ser 777              | CKII             |
| 10   | Thr 785              | PKC, ATM         |
Table 4(on next page)

Physiological parameters of Npas4.
| Property             | Value        |
|----------------------|--------------|
| Molecular Weight     | 87116.58     |
| Theoretical PI       | 4.53         |
| Formula              | $C_{3894}H_{5965}N_{1005}O_{1220}S_{23}$ |
| Total no. of Atoms   | 12107        |
| Estimated half life  | 30 hours     |
Table 5 (on next page)

Models of Npas4 protein with different methods and their evaluation results.
| Model/Tool | Method               | PROSA | Qmean | Ramachandran plot |
|------------|----------------------|-------|-------|-------------------|
|            |                      |       |       | Outlier | Allowed region | Favored region |
| I-TASSER-1 | Multiple-threading alignments | -3.6 | -11.55 | 9.4%     | 18.6%   | 72.0%   |
| I-TASSER-2 | Multiple-threading alignments | -1.28 | -12.31 | 8.4%     | 17.4%   | 74.4%   |
| I-TASSER-3 | Multiple-threading alignments | -2.59 | -16.61 | 18.5%    | 28.0%   | 53.5%   |
| I-TASSER-4 | Multiple-threading alignments | -4.99 | -16.83 | 19.2%    | 24.6%   | 56.1%   |
| I-TASSER-5 | Multiple-threading alignments | -5.07 | -13.55 | 14.2%    | 27.8%   | 58.0%   |
| LEMOT-1    | Local meta-threading  | -2.89 | -3.34  | 1.6%     | 2.2%    | 96.1%   |
| LEMOT-2    | Local meta-threading  | -3.07 | -3.22  | 1.4%     | 2.9%    | 95.8%   |
| LEMOT-3    | Local meta-threading  | -3.66 | -3.08  | 0.8%     | 2.4%    | 96.9%   |
| LEMOT-4    | Local meta-threading  | -3.72 | -3.66  | 1.6%     | 2.4%    | 96.0%   |
| LEMOT-5    | Local meta-threading  | 15.32 | -10.17 | 10.2%    | 14.2%   | 75.6%   |
| LEMOT-6    | Local meta-threading  | -3.11 | -3.96  | 2.0%     | 5.9%    | 92.1%   |
| LEMOT-7    | Local meta-threading  | -2.61 | -2.93  | 1.6%     | 4.5%    | 93.9%   |
| LEMOT-8    | Local meta-threading  | -3.33 | -2.75  | 1.5%     | 2.9%    | 95.6%   |
| LEMOT-9    | Local meta-threading  | -3.63 | -2.61  | 0.6%     | 3.2%    | 96.1%   |
| LEMOT-10   | Local meta-threading  | 3.09  | -20.65 | 7.5%     | 15.9%   | 76.7%   |
|          | Threading          | -5.42 | -5   | 2.8% | 4.7% | 92.5% |
|----------|--------------------|-------|------|------|------|-------|
| MUSTER-1 | Multi-Source Threading | -1.57 | -12.86 | 4.5% | 8.2% | 87.2% |
| MUSTER-2 | Multi-Source Threading | -1.51 | -20.75 | 6.2% | 12.2% | 81.5% |
| MUSTER-3 | Multi-Source Threading | 1.09  | -2.38 | 7.2% | 11.2% | 81.2% |
| MUSTER-4 | Multi-Source Threading | -3.19 | -3.91 | 1.8% | 1.9% | 96.4% |
| MUSTER-5 | Multi-Source Threading | -3.35 | -2.88 | 1.1% | 2.6% | 96.2% |
| MUSTER-6 | Multi-Source Threading | -2.87 | -20.70 | 4.8% | 10.0% | 85.2% |
| MUSTER-7 | Multi-Source Threading | -0.34 | -14.63 | 4.8% | 9.5% | 85.8% |
| MUSTER-8 | Multi-Source Threading | -2.44 | -4.14 | 2.1% | 3.4% | 94.5% |
| MUSTER-9 | Multi-Source Threading | 0.53  | -9.44 | 7.1% | 12.0% | 80.9% |
| MUSTER-10| Multi-Source Threading | -1.46 | -12.05 | 4.5% | 12.9% | 82.6% |
| Phyre    | Homology detection method | -4.13 | -8.70 | 7.9% | 12.8% | 79.4% |
| SPARSK-1 | Fold Recognition     | -3.9  | -9.59 | 5.5  | 7.8  | 86.7% |
| SPARSK-2 | Fold Recognition     | -4.57 | -6.67 | 2.2  | 3.9  | 93.9% |
| SPARSK-3 | Fold Recognition     | -4.02 | -6.12 | 1.5  | 3.6  | 94.9% |
| SPARSK-4 | Fold Recognition     | -5.58 | -6.05 | 1.6  | 6.1  | 92.2% |
| SPARSK-5 | Fold Recognition     | -5.33 | -7.03 | 1.8  | 4.4  | 93.9% |
| SPARSK-6 | Fold Recognition     | -4.77 | -7.39 | 2.2  | 4.2  | 93.5% |
| SPARSK  | Fold Recognition | -4.05 | -6.53 | 1.4   | 3.6   | 95.0% |
|---------|------------------|-------|-------|-------|-------|-------|
| SPARSK 8 | Fold Recognition  | 0.7   | -11.0 | 6.6   | 7.4   | 86.0% |
| SPARSK 9 | Fold Recognition  | -4.3  | -7.51 | 2.4   | 3.6   | 94.0% |
| SPARSK 10 | Fold Recognition | -4.71 | -6.99 | 1.8   | 5.0   | 93.2% |
Table 6 (on next page)

The validated parameters for refined models of Npas4 Protein.
| Model/Tool | Method                 | PROSA | Qmean | Ramachandran plot |
|-----------|------------------------|-------|-------|-------------------|
|           |                        |       |       | Outlier | Allowed Region | Favored Region |
| MUSTER-4  | Multi-Source Threading | -3.19 | -3.91 | 1.8%    | 1.9%            | 96.4%          |
| MUSTER-5  | Multi-Source Threading | -3.35 | -2.88 | 1.1%    | 2.6%            | 96.2%          |
| LEMOT-1   | Local meta-threading   | -2.89 | -3.34 | 1.6%    | 2.2%            | 96.1%          |
| LEMOT-3   | Local meta-threading   | -3.66 | -3.08 | 0.8%    | 2.4%            | 96.9%          |
| LEMOT-4   | Local meta-threading   | -3.72 | -3.66 | 1.6%    | 2.4%            | 96.0%          |
| LEMOT-9   | Local meta-threading   | -3.63 | -2.61 | 0.6%    | 3.2%            | 96.1%          |