Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0510.v1

Review

The participation of the intrinsically disordered regions of the bHLH-PAS transcription factors in disease development

Marta Kolonko [1], Beata Greb-Markiewicz [1]*

[1] Department of Biochemistry, Molecular Biology and Biotechnology, Faculty of Chemistry, Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

* Correspondence: beata.greb-markiewicz@pwr.edu.pl; Tel.: 0048713206226

Abstract: The bHLH-PAS proteins are a family of transcription factors regulating expression of a wide range of genes involved in different functions, from differentiation and development control, by oxygen and toxins sensing to circadian clock setting. In addition to the well-preserved DNA-binding bHLH and PAS domains, bHLH-PAS proteins contain long intrinsically disordered C-terminal regions, responsible for their activity regulation. Our aim was to analyse the potential connection between disordered regions of the bHLH-PAS transcription factors with posttranscriptional modifications and liquid-liquid phase separation in the context of the disease-associated missense mutations. Highly flexible disordered regions, enriched in short more ordered motives, are responsible for wide spectrum of interactions with transcriptional co-regulators. Based on our *in silico* analysis and taking into account fact that transcription factors functions can be modulated by posttranslational modifications and spontaneous phase separation, we assume that the location of missense mutations inducing disease states, is clearly related to sequences directly undergoing these processes or to sequences responsible for their activity regulation.

Keywords: disease-associated mutation; IDR; intrinsically disordered region; LLPS; phase separation; PTM; Ahr; AhRR; SIM1; SIM2; Hif-2α; NPAS4; ARNT2; BMAL1; disorder prediction; LLPS prediction; cancer; HuVarBase,, catGranule prediction

1. Introduction

The basic helix–loop–helix/Per-ARNT-SIM (bHLH–PAS) proteins are an important class of transcription factors (TFs) responsible for the regulation of developmental and physiological events occurring in mammals [1]. Representants of this family perform a wide spectrum of functions. Starting with Aryl hydrocarbon receptor (AHR) acting as receptor for environmental stimuli including highly toxic dioxins [2] by Clock and Bmal1 regulating circadian rhythms of organism [3] to Hypoxia inducible factor 1α (Hif-1α) [4] being an specific oxygen sensor in cells.HifF-1α translocate in hypoxia conditions from cytoplasm to the nucleus, binds to Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) and induces the expression of regulated genes related to angiogenesis, cell proliferation, glucose and iron metabolism [5]. The incorrect control of these processes is commonly connected with the genesis of many diseases, including cancer, strokes and heart diseases [4].

bHLH-PAS proteins are commonly divided into two classes. Class I proteins form functional heterodimers with constitutively expressed class II family members. In contrast to the latter, expression of class I proteins is specifically regulated by physiological states and/or environmental signals. Class I comprises mammalian AhR, AhRR, SIM1-2, Hif1-3α and NPAS1-4 TFs, while class II

contains general partners: ARNT, ARNT2, and brain and muscle ARNT-like 1 and 2 proteins (BMAL1 and BMAL2) [6,7].

Despite mediating highly differentiated signalling pathways, bHLH-PAS proteins present conserved domain structure. The bHLH domain, typically located at the N-terminus of protein, is responsible for DNA binding and dimerization [8]. Following PAS domain comprises two structurally conserved regions: PAS1 and PAS2, separated by a poorly conserved link [9][1]. While the PAS1 region is linked to the selection of dimerization partner and specificity of target genes activation [10], PAS2 region is responsible for ligands/cofactors binding. Each binding case may affect proteins conformation and thus, its activity [10]. In contrast to defined domains, the C-termini of bHLH–PAS proteins present significant primary structure variability and are considered as highly important and unique parts of the proteins responsible for the specific modulation of the bHLH–PAS protein action [10]. They usually comprise specific regions responsible for protein-protein interaction known as transcription activation/repression domains (TAD/RPD) [11][12]. Importantly, most of the bHLH-PAS proteins C-termini were predicted as intrinsically disordered regions (IDRs) [13].

IDRs and intrinsically disordered proteins (IDPs) do not possess a stable tertiary structure in physiological conditions [14] contradicting the fundamental paradigm of biochemistry and structural biology that the function of a protein, results directly from its stable tertiary structure [15]. Currently, more than 20-30% of eukaryotic proteins have been found to exhibit features of IDPs, and over 70% of proteins involved in signal transduction cascades have long IDRs. IDPs were identified as important elements of the cell cycle, cell differentiation, regulation of transcription, mRNA processing and apoptosis control [16][17][18].

The lack of a defined structure is critical for the IDPs and IDRs functionality [17]. The conformational plasticity allows IDPs to interact with several proteins/ligands which is highly advantageous in the molecular recognition processes [19]. From this reason IDPs are commonly involved in one-to-many and in many-to-one binding and can function as hub proteins responsible for the cross-talk of different pathways [20]. The elongated conformation and low compactness make IDPs excellent targets for post-translational modifications (PTMs) and proteolytic degradation, which are typical ways of proteins activity regulation [21]. Recently, additional role of IDR in formation of self-assembled, membrane-less organelles through liquid-liquid phase separation (LLPS) was presented. Interestingly, protein-protein interaction (PPI) could lead to LLPS formation, which enable the partition of specific functions, but in contrast LLPS may also prevent from protein interactions [22,23][24]. In context of transcription factors very interesting is the putative role of LLPS in fast cellular responses to external stimuli [25]. The ability of protein to LLPS formation may be regulated by a wide spectrum of mentioned PTMs and alternative splicing [26].

Proteome-wide analyses of disease-related mutations have shown that gain or loss of post-translational modification sites, which are generally found in IDRs, contributes to human diseases. Moreover, IDRs are enriched in proteins considered as participating in human diseases, for instance, 80% of human cancer-associated proteins containing IDRs [27]. We were interested in answering the question if there is a connection between missense mutation localization and diseases development in the case of bHLH-PAS proteins family. That's why we performed prediction of IDRs presence and LLPS propensity simultaneously with analyzing human polymorphism and posttranslational modification database. We found that most of disease-associated missense mutations are located in IDRs of analyzed bHLH-PAS family representants.

2. bHLH-PAS proteins and diseases.

2.1 AhR and AhRR

AhR, best known as a mediator of environmental pollutants toxicity, contributes additionally to the proper functioning of the liver, cardiovascular, immune, and reproductive systems [28]. AhR was also connected with normal skin formation during fetal development and with pathological states such as epidermal wound healing and skin carcinogenesis as well [29]. Recently, AhR has been recognized as an important modulator of diseases driven by immune/inflammatory processes [30]. The ligand bound AhR translocate to the nucleus and mediates biological response to toxins resulting in wasting syndrome, hepatotoxicity, teratogenesis, and tumour promotion [2]. Activation of AhR was linked to chronic kidney and cardiovascular diseases [28]. The overexpression and constitutive AhR activation have been assigned to various types of tumours [31], i.e. brain tumors: gliomas, meningiomas, medulloblastomas, and neuroblastomas [32]. AhR activation is linked with renal damage, diabetic nephropathy and urinary system-associated cancers [33]. AhR as heterodimer with ARNT functions as co-regulator of estrogen signaling mediated by estrogen receptor (ER) [34] and is considered as responsible for connecting inflammation process with breast cancer [35].

Interestingly, AhR self-regulates its activity by activation of the repressor, Aryl hydrocarbon Receptor Repressor (AhRR). In comparison to AhR, AhRR presents higher tissue-specificity. The highest concentration of this protein was observed in testis, lung, spleen, heart, and kidney [36]. The repressor competes with AhR for the binding with ARNT and forms an inactive heterodimer AhRR/ARNT [35]. AhRR is not able to bind AhR ligands because it does not possess the PASB domain in the N-terminal region. Additionally AhRR contains the C-terminal trans-repression domain (instead of transactivation domains in AhR C-terminus), which allows to bind corepressors involved in a negative regulatory loop [37]. Zudaire with colleagues [38] demonstrated downregulation of AhRR expression in human malignant tissue isolated from different anatomical origins like colon, breast, lung, stomach, cervix, and ovary. Also, genetic polymorphisms of AhRR was related to susceptibility to advanced endometriosis [39][40]. Interestingly, it was observed that AhRR splice variant is able to inhibit transcription activated by Hif-1 what is vital for cancer progression [41].

2.2 Single minded protein (SIM)

The mammalian SIM exists in two isoforms: SIM1 and SIM2, with a high amino acid identity in their N-terminal regions (90% identity in the bHLH and PAS domains) and high diversity in their C-termini [42]. While SIM1 is responsible for the activation of specific genes expression, SIM2 is defined as an inhibitor. The opposite transcriptional effect results from the presence of two repression domains within the SIM2 C-terminal sequence [43][44]. These confirms the responsibility of the C-terminal regions for the bHLH–PAS proteins functions and activity [10]. SIM1 dimerizes with ARNT and activates specific genes related to the development, terminal differentiation, and post-development functioning of neuronal cells, especially in the paraventricular nucleus of the hypothalamus (PVN) [45]. Importantly, PVN is responsible for several autonomic processes, including response to stress, metabolism control, growth, reproduction and appetite regulation [45]. Since the SIM1 plays role in the long-term regulation of food intake and energy expenditure [46], its reduced activity manifests phenotypically with profound obesity and increased linear growth. The weight gain is connected high food consumption, since measured energy expenditure is usual [46]. It was shown, that SIM1 haploinsufficiency in mice induces hyperphagia (abnormally increased appetite for consumption of food) [47] leading to the obesity and developmental abnormalities of the brain [48]. As shown, transgenic mice with overexpressed SIM1 are resistant to the diet-induced obesity, what supports a post developmental, physiologic role for SIM1 in feeding regulation [49]. Contrary, induced SIM1 overexpression contributes to a decreased food intake [50].

2.3 Hypoxia inducible factor 2α (Hif-2α)

Functional Hypoxia-inducible factors are heterodimers comprising one of the three known α-subunits regulated by oxygen (HifF-1α, Hif-2α and HifF-3α), and constitutively expressed ARNT (known also as Hif-1β) [51]. Hif-2α, named also as endothelial PAS-1 (EPAS1) was isolated for the first time from endothelial cells [52]. Interestingly, some contrary results concerning tissue specificity of Hif-2α were reported [53]. Hif-2α shares approximately 50% sequence identity with ubiquitously expressed Hif-1α and both proteins are regulated by oxygen level. Under normal level, two proline residues in the oxygen-dependent degradation domain located in C-termini of Hif-1α/Hif-2α are hydroxylated and targeted to the ubiquitin–proteasome (26S) pathway for degradation, also arginine residues is hydroxylated to prevent interaction with coactivator protein p300 [54]. Hif-2α, similarly to HifF-1α, was shown to induce the expression of genes stimulating cell cycle progression, proliferation, apoptosis promotion, autophagy and angiogenesis [51]. Hif-2α regulates erythropoietin level. It is also involved in embryonic development and metastasis [55][56]. Interestingly Hif-2α presents punctate localization within the nucleus in contrast to homogenously in nucleus distributed Hif-1α. Distinct subnuclear localization of the alpha subunits was proposed to contribute to the differences in the regulation and activity of these two TFs [57]. It was shown that Hif-2α shuttling is regulated by phosphorylation [58]. Some studies of kidney cancer suggested oncogenic role for Hif-2α, in contrast to Hif-1α which manifested tumor suppressor properties [59]. Missense mutations within the bHLH and PAS domains Hif-1α/Hif-2α proteins have linked to cancers including stomach adenocarcinomas, endometrial carcinomas, brain gliomas, lung adenocarcinomas, hepatocellular carcinomas and skin melanomas [54]. The conserved between all known Hif-2α proteins mutation of Gly537, located close to the primary oxygenation site, results in familial erythrocytosis characterized by an increased number of red blood cells. The familial erythrocytosis symptoms are headaches, dizziness, nosebleeds, and shortness of breath. Additionally, the red blood cells excess increase the risk of developing abnormal blood clots [60].

2.4 Neuronal PAS-domain containing protein 4 (NPAS4)

Initially, it was shown that the expression and activity of NPAS4 protein occurs mainly in the nervous system [61]. However, the following studies have shown that NPAS4 is also expressed in β cells of pancreatic islets affecting significantly pancreatic cells. In this case, NPAS4 expression is induced by endoplasmic reticulum stressors and prevent the death of β cells [62,63]. In the nervous system, NPAS4 is responsible for the regulation of the development of GABAergic inhibitory neurons [64]. NPAS4 was shown to be able to inhibit seizure attacks in pilocarpine-induced epileptic rats [65]. Importantly, increased level of NPAS4 expression has been linked to brain protection in focal and generalized ischemic strokes of the brain, where it prevents necrosis and leads to cell apoptosis [66,67]. It was additionally demonstrated that NPAS4 is involved in the structural plasticity of the nervous system and plays an important role in the formation of long-term memory. Its expression is also induced during the learning process [68][69]. Interestingly, NPAS4 over-expression can reverse the tau protein aggregation [70]. Finally, NPAS4 expression was also detected in endothelial cells, where, similarly to pancreatic beta-cells, NPAS4 promoted pro-angiogenic cell functions such as migration or sprout formation [71]. Interestingly, for human NPAS4 was proposed isoform 2 of NPAS4 comprising aa 1-234 (only bHLH and PAS-1 domains) with V234G substitution, however without a proof at the protein translation level and with no known function [72]. To date, only few dimerization partners for NPAS4 were identified: ARNT and ARNT2 being general partners for the class I bHLH-PAS TFs in the brain [73] and the melanoma-associated antigen D1 (MAGED1), expressed ubiquitously in both developing and adult tissues but particularly abundant in the brain. MAGED1 functions in various pathways, like neuronal precursor apoptosis, differentiation, in stabilizing periodicity in the circadian rhythm and in learning and memory formation [74]. It was shown, that developmental down-regulation of NPAS4 in the frontal cortex caused behavioural abnormalities observed in the case of

 doi:10.20944/preprints202010.0510.v1

neurodevelopmental disorders like schizophrenia or autism [75]. NPAS4 was also linked to a number of other serious psychiatric disorders like depression, Huntington's disease, Down syndrome, neurodegenerative diseases (e.g. Alzheimer's disease) [76].

2.5 Aryl hydrocarbon receptor nuclear translocator 2 (ARNT2) and BMAL1

ARNT2 is the representant of the second class of the bHLH-PAS TFs. It is constitutively expressed and acts as general heterodimerization partner for multiple class I members, including SIM1 [77] and NPAS4 [78,79]. In contrast to ARNT expressed equally in all tissues and interacting with wide spectrum of physiological partners (ARNT is indispensable for AHR and Hif signalling) [80], ARNT2 is highly expressed in brain (in the developing CNS), kidney, urinary tract and embryos [81][82]. ARNT2 deficiency leads to the secondary microcephaly within the first few months of human life with a specific frontal and temporal lobe hypoplasia [83]. Secondary microcephaly indicates a progressive neurodegenerative condition caused by decreased number of dendritic connections and/or neurons activity [84]. The hypothalamic insufficiency can cause obesity, diabetes and combined pituitary hormone deficiency [83]. The latter seems to be consistent with a key role for ARNT2 in the development of specific neurosecretory neurons in the human hypothalamus [83]. Some of Arnt2 mutants are also considered as causing hyperphagic obesity, diabetes and hepatic steatosis [85]. ARNT2 was shown also as an important component of protein complex located at a node of TFs network controlling glioblastoma cell aggressiveness [86].

BMAL1 together with its heterodimerization partner CLOCK creates the core of regulatory mechanism of mammalian circadian rhythms. The BMAL1 transactivation domain (TAD) located in C-terminus acts as a regulatory hub that interacts with positive/negative transcriptional regulators as a function of circadian time to control the activation state of CLOCK-BMAL1 dimer [87]. The conformational switch of TAD is caused by cis/trans isomerization about a highly conserved W624-P625 imide bond [88]. BMAL1 polySUMOylation leads to ubiquitination and binding of CREB-binding protein (CBP) in discrete nuclear foci which potentiates its transcriptional activity. Formation of nuclear bodies containing BMAL1/CBP provides transcriptionally active sites of target genes [89] and support our thesis about putative role of BMAL1 in LLPS formation. Similarly to other bHLH-PAS TFs Bmal 1 is shuttling protein [90] and phosphorylation regulates localization signal activity [91]. Interestingly, BMAL1 was shown to stimulate also translation by interactions with the translational machinery in the cytosol after S42 phosphorylation [92]. Geyfman with co-workers [93] reported that BMAL1-dependent circadian variation in DNA sensitivity to UVB induced damage, pointing connection of circadian mechanisms with epidermal carcinogenesis.

3. Most of missense mutations associated with disease are located in IDRs of bHLH-PAS proteins.

To date, the structural characterization of bHLH-PAS TFs was limited to the bHLH-PAS regions, excluding the C-terminal region. It can be explained by the difficulties associated with the full-length proteins expression and purification caused by their long, disordered C-termini. We discussed this research area in details previously [13]. Previously documented analysis of the location of the missense mutations linked with cancers was limited to bHLH-PAS domains of selected bHLH-PAS members (HifF-1α and Hif-2α) [54].

Taking into account the relationship of bHLH-PAS TFs with some serious disorders presented in previous chapter, we asked the question about localization of known missense mutations associated with diseases on the entire proteins' length, also in regions enriched in IDRs. For this purpose, we have reviewed the literature and analyzed Human Variants Database (HuVarBase) https://www.iitm.ac.in/bioinfo/huvarbase/mas18srch.php [94]. HuVarBase, is a comprehensive database on human genome variants reported in the databases: Humsavar (Human polymorphisms and disease mutations), 1000 Genomes (genetic variants occurring at least in 1% of studied populations), SwissVar (portal to search variants in Swiss-Prot entries of the UniProt

Knowledgebase), ClinVar (aggregates information about genomic variation and its relationship to human health) and COSMIC (the Catalogue Of Somatic Mutations In Cancer). We have selected AhR, AhRR, SIM1, SIM2, Hif-2α , ARNT and BMAL1 as our analysis subjects.

Next, we performed *in silico* analysis using the predictors of intrinsically disordered regions: PONDR-VLXT [95], PONDR-VLS2 [96] and IUPred [97] and computational analyses of the putative propensity to undergo LLPS using catGranule program, (http://service.tartaglialab.com/update_submission/216885/dd56e32a89) [98]. Because results of protein disorder prediction were consistent for all used predictors, we present only results obtained for PONDR-VLS2, for simplicity sake. The range of the propensity score is not determined precisely in catGranule prediction so we used as control performed previously prediction for nucleophosmin and estrogen receptor, which were experimentally documented as undergoing LLPS [99].

We used PhophoSitePlus database (https://www.phosphosite.org/homeAction) to view known posttranlational modifications sites [100]. We have also performed prediction of multiple types of PTM site including phosphorylation, lysine acetylation, ubiquitination, sumoylation, methylation, O-GalNAc, O-GlcNAc, sulfation and proteolytic cleavage using PTM-ssMP web server http://bioinformatics.ustc.edu.cn/PTM-ssMP/index/ [101]. Results of our analysis we present in detail for each selected protein.

3.1 AhR and AhRR

According to PONDR server prediction, most of the documented (Fig. 1A a) and predicted sites underlying PTMs (Fig. 1A b) are located in disordered regions of AhR, which were noticeable at the short N-terminal fragment preceding bHLH domain (1-26 aa), the linker between PAS1 and PAS2 domains (182-274 aa) and the long C-terminus of the protein (387-848 aa) (Fig. 1A c). In contrast, the regions comprising preserved domains are mainly ordered, what is common characteristic for bHLH-PAS proteins. The missense mutations in IDRs are linked mainly to large intestine cancer (T199P, P260L, N505S, T507I, P838S), soft tissue cancer (R554K), thyroid cancer (V570I), kidney cancer (E488K), and liver cancer (P18L) (Supplementary Materials). The proximity of missense mutation location (Fig. 1A b) and PTMs seems to be crucial for diseases development. Prediction of LLPS propensity results in positive maximal score in the C-terminal fragment (500-600aa) (Fig. 4A d) in the area enriched in disease associated mutations (see Fig. 1A b). The additional local positive maximum is observed in the predicted as locally disordered linker between bHLH and PAS domain.

In the case of AhRR, all documented (Fig. 1B a) and predicted sites underlying PTMs (Fig. 1B b) localize in the IDRs. What important, AhRR undergo many rather not common modifications like SUMOylation (Fig. 1B a, black dots). AhRR as repressor lacks ligand binding PAS2 domain and is highly disordered not only at the N-and C-termini (1-27 aa and 183-700 aa) but also in the linker between bHLH and PAS domains (82-111 aa) (Fig. 1B c). AhRR presents defined ordered structure only in the middle of of bHLH domain and in the entire PAS domain. LLPS propensity score shows positive maximum for the central part of protein (approximately 340-440aa) (Fig. 4A d) surrounded by PTM sites and coincides with the area of disordered C-terminus. We can observe that entire AhRR C-terminus is rich in the disease-associated mutations in contrast to the bHLH and PAS domains. Diseases linked with mutations are represented mainly by intestine cancer (I226V, R230C, R285W, A300T, T419I, R485W, R491W, G494S, V553M, D645H,), skin cancer (P283S, A301V, G427E), prostate cancer (R491Q, D645H) and liver cancer (C545F, A674S). The other single mutations are connected to endometrium cancer (A371T), CNS cancer (P189A) and esophagus cancer (G612S) (Supplementary Materials).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0510.v1

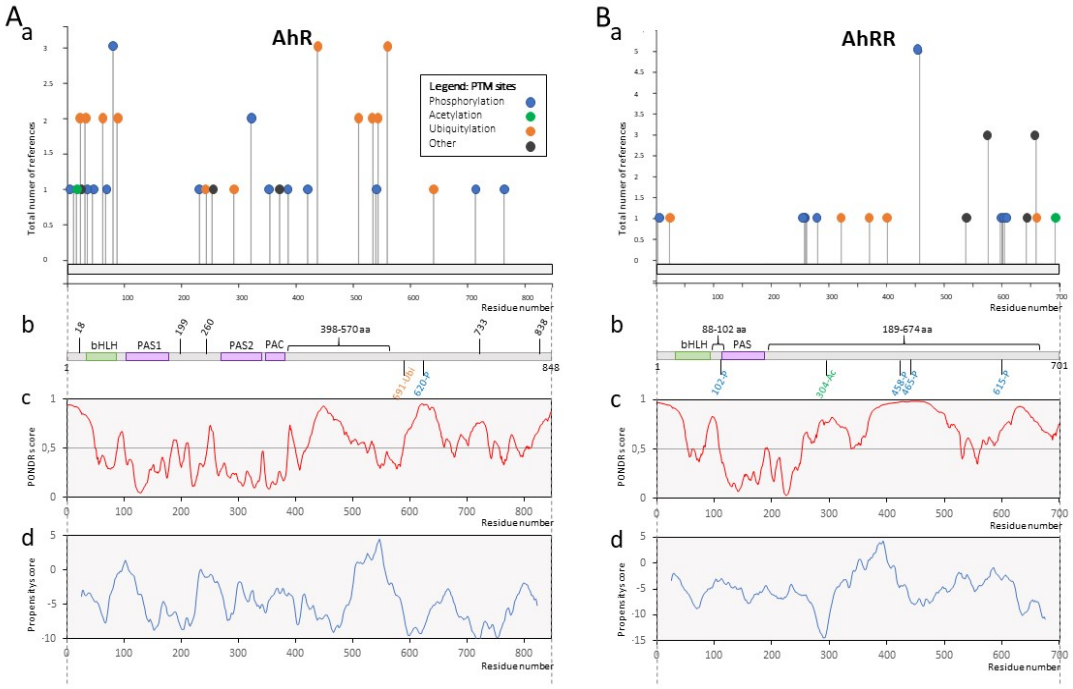

Figure 1. Schematic presentation of results for AhR (P35869) (A) and AhRR (A9YTQ3) (B) analysis.
(a) Post-translation modifications based on PhosphoSitePlus server [100], **(b)** the domain structure of protein, green indicates the bHLH domain (27-80aa AhR; 28-81aa AhRR), whereas purple represents PAS domains (111-181aa PAS1, 275-342aa PAS2 348-386aa PAC; 112-182aa PAS AhRR). Disease-associated missense mutations localized in disordered regions and annotated in HuVarBase database (62) are marked above the diagram. Post-translation modifications predicted with PTM-ssMP server [101] are marked below the diagram (blue – phosphorylation, orange – ubiquitylation), **(c)** disorder regions predictions based on protein amino acids sequence using PONDR-VLS2 server [96]. A score over 0.5 indicates a high probability of disorder. **(d)** LLPS propensity prediction based on catGranule server http://service.tartaglialab.com/update_submission/216885/dd56e32a89 [98].

3.2 SIM1 and SIM2

According to PONDR server prediction, all documented (Fig. 2A a) and predicted (Fig. 2A b) sites of PTMs of SIM1 are located in disordered regions, predicted in the linker between bHLH and PAS1 domains (64-76 aa) and at the long C-terminus of SIM1 (336-766 aa) (Fig. 2A c). bHLH and PAS domains and additionally short regions observed in the SIM1 C-terminus (450-500aa, 700-740aa, Fig. 2Ac) are predicted as more structured. What important, the short-ordered regions in the middle of disordered C-termini are described as characteristic for bHLH-PAS class I proteins [13]. The disease-associated missense mutations are located in disordered linker and C-terminus (Fig. 2A b, Supplementary Materials). Prediction of LLPS propensity results in local positive maximum in the linker between bHLH and PAS, linker between PAS1 and PAS2 and in N-terminal region of C-terminus (390aa). The 310-450 aa area deserves special attention. It presents high disorder propensity simultaneously with local LLPS maximum and is enriched in PTM sites. Also many disease-associated mutations are reported in this area. According to HuVarBase, SIM1 missense mutations are linked mainly with skin cancer (H394Y, H402Y, D424N, S428F, S454L, R471Q, R493C, R550C, P588L, S603F, P661L, R665C). The others: lung cancer (R192H, G392R, E530K, A570G, N650Y, S701C), breast cancer (P352T, A494T), liver cancer (H559Q,, G448C, Q704H) large intestine cancer (L217P, A371V, C472W, R548Q, S663L) stomach cancer (S541L), hematopoietic and lymphoid tissue cancer (G408R, T481M), CNS cancer (P539R) esophagus cancer (E725K) and Schaaf-Yang syndrome (Q704L) (Supplementary Materials).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0510.v1

In the case of SIM2 almost all PTM sites, both documented and predicted (Fig. 2B a, b) are placed along the long highly disordered C-terminus (336-667 aa)(Fig. 2B c). The only modification documented for this protein is phosphorylation. Similarly to previously analysed bHLH-PAS TFs, most of missense disease-associated mutations are observed in the IDRs(Fig. 2B b). Predicted LLPS propensity results in local positive maximum in the linker between bHLH and PAS (54-76 aa), also predicted as disordered. What important, this area seems to be not susceptible to PTMs, however contains high number of missense mutations. According to HuVarBase, SIM2 missense mutations are linked mainly with lung cancer (S343Y, S355F, P385H, T646P, Q469P), skin cancer (P57S, M164I, E339K, E345K,, M377I, P448S, D450N, F454S), liver cancer (F56L, A70T, G174S, F394S) and large intestine cancer (A63V, A169V, D202N, T433M), The others: endometrium cancer (K190N), cervix cancer (K368N), fallopian tube cancer (C489G), hematopoietic and lymphoid tissue cancer (A350S), bone cancer (S199Y), thyroid cancer (L483M) and upper aerodigestive tract cancer (S502W) (Supplementary Materials).

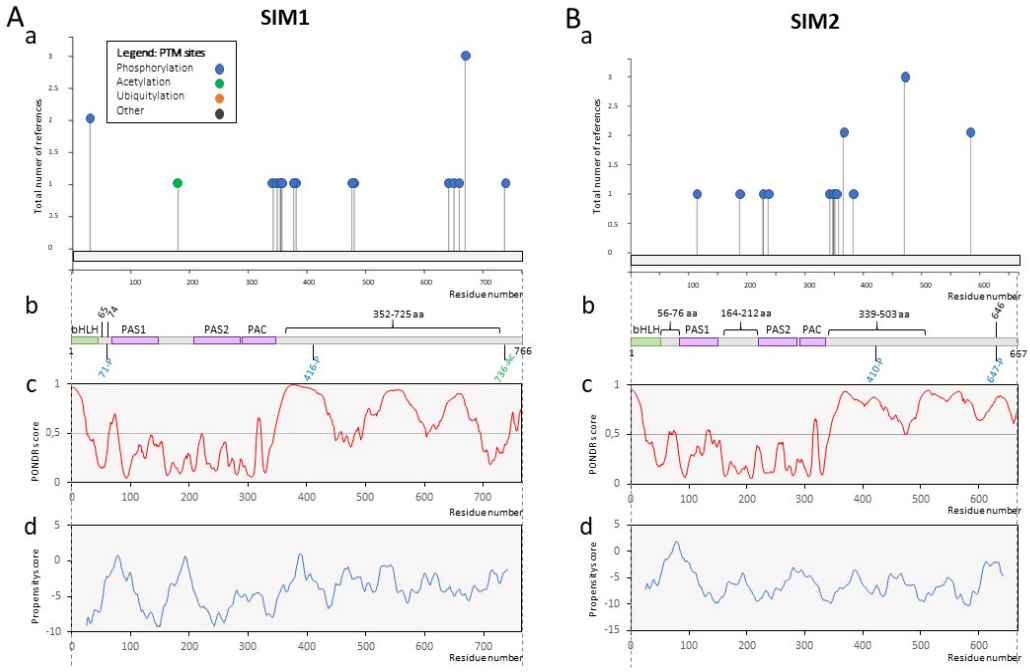

Figure 2. Schematic presentation of results for SIM1 (P81133) (A) and SIM2 (Q14190) (B) analysis.
(a) Post-translation modifications based on PhosphoSitePlus server [100], **(b)** the domain structure of protein, green indicates the bHLH domain (1-63aa SIM1; 1-53aa SIM2), whereas purple represents PAS domains 77-147aa PAS1 SIM1, 77-149aa PAS1 SIM2, 218-288aa PAS2 and 292-335aa PAC SIM1/2). Disease-associated missense mutations localized in disordered regions and annotated in HuVarBase database (62) are marked above the diagram. Post-translation modifications predicted with PTM-ssMP server [101] are marked below the diagram (blue – phosphorylation, orange – ubiquitylation), **(c)** disorder regions predictions based on protein amino acids sequence using PONDR-VLS2 server [96]. A score over 0.5 indicates a high probability of disorder. **(d)** LLPS propensity prediction based on catGranule server http://service.tartaglialab.com/update_submission/216885/dd56e32a89 [98].

3.3 Hif-2α

For Hif-2α, most of documented (Fig. 3A a) and predicted (Fig. 3A b) PTM targets are placed along the long C-terminus (348-870 aa) predicted as IDR (Fig. 3A c). Similarly to previously described proteins, most of missense mutations in the Hif-2α sequence are located in the C-terminus and in the linker between bHLH and PAS1 domains (48-83 aa) – in disordered regions (Fig. 3A b). At

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0510.v1

the same time, some of Hif-2α documented PTMs are observed in region comprising defined domains. It is caused by significantly higher in comparison to AhR or SIM proteins relaxation in area. Hif-2α in addition to phosphorylation is highly targeted by ubiquitination. Predicted LLPS propensity results in many positive maxima, on the entire protein length (Fig. 3A d). What important, these areas coincide with predicted disordered fragments. Hif-2α missense mutations are mostly linked with familial erythrocytosis (A410T, M535V, M535T, G537R, G537W, F540L, F608L, S703A, T766P, P785T, I789V, R798G, R825Q, E832D). The others: autonomic ganglia cancer (L529P, A530T, A530E ,D539Y), large intestine cancer (S372N, Y489H, S672Y, N768T) adrenal gland cancer (P531L, P531S, Y532C), pancreas cancer (T776P, A530T), hematopoietic and lymphoid tissue cancer (E82K), ovary cancer (S723N), stomach cancer (S474T), prostate cancer (M507T), lung cancer (S72L), liver cancer (L542R) and esophagus cancer (D753E) (Supplementary Materials).

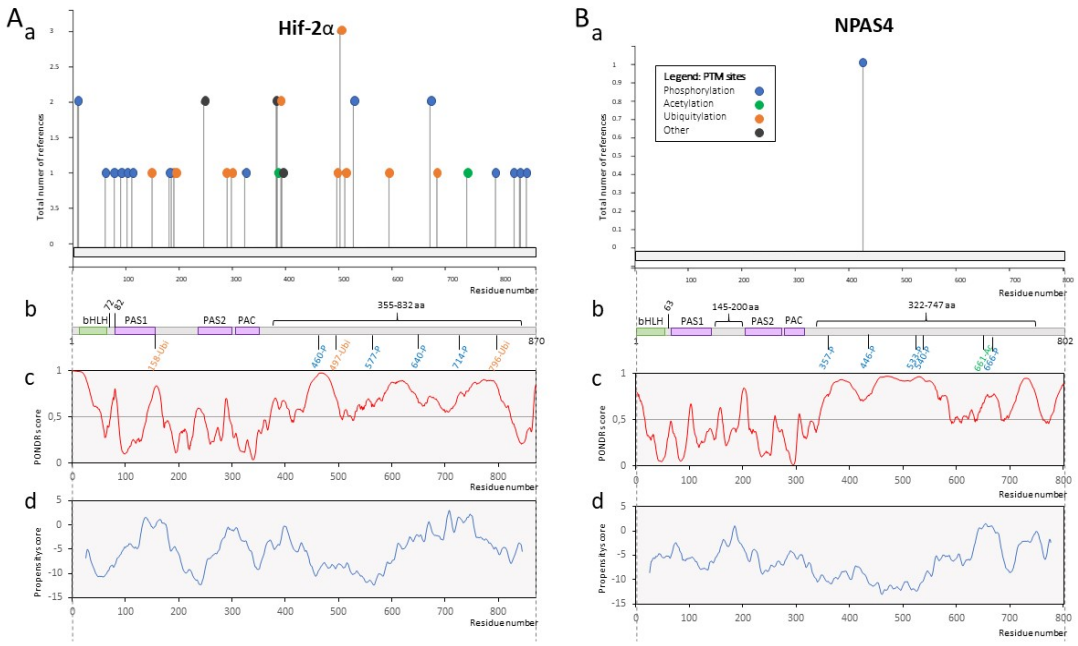

Figure 3. Schematic presentation of results for Hif-2α (Q99814) (A) and NPAS4 (Q8IUM7) (B) analysis.
(a) Post-translation modifications based on PhosphoSitePlus server [100], **(b)** the domain structure of protein, green indicates the bHLH domain (14-47aa Hif--2α; 1-53aa NPAS4), whereas purple represents PAS domains (84-154aa PAS1, 230-300aa PAS2, 304-347aa PAC HIF-2α; 70-144aa PAS1, 203-273aa PAS2, 278-317aa PAC NPAS4). Disease-associated missense mutations localized in disordered regions and annotated in HuVarBase database (62) are marked above the diagram. Post-translation modifications predicted with PTM-ssMP server [101] are marked below the diagram (blue – phosphorylation, orange – ubiquitylation), **(c)** disorder regions predictions based on protein amino acids sequence using PONDR-VLS2 server [96]. A score over 0.5 indicates a high probability of disorder. **(d)** LLPS propensity prediction based on catGranule server http://service.tartaglialab.com/update_submission/216885/dd56e32a89 [98].

3.4 NPAS4

NPAS4 is one of the immediate early genes (IEG) that activate mechanisms related to the first defense against many cellular stresses [102]. Importantly, immediate early genes are regulated by a specific stimulus without de novo synthesis protein [103]. Up to date, there is only one documented NPAS4 modification – phosphorylation (Fig. 3B a). All predicted PTMs (Fig. 3B b) are located in the C-terminus. Results of disorder prediction indicated the presence of the long IDR in C-terminal part of protein (318-802 aa) and additional short IDRs in the N-terminal part of NPAS4, comprising bHLH and PAS domains especially in the PAS1/PAS2 linker (145-202a) and in lesser extend in bHLH/PAS1 linker (54-69) (Fig. 3B c). Interestingly, catGRANULE prediction of propensity to LLPS

(Fig.3B d) showed results mostly compatible with IDRs prediction. An exception is the central part of protein (approximately 350-600 aa) with lower LLPS propensity score though high probability to be disordered both linker region. Similarly to previous analyzed proteins, in accordance to HuVarBase, disease-associated missense mutations of NPAS4 are located in IDRs, mostly (but not only) predicted also as presenting putative ability for LLPS formation. Especially interesting is fragment of C-terminal part (approximately 650-700 aa) as IDR with positive score of LLPS propensity and with predicted modifications like acetylation or phosphorylation. NPAS4 missense mutations are linked predominantly to liver cancer (R150L, P194L, Q332K, P405L, Q547H, I639V, D647N, P679L, S683I, S747F), skin cancer (R145C, P194S, D419N, L455F, P533S, P533L, S544N, T558I, D716N, E725K, D730N), large intestine cancer (R159C, R172Q, P199H, L322I, L351I) and esophagus cancer (A175T, A592V, V710M). The other reported cancers: upper aerodigestive tract (S453C, Q469H), breast (R200H, E628G), kidney (R595W), stomach (T708M), endometrium (P597S), thyroid (S493L), pancreas (R634H) cervix (Q629H), bone (E724K) and CNS (T587M) (Supplementary Materials).

3.5 ARNT2 and BMAL1

To compare different classes of bHLH-PAS TFs, we performed similar to previous analysis for ARNT2 and BMAL1 as representants of class II bHLH-PAS proteins. For ARNT2, most of documented (Fig. 4A a) and predicted (Fig. 4A b) PTMs are located in the C-terminus. This region, along with long N-terminus and the long linker between bHLH and PAS1 domains are predicted as highly disordered (Fig. 4A c). The longer regions of protein structure relaxation in the central part in comparison to described class I members, could explain the ability of class II proteins to serve as an interaction partner for different class I proteins. Most of missense mutations in the protein' sequence are located in the C-terminus and in other regions predicted as disordered. Predicted LLPS propensity results in many positive maxima, on the entire protein length (Fig. 4A d), what seems to be characteristic for class II bHLH-PAS TFs. Again, LLPS positive areas coincide with disordered fragments. ARNT2 disease-associated missense variants are linked to large intestine cancer (A28V, R47C, R240K, P579S, T602M), skin cancer (S458L, P529S), CNS cancer (Y430N), lung cancer (A25T, V683L), liver cancer (D191G, G710A), hematopoietic and lymphoid tissue cancer (H543R), pancreas cancer (P269S) and stomach cancer (G31R) (Supplementary Materials).

In the case of BMAL1 almost all PTM sites (Fig. 4B a, b) are distributed along the long C-terminus (445-626 aa), N-terminus (1-71 aa) and the linker between PAS1 and PAS2 domains (216-325 aa). All these fragments are predicted as highly disordered (Fig. 4B c). Importantly, similarly to ARNT2, the longer regions of disorder in the middle part of BMAL1 in comparison to described class I members is preserved and suggest more general characteristic of class II members with functional significance of interacting with wide spectrum of partners from class I. In contrast to all previously analysed bHLH-PAS proteins, no disease-associated missense mutation was reported in the disordered C-terminal region of BMAL1, instead, missense mutations cumulated in disordered N-terminal part. It was unexpected as the C-terminal TAD plays important role in mammalian clock regulation. Predicted LLPS propensity results in local positive maxima in the N- and C-termini in accordance to IDR prediction. Similarly . BMAL1 seems to be the subject of wider spectrum of PTMs (phosphorylation, ubiquitination, acetylation and SUMOylation) in comparison to ARNT2. BMAL1 disease-associated missense mutations are linked predominantly to large intestine cancer (D22N, S27Y, R37C, R37H, R244Q, V260A). The others: oesophagus cancer (E62Q), genital tract cancer (E65K), thyroid cancer (H66P, C249R)), skin cancer (P234H), cervix cancer (S246C), pancreas cancer (P292T), stomach cancer (T224S), breast cancer (T140S) and liver cancer (Q4L) (Supplementary Materials).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0510.v1

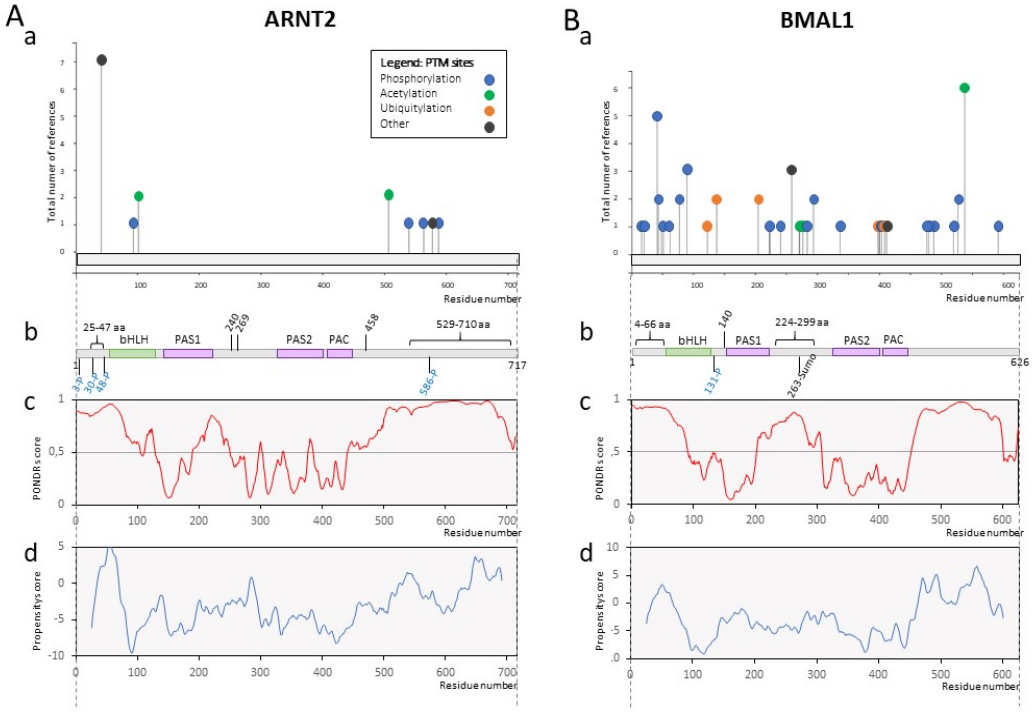

Figure 4. Schematic presentation of results for ARNT2 (Q9HBZ2) (A) and BMAL1 (O00327) (B) analysis. **(a)** Post-translation modifications based on PhosphoSitePlus server [100], **(b)** the domain structure of protein, green indicates the bHLH domain (63-116aa ARNT2;72-125aa BMAL1) , whereas purple represents PAS domains (134-209aa PAS1, 323-393aa PAS2, 398-441aa PAC ARNT2; 143-215aa PAS1, 326-396aa PAS2, 401-444aa PAC BMAL1). Disease-associated missense mutations localized in disordered regions and annotated in HuVarBase database (62) are marked above the diagram. Post-translation modifications predicted with PTM-ssMP server [101] are marked below the diagram (blue – phosphorylation, orange – ubiquitylation), **(c)** disorder regions predictions based on protein amino acids sequence using PONDR-VLS2 server [96]. A score over 0.5 indicates a high probability of disorder. **(d)** LLPS propensity prediction based on catGranule server http://service.tartaglialab.com/update_submission/216885/dd56e32a89 [98].

4. Discussion

Functional analysis concerning proteins involved in the cross between different signaling pathways and simultaneously interacting with multiple partners (hub proteins), proved that the intrinsically disordered character of interacting regions is highly important [20]. Also the DNA-binding proteins in Eukaryotes were shown to be significantly enriched in disordered domains [104]. As mentioned previously, bHLH-PAS proteins play both functions as transcription factors interacting with many physiological partners with ability to bind DNA.

As we presented, bHLH-PAS TFs in addition to long IDRs in their C-terminal regions have short IDRs localized usually prior to the bHLH domain and in the PAS domains linker. An example that illustrates the importance of disordered regions for TF function is other family member NPAS3. V304I and A552P were identified as NPAS3 missense variants associated with psychiatric disorders. V304I mutation located in PAS linker did not altered protein' molecular function, however was sequestered in the insoluble fraction of cell lysates, suggesting aggregation of protein. The second mutation A552P located in C-terminus, was documented to be sufficient for activation of reporter gene expression [105].

Making use of HuVarBase data with *in silico* analysis of selected representants of bHLH-PAS family allowed us to show that missense mutations associated with disease are located mostly in intrinsically disordered regions. For most of analysed proteins (AhRR, SIM1, Hif-2a, NPAS4) also positive propensity score to LLPS was predicted in some of disordered fragments. Additionally,

 doi:10.20944/preprints202010.0510.v1

there are often located at or in close proximity to sites undergoing PTMs. Analyzing data we observed that often S being target of phosphorylation was substituted by a residue without such ability or opposite other residue substituted by S, what suggest impact of changed protein PTM pattern in IDRs for disease development. In some cases we observed also G-A substitution which could influence IDR folding propensity. In some case mutation could obviously impact protein chain configuration, for example E/K substitution with the change of the amino acid residue charge. In some cases however, for example L/I or R/K substitution impact was not so obvious, though substitution resulted in deleterious effect as well.

Protein-protein interaction (PPI) networks relay on the intrinsically disordered nature of the proteins functionally assembling. This allows connection of different signaling pathways and creation of larger networks [106]. bHLH-PAS TFs usually function as hub proteins on the cross of many signaling pathways, what is extremally important for their activity. IDPs and IDRs functionality may depend on the ability of a disorder to order transition after binding to the partner [107] Disease-associated missense mutations most often were found in PPI regions [108], known as short linear motifs (SLiMs) [27]. The reason could be difficulties in obtaining by SLiM expected conformation after missense mutation. Recently, It was shown that transition of the peptide mimicking SLiM to a conformation with pronounced α-helical structure could be broken by an amino acid substitution by proline as a helix breaker [109]. Activity of SLiMs responsible for PPI or protein localization is regulated also by PTMs which induce protein conformational changes. If so, the missense mutation of PTMs target aa residues were suggested as important sites involved in disease induction after substitution [110].

bHLH-PAS TFs activity depends on nucleocytoplasmic shuttling, which occurs as result of SLiMs recognition by proteins responsible for nuclear export/import. Nuclear localization signal (NLS) or nuclear export signal (NES) signals were defined in the bHLH and PAS domains as well as in the C-terminal unstructured region of AhR (NLS and NES). Also, C-termini of Hif-1α and Hif-2α contain conserved NLS and NES. For SIM2 C-terminal region cytoplasmic localization was documented [111]. Finally, we demonstrated previously the presence of overlapping NES and NLS in the C-terminal region of NPAS4 [112]. PTMs like phosphorylation, especially close to the NLS/NES were shown to regulate the intracellular distribution of proteins via activation/deactivation of the localization motifs [113]. This suggest that disease-associated missense mutations located in C-termini could be linked with the change of NLS/NES activity by substitution of aa residue in signal sequence or by substitution of aa residue located close to the signal sequence and underlying modification important for this signal activity.

It was shown that cells organize many biochemical processes in membrane-less compartments arising by a process of liquid-liquid phase separation (LLPS). Interestingly, LLPS of disease-causing mutant of heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1, D262V) was shown to promote fibrillization, while the wild type protein - not [114]. Pathological neurodegeneration related to age or disease and protein aggregation has been also linked to LLPS-driven processes [22]. Proteins containing long IDRs represent an abundant class of macromolecules that can phase separate under physiological conditions. IDRs do not have stable 3D structure and often contain repeated sequence elements providing the basis for multivalent weakly adhesive intermolecular interactions responsible for LLPS formation [115]. Recently, we discussed bHLH TFs as factors putatively engaged in formation of LLPS during transcription process [99]. We propose that aberrant regulation of LLPS processes by disease-associated bHLH-PAS variants with specific missense mutations could result in diseases development.

Performed predictions of IDRs presence and LLPS propensity simultaneously with human polymorphism and posttranslational modification database analysis led us to conclude, that most of disease-associated missense mutations are in IDRs of analyzed bHLH-PAS family members and are located in close proximity to the regions important for LLPS regulation, or susceptible to PTMs. PTMs pattern change can affect protein's interaction network or protein's subcellular localization signals activity. All those can modify proteins function and induce specific disease states.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Supplementary Materials

Author Contributions: Conceptualization, B.GM. writing—original draft preparation, M.KA. and B.GM.; writing—review and editing, M.KA. and B.GM.; All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by a subsidy from The Polish Ministry of Science and High Education for the Faculty of Chemistry of Wroclaw University of Science and Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crews ST. Control of cell lineage-specific development and transcription by bHLH-PAS proteins. Genes Dev. Cold Spring Harbor Laboratory Press; 1998;12: 607–20.
2. Petrulis JR, Kusnadi A, Ramadoss P, Hollingshead B, Perdew GH. The hsp90 Co-chaperone XAP2 Alters Importin β Recognition of the Bipartite Nuclear Localization Signal of the Ah Receptor and Represses Transcriptional Activity. J Biol Chem. 2003;278: 2677–2685. doi:10.1074/jbc.M209331200
3. Gustafson CL, Partch CL. Emerging Models for the Molecular Basis of Mammalian Circadian Timing. Biochemistry. 2015;54: 134–149. doi:10.1021/bi500731f
4. Lee J-W, Bae S-H, Jeong J-W, Kim S-H, Kim K-W. Hypoxia-inducible factor (HIF-1)α: its protein stability and biological functions. Exp Mol Med. 2004;36: 1–12. doi:10.1038/emm.2004.1
5. Ema M, Hirota K, Mimura J, Abe H, Yodoi J, Sogawa K, et al. Molecular mechanisms of transcription activation by HLF and HIF1alpha in response to hypoxia: their stabilization and redox signal-induced interaction with CBP/p300. EMBO J. 1999;18: 1905–1914. doi:10.1093/emboj/18.7.1905
6. Fribourgh JL, Partch CL. Assembly and function of bHLH-PAS complexes. Proc Natl Acad Sci U S A. National Academy of Sciences; 2017;114: 5330–5332. doi:10.1073/pnas.1705408114
7. Michael AK, Partch CL. bHLH-PAS proteins: functional specification through modular domain architecture. OA Biochem. 2013; 1(2):16. Available: https://www.oapublishinglondon.com/article/1123%5Cnhttps://www.oapublishinglondon.com/article/1123#Abstract
8. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, et al. Genome-Wide Analysis of Basic/Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis. PLANT Physiol. 2006;141: 1167–1184. doi:10.1104/pp.106.080580
9. Ponting CP, Aravind L. PAS: a multifunctional domain family comes to light. Curr Biol. Cell Press; 1997;7: R674–R677. doi:10.1016/S0960-9822(06)00352-6
10. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. Int J Biochem Cell Biol. Pergamon; 2004;36: 189–204. doi:10.1016/S1357-2725(03)00211-5
11. Wu D, Rastinejad F. Structural characterization of mammalian bHLH-PAS transcription factors. Curr Opin Struct Biol. 2017;43: 1–9. doi:10.1016/j.sbi.2016.09.011
12. Partch CL, Gardner KH. Coactivator recruitment: a new role for PAS domains in transcriptional regulation by the bHLH-PAS family. J Cell Physiol. 2010;223: 553–7. doi:10.1002/jcp.22067
13. Kolonko M, Greb-Markiewicz B. bHLH–PAS Proteins: Their Structure and Intrinsic Disorder. Int J Mol Sci. 2019;20: 3653. doi:10.3390/ijms20153653
14. Uversky VN. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. Curr Opin Struct Biol. 2017;44: 18–30. doi:10.1016/j.sbi.2016.10.015
15. Mirsky AE, Pauling L. On the Structure of Native, Denatured, and Coagulated Proteins. Proc Natl Acad Sci U S A. 1936;22: 439–47.
16. Uversky VN. The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of Any Given Proteome. J Biomed Biotechnol. 2010;2010: 1–14. doi:10.1155/2010/568068
17. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci. 2002;27: 527–533. doi:10.1016/S0968-0004(02)02169-2
18. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000;41: 415–27.
19. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. Protein Sci. 2002;11: 739–756. doi:10.1110/ps.4210102
20. Hu G, Wu Z, Uversky VN, Kurgan L. Functional analysis of human hub proteins and their interactors

involved in the intrinsic disorder-enriched interactions. Int J Mol Sci. 2017;18: 1–40. doi:10.3390/ijms18122761

21. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;16: 18–29. doi:10.1038/nrm3920
22. Alberti S, Gladfelter A, Mittag T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell. Elsevier Inc.; 2019;176: 419–434. doi:10.1016/j.cell.2018.12.035
23. Mitrea DM, Kriwacki RW. Phase separation in biology; functional organization of a higher order. Cell Commun Signal. 2016;14: 1. doi:10.1186/s12964-015-0125-7
24. Bugge K, Brakti I, Fernandes CB, Dreier JE, Lundsgaard JE, Olsen JG, et al. Interactions by Disorder – A Matter of Context. Front Mol Biosci. 2020;7: 1–16. doi:10.3389/fmolb.2020.00110
25. Yoo H, Triandafillou C, Drummond DA. Cellular sensing by phase separation: Using the process, not just the products. J Biol Chem. 2019;294: 7151–7159. doi:10.1074/jbc.TM118.001191
26. Uversky VN. Supramolecular fuzziness of intracellular liquid droplets: Liquid–liquid phase transitions, membrane-less organelles, and intrinsic disorder. Molecules. 2019. doi:10.3390/molecules24183265
27. Uyar B, Weatheritt RJ, Dinkel H, Davey NE, Gibson TJ. Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer? Mol Biosyst. Royal Society of Chemistry; 2014;10: 2626–2642. doi:10.1039/c4mb00290c
28. Zhao H, Chen L, Yang T, Feng YL, Vaziri ND, Liu BL, et al. Aryl hydrocarbon receptor activation mediates kidney disease and renal cell carcinoma [Internet]. Journal of Translational Medicine. BioMed Central Ltd.; 2019. pp. 1–14. doi:10.1186/s12967-019-2054-5
29. Ikuta, T, Namiki T F. AhR protein trafficking and function in the skin. Biochem Pharmacol. Elsevier; 2009;77: 588–596. doi:10.1016/J.BCP.2008.10.003
30. Neavin D, Liu D, Ray B, Weinshilboum R. The Role of the Aryl Hydrocarbon Receptor (AHR) in Immune and Inflammatory Diseases. Int J Mol Sci. 2018;19: 3851. doi:10.3390/ijms19123851
31. Xue P, Fu J, Zhou Y. The Aryl Hydrocarbon Receptor and Tumor Immunity. Front Immunol. 2018;9: 286. doi:10.3389/fimmu.2018.00286
32. Perepechaeva ML, Grishanova AY. The role of aryl hydrocarbon receptor (AHR) in brain tumors. Int J Mol Sci. 2020;21. doi:10.3390/ijms21082863
33. Zhao H, Chen L, Yang T, Feng YL, Vaziri ND, Liu BL, et al. Aryl hydrocarbon receptor activation mediates kidney disease and renal cell carcinoma. J Transl Med. BioMed Central; 2019;17: 1–14. doi:10.1186/s12967-019-2054-5
34. Ohtake F, Takeyama K, Matsumoto T, Kitagawa H, Yamamoto Y, Nohara K, et al. Modulation of oestrogen receptor signalling by association with the activated dioxin receptor. Nature. 2003;423: 545–50. doi:10.1038/nature01606
35. Guarnieri T. Aryl hydrocarbon receptor connects inflammation to breast cancer. Int J Mol Sci. 2020;21: 1–17. doi:10.3390/ijms21155264
36. Hahn ME, Allan LL, Sherr DH. Regulation of constitutive and inducible AHR signaling : Complex interactions involving the AHR repressor §. 2009;77: 485–497. doi:10.1016/j.bcp.2008.09.016
37. Larigot L, Juricek L, Dairou J, Coumoul X. AhR signaling pathways and regulatory functions. Biochim open. Elsevier; 2018;7: 1–9. doi:10.1016/j.biopen.2018.05.001
38. Zudaire E, Cuesta N, Murty V, Woodson K, Adams L, Gonzalez N, et al. The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. J Clin Invest. 2008;118: 640–650. doi:10.1172/JCI30024
39. Kim SH, Choi YM, Lee GH, Hong MA, Lee KS, Lee BS, et al. Association between susceptibility to advanced stage endometriosis and the genetic polymorphisms of aryl hydrocarbon receptor repressor and glutathione-S-transferase T1 genes. Hum Reprod. 2007;22: 1866–1870. doi:10.1093/humrep/dem112
40. Tsuchiya M, Katoh T, Motoyama H, Sasaki H, Tsugane S, Ikenoue T. Analysis of the AhR, ARNT, and AhRR gene polymorphisms: Genetic contribution to endometriosis susceptibility and severity. Fertil Steril. 2005;84: 454–458. doi:10.1016/j.fertnstert.2005.01.130
41. Vogel CFA, Haarmann-stemmann T. ScienceDirect Toxicology The aryl hydrocarbon receptor repressor – More than a simple feedback inhibitor of AhR signaling : Clues for its role in inflammation and cancer. Curr Opin Toxicol. Elsevier Ltd; 2017;2: 109–119. doi:10.1016/j.cotox.2017.02.004
42. Woods SL, Whitelaw ML. Differential Activities of Murine Single Minded 1 (SIM1) and SIM2 on a Hypoxic Response Element. J Biol Chem. 2002;277: 10236–10243. doi:10.1074/jbc.M110752200
43. Moffett P, Reece M, Pelletier J. The murine Sim-2 gene product inhibits transcription by active repression and functional interference. Mol Cell Biol. American Society for Microbiology (ASM);

1997;17: 4933–47.
44. Moffett P, Pelletier J. Different transcriptional properties of mSim-1 and mSim-2. FEBS Lett. 2000;466: 80–86. doi:10.1016/S0014-5793(99)01750-0
45. Blackburn PR, Sullivan AE, Gerassimou AG, Kleinendorst L, Bersten DC, Cooiman M, et al. Functional Analysis of the SIM1 Variant p.G715V in 2 Patients With Obesity. J Clin Endocrinol Metab. NLM (Medline); 2020;105. doi:10.1210/clinem/dgz192
46. Holder JL, Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. Hum Mol Genet. Oxford Academic; 2000;9: 101–108. doi:10.1093/hmg/9.1.101
47. Michaud JL, Boucher F, Melnyk A, Gauthier F, Goshu E, Lévy E, et al. Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. Hum Mol Genet. Oxford University Press; 2001;10: 1465–1473. doi:10.1093/hmg/10.14.1465
48. Bonnefond A, Raimondo A, Stutzmann F, Ghoussaini M, Ramachandrappa S, Bersten DC, et al. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. J Clin Invest. American Society for Clinical Investigation; 2013;123: 3037–3041. doi:10.1172/JCI68035
49. Kublaoui BM, Holder JL, Tolson KP, Gemelli T, Zinn AR. SIM1 Overexpression Partially Rescues Agouti Yellow and Diet-Induced Obesity by Normalizing Food Intake. Endocrinology. Oxford Academic; 2006;147: 4542–4549. doi:10.1210/en.2006-0453
50. Yang C, Gagnon D, Vachon P, Tremblay A, Levy E, Massie B, et al. Adenoviral-mediated modulation of Sim1 expression in the paraventricular nucleus affects food intake. J Neurosci. J Neurosci; 2006;26: 7116–7120. doi:10.1523/JNEUROSCI.0672-06.2006
51. Camuzi D, de Amorim ÍSS, Ribeiro Pinto LF, Oliveira Trivilin L, Mencalha AL, Soares Lima SC. Regulation Is in the Air: The Relationship between Hypoxia and Epigenetics in Cancer. Cells. Multidisciplinary Digital Publishing Institute (MDPI); 2019;8. doi:10.3390/cells8040300
52. Tian H, McKnight SL, Russell DW. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. Genes Dev. 1997;11: 72–82. doi:10.1101/gad.11.1.72
53. Wiesener MS, Jürgensen JS, Rosenberger C, Scholze CK, Hörstrup JH, Warnecke C, et al. Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. FASEB J. 2003;17: 271–273. doi:10.1096/fj.02-0445fje
54. Wu D, Potluri N, Lu J, Kim Y, Rastinejad F. Structural integration in hypoxia-inducible factors. Nature. 2015;524: 303–308. doi:10.1038/nature14883
55. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. J Proteome Res. 2006;5: 879–887. doi:10.1021/pr060048x
56. Percy MJ. A Gain-of-Function Mutation in the HIF2A Gene in Familial Erythrocytosis. Eng J Med. 2008;23: 1–7. doi:10.1038/jid.2014.371
57. Taylor SE, Bagnall J, Mason D, Levy R, Fernig DG, See V. Differential sub-nuclear distribution of hypoxia-inducible factors (HIF)-1 and -2 alpha impacts on their stability and mobility. Open Biol. The Royal Society; 2016;6: 160195. doi:10.1098/rsob.160195
58. Gkotinakou I-M, Befani C, Simos G, Liakos P. ERK1/2 phosphorylates HIF-2α and regulates its activity by controlling its CRM1-dependent nuclear shuttling. J Cell Sci. The Company of Biologists Ltd; 2019;132: jcs225698. doi:10.1242/jcs.225698
59. Smythies JA, Sun M, Masson N, Salama R, Simpson PD, Murray E, et al. Inherent DNA-binding specificities of the HIF-1α and HIF-2α transcription factors in chromatin. EMBO Rep. 2019;20: e46401. doi:10.15252/embr.201846401
60. Hussein K, Percy M, McMullin MF. Clinical utility gene card for: Familial erythrocytosis. Eur J Hum Genet. Nature Publishing Group; 2012;20: 593. doi:10.1038/ejhg.2011.252
61. Ooe N, Saito K, Mikami N, Nakatuka I, Kaneko H. Identification of a Novel Basic Helix-Loop-Helix-PAS Factor, NXF, Reveals a Sim2 Competitive, Positive Regulatory Role in Dendritic-Cytoskeleton Modulator Drebrin Gene Expression. Mol Cell Biol. 2003;24: 608–616. doi:10.1128/MCB.24.2.608-616.2004
62. Sabatini P V., Krentz NAJ, Zarrouki B, Westwell-Roper CY, Nian C, Uy RA, et al. Npas4 Is a novel activity-Regulated cytoprotective factor in pancreatic β-Cells. Diabetes. 2013;62: 2808–2820. doi:10.2337/db12-1527
63. Sabatini P V., Lynn FC. All-encomPASsing regulation of β-cells: PAS domain proteins in β-cell dysfunction and diabetes. Trends Endocrinol Metab. 2015;26: 49–57. doi:10.1016/j.tem.2014.11.002
64. Furukawa-Hibi Y, Yun J, Nagai T, Yamada K. Transcriptional suppression of the neuronal PAS domain 4 (Npas4) gene by stress via the binding of agonist-bound glucocorticoid receptor to its promoter. J Neurochem. 2012;123: 866–875. doi:10.1111/jnc.12034

65. Wang D, Ren M, Guo J, Yang G, Long X, Hu R, et al. The inhibitory effects of Npas4 on seizures in pilocarpine-induced epileptic rats. PLoS One. 2014;9: 1–15. doi:10.1371/journal.pone.0115801
66. Choy FC, Klarić TS, Koblar SA, Lewis MD. The Role of the Neuroprotective Factor Npas4 in Cerebral Ischemia. Int J Mol Sci. 2015;16: 29011–28. doi:10.3390/ijms161226144
67. Choy FC, Klarić TS, Leong WK, Koblar SA, Lewis MD. Reduction of the neuroprotective transcription factor Npas4 results in increased neuronal necrosis, inflammation and brain lesion size following ischaemia. J Cereb Blood Flow Metab. SAGE PublicationsSage UK: London, England; 2016;36: 1449–1463. doi:10.1177/0271678X15606146
68. Ramamoorthi K, Fropf R, Belfort GM, Fitzmaurice HL, McKinney RM, Neve RL, et al. Npas4 Regulates a Transcriptional Program in CA3 Required for Contextual Memory Formation. Science (80-). 2011;334: 1669–1675.
69. Ploski JE, Monsey MS, Nguyen T, DiLeone RJ, Schafe GE. The Neuronal PAS Domain Protein 4 (Npas4) Is Required for New and Reactivated Fear Memories. Izquierdo I, editor. PLoS One. Public Library of Science; 2011;6: e23760. doi:10.1371/journal.pone.0023760
70. Fan W, Long Y, Lai Y, Wang X, Chen G, Zhu B. NPAS4 Facilitates the Autophagic Clearance of Endogenous Tau in Rat Cortical Neurons. J Mol Neurosci. 2016;58: 401–410. doi:10.1007/s12031-015-0692-5
71. Esser JS, Charlet A, Schmidt M, Heck S, Allen A, Lother A, et al. The neuronal transcription factor NPAS4 is a strong inducer of sprouting angiogenesis and tip cell formation. Cardiovasc Res. Oxford University Press; 2017;113: 222–223. doi:10.1093/cvr/cvw248
72. Gerhard DS, Wagner L, Feingold EA, Shenmen CM, Grouse LH, Schuler G, et al. The status, quality, and expansion of the NIH full-length cDNA project. Genome Res. 2004;14: 2121–2127. doi:10.1101/gr.2596504.2
73. Ooe N, Saito K, Kaneko H. Characterization of functional heterodimer partners in brain for a bHLH-PAS factor NXF. Biochim Biophys Acta. 2009;1789: 192–7. doi:10.1016/j.bbagrm.2009.01.003
74. Sullivan AE, Peet DJ, Whitelaw ML. MAGED1 is a novel regulator of a select subset of bHLH PAS transcription factors. FEBS J. 2016;283: 3488–3502. doi:10.1111/febs.13824
75. Shepard R, Heslin K, Hagerdorn P, Coutellier L. Downregulation of Npas4 in parvalbumin interneurons and cognitive deficits after neonatal NMDA receptor blockade: relevance for schizophrenia. Transl Psychiatry. Springer US; 2019;9. doi:10.1038/s41398-019-0436-3
76. Coutellier L, Beraki S, Ardestani PM, Saw NL, Shamloo M. Npas4: a neuronal transcription factor with a key role in social and cognitive functions relevant to developmental disorders. PLoS One. 2012;7: e46604. doi:10.1371/journal.pone.0046604
77. Michaud JL, Derossi C, May NR, Holdener BC, Fan CM. ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus. Mech Dev. Mech Dev; 2000;90: 253–261. doi:10.1016/S0925-4773(99)00328-7
78. Sharma N, Pollina EA, Nagy MA, Yap EL, DiBiase FA, Hrvatin S, et al. ARNT2 Tunes Activity-Dependent Gene Expression through NCoR2-Mediated Repression and NPAS4-Mediated Activation. Neuron. Elsevier Inc.; 2019;102: 390-406.e9. doi:10.1016/j.neuron.2019.02.007
79. Okur Z, Scheiffele P. The Yin and Yang of Arnt2 in Activity-Dependent Transcription. Neuron. Elsevier Inc.; 2019;102: 270–272. doi:10.1016/j.neuron.2019.04.006
80. Dougherty EJ, Pollenz RS, Florida S. 2 . 13 ARNT : A Key bHLH / PAS Regulatory Protein Across Multiple Pathways. 2010;
81. Aitola MH, Pelto-Huikko MT. Expression of Arnt and Arnt2 mRNA in developing murine tissues. J Histochem Cytochem. 2003;51: 41–54. doi:10.1177/002215540305100106
82. Hirose K, Morita M, Ema M, Mimura J, Hamada H, Fujii H, et al. cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt). Mol Cell Biol. 1996;16: 1706–13. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231157&tool=pmcentrez&rendertype=abstract
83. Webb EA, Almutair A, Kelberman D, Bacchelli C, Chanudet E, Lescai F, et al. ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. Brain. Oxford University Press; 2013;136: 3096–3105. doi:10.1093/brain/awt218
84. Woods CG. Human microcephaly. Current Opinion in Neurobiology. Elsevier Ltd; 2004. pp. 112–117. doi:10.1016/j.conb.2004.01.003
85. Turer EE, Miguel MS, Wang K wen, McAlpine W, Ou F, Li X, et al. A viable hypomorphic Arnt2 mutation causes hyperphagic obesity, diabetes and hepatic steatosis. DMM Dis Model Mech. 2018;11. doi:10.1242/dmm.035451

86. Bogeas A, Morvan-Dubois G, El-Habr EA, Lejeune FX, Defrance M, Narayanan A, et al. Changes in chromatin state reveal ARNT2 at a node of a tumorigenic transcription factor signature driving glioblastoma cell aggressiveness. Acta Neuropathol. Springer Berlin Heidelberg; 2018;135: 267–283. doi:10.1007/s00401-017-1783-x
87. Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science (80-). 2012;338: 349–354. doi:10.1126/science.1226339
88. Gustafson CL, Parsley NC, Asimgil H, Lee HW, Ahlbach C, Michael AK, et al. A Slow Conformational Switch in the BMAL1 Transactivation Domain Modulates Circadian Rhythms. Mol Cell. Elsevier Inc.; 2017;66: 447-457.e7. doi:10.1016/j.molcel.2017.04.011
89. Lee Y, Lee J, Kwon I, Nakajima Y, Ohmiya Y, Son GH, et al. Coactivation of the CLOCK-BMAL1 complex by CBP mediates resetting of the circadian clock. J Cell Sci. 2010;123: 3547–3557. doi:10.1242/jcs.070300
90. Zheng X, Zhao X, Zhang Y, Tan H, Qiu B, Ma T, et al. RAE1 promotes BMAL1 shuttling and regulates degradation and activity of CLOCK: BMAL1 heterodimer. Cell Death Dis. Springer US; 2019;10. doi:10.1038/s41419-019-1346-2
91. Tamaru T, Hirayama J, Isojima Y, Nagai K, Norioka S, Takamatsu K, et al. CK2α phosphorylates BMAL1 to regulate the mammalian clock. Nat Struct Mol Biol. Nature Publishing Group; 2009;16: 446–448. doi:10.1038/nsmb.1578
92. Lipton JO, Yuan ED, Boyle LM, Ebrahimi-Fakhari D, Kwiatkowski E, Nathan A, et al. The Circadian Protein BMAL1 Regulates Translation in Response to S6K1-Mediated Phosphorylation. Cell. NIH Public Access; 2015;161: 1138–1151. doi:10.1016/j.cell.2015.04.002
93. Geyfman M, Kumar V, Liu Q, Ruiz R, Gordon W, Espitia F, et al. Brain and muscle Arnt-like protein-1 (BMAL1) controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis. Proc Natl Acad Sci U S A. 2012;109: 11758–63. doi:10.1073/pnas.1209592109
94. Ganesan K, Kulandaisamy A, Binny Priya S, Michael Gromiha M. HuVarbase: A human variant database with comprehensive information at gene and protein levels. PLoS One. 2019;14: 1–7. doi:10.1371/journal.pone.0210475
95. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta. 2010;1804: 996–1010. doi:10.1016/j.bbapap.2010.01.011
96. Li, Romero, Rani, Dunker, Obradovic. Predicting Protein Disorder for N-, C-, and Internal Regions. Genome Inform Ser Workshop Genome Inform. 1999;10: 30–40. Available: http://www.ncbi.nlm.nih.gov/pubmed/11072340
97. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005;21: 3433–3434. doi:10.1093/bioinformatics/bti541
98. Bolognesi B, Gotor NL, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, et al. A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. Cell Rep. The Author(s); 2016;16: 222–231. doi:10.1016/j.celrep.2016.05.076
99. Tarczewska A, Greb-Markiewicz B. The Significance of the Intrinsically Disordered Regions for the Functions of the bHLH Transcription Factors. Int J Mol Sci. 2019;20. doi:10.3390/ijms20215306
100. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43: D512–D520. doi:10.1093/nar/gku1267
101. Liu Y, Wang M, Xi J, Luo F, Li A. PTM-ssMP: A web server for predicting different types of post-translational modification sites using novel site-specific modification profile. Int J Biol Sci. 2018;14: 946–956. doi:10.7150/ijbs.24121
102. Fowler T, Sen R, Roy AL. Regulation of primary response genes. Mol Cell. NIH Public Access; 2011;44: 348–60. doi:10.1016/j.molcel.2011.09.014
103. Greenberg ME, Hermanowski AL, Ziff EB. Effect of protein synthesis inhibitors on growth factor activation of c-fos, c-myc, and actin gene transcription. Mol Cell Biol. American Society for Microbiology (ASM); 1986;6: 1050–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/2431274
104. Wang C, Uversky VN, Kurgan L. Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. Proteomics. 2016;16: 1486–1498. doi:10.1002/pmic.201500177
105. Luoma LM, Berry FB. Molecular analysis of NPAS3 functional domains and variants. BMC Mol Biol. BioMed Central; 2018;19: 1–19. doi:10.1186/s12867-018-0117-4
106. Emily Bowler1, Zhenghe Wang2 and RME. How do oncoprotein mutations rewire protein-protein

interaction networks? Perspectives and techniques. Physiol Behav. 2017;176: 139–148. doi:10.1586/14789450.2015.1084875.How

107. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol. 2002;12: 54–60. doi:10.1016/S0959-440X(02)00289-0
108. Wong ETC, So V, Guron M, Kuechler ER, Malhis N, Bui JM, et al. Protein–protein interactions mediated by intrinsically disordered protein regions are enriched in missense mutations. Biomolecules. 2020;10: 1–19. doi:10.3390/biom10081097
109. Sharma N, Fonin A V., Shpironok OG, Silonov SA, Turoverov KK, Uversky VN, et al. Folding perspectives of an intrinsically disordered transactivation domain and its single mutation breaking the folding propensity. Int J Biol Macromol. Elsevier B.V.; 2020;155: 1359–1372. doi:10.1016/j.ijbiomac.2019.11.111
110. Huang Q, Chang J, Cheung MK, Nong W, Li L, Lee MT, et al. Human proteins with target sites of multiple post-translational modification types are more prone to be involved in disease. J Proteome Res. 2014;13: 2735–2748. doi:10.1021/pr401019d
111. Greb-Markiewicz B, Kolonko M. Subcellular Localization Signals of bHLH-PAS Proteins: Their Significance, Current State of Knowledge and Future Perspectives. Int J Mol Sci. Multidisciplinary Digital Publishing Institute; 2019;20: 4746. doi:10.3390/ijms20194746
112. Greb-Markiewicz B, Zarębski M, Ożyhar A. Multiple sequences orchestrate subcellular trafficking of neuronal PAS domain-containing protein 4 (NPAS4). J Biol Chem. American Society for Biochemistry and Molecular Biology; 2018; jbc.RA118.001812. doi:10.1074/jbc.RA118.001812
113. Jans DA, Hübner S. Regulation of protein transport to the nucleus: central role of phosphorylation. Physiol Rev. American Physiological Society; 1996;76: 651–85. Available: http://www.ncbi.nlm.nih.gov/pubmed/8757785
114. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, et al. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell. Elsevier Inc.; 2015;163: 123–133. doi:10.1016/j.cell.2015.09.015
115. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: Organizers of cellular biochemistry. Nat Rev Mol Cell Biol. Nature Publishing Group; 2017;18: 285–298. doi:10.1038/nrm.2017.7