Supplementary Material

Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR

Nan Hao[1,2], Murray L. Whitelaw[1,2], Keith E. Shearwin[1], Ian B. Dodd[1], & Anne Chapman-Smith[1,2] *

[1]School of Molecular and Biomedical Science (Biochemistry), University of Adelaide, Adelaide, SA 5005, Australia, and [2]Australian Research Council Special Centre for the Molecular Genetics of Development, University of Adelaide.

* e-mail: anne.chapmansmith@adelaide.edu.au

Supplementary Figure Legends

Supplementary Figure 1: bHLH.PAS transcription factor networks.

Network interactions within the basic Helix-Loop-Helix (bHLH)/Per-Arnt-Sim homology (PAS) transcription factor family are centered around two hub proteins (ovals), Brain and Muscle Arnt-like (BMAL) and Aryl hydrocarbon Receptor Nuclear Translocator (Arnt), which can homodimerize (open arrowheads) as well as heterodimerize (closed arrowheads). Heterodimeric partner proteins (rectangles) interact with the hub proteins (closed arrowheads) and do not homodimerize or interact with each other, with the exception of the negative regulator, Inhibitory PAS (IPAS).

The BMAL cluster regulates circadian rhythm and includes BMAL 1 & 2, Clock 1 & 2 and PERIOD (PER) proteins. The latter contain PAS domains but not the bHLH region (triangles), and are negative regulators of transcription.

The Arnt cluster senses and regulates response to environmental cues such as xenobiotics (Aryl hydrocarbon Receptor (AhR)) and molecular oxygen (Hypoxia Inducible Factor-αs (HIF-1α, 2α and 3α)). Members in the Arnt cluster also have roles in many developmental, adaptive and disease processes, including development of the liver (AhR), kidney and lung (Single-Minded 2 (Sim2) Neuronal PAS proteins 1 and 3 (NPAS1 and NPAS3)), fertility (AhR); angiogenesis, erythropoiesis and vasculogenesis (HIF-αs; Inhibitory PAS protein (IPAS)); feeding response (Single Minded 1 (Sim1)) as well as neurogenesis, neurosignaling (Sim1, Sim2, NPAS1 and NPAS3) and synaptic plasticity (NPAS4) (1-4,5 and references therein).

The hub protein Arnt is ubiquitously expressed while the Arnt2 homologue has

tissue-restricted expression and is less well studied. Although the interaction between Arnt2 and most other bHLH.PAS proteins in the Arnt network has been demonstrated in vitro, it is thought that the role of Arnt2 is restricted to bHLH.PAS TFs expressed in brain in vivo, e.g. NPAS4, Sim1 and Sim2.

Supplementary Figure 2: CD spectra of Arnt362 mutants

CD spectra were recorded in 10 mM sodium phosphate pH 8.0 at 20^{o}C on a Jasco J-815 spectrometer, from 260-185 nm in 0.2 nm steps, with 1 sec response time and 1 nm bandwidth, recording 7 accumulations per spectrum, with buffer correction. Data were normalised for differences in protein concentration using ellipticity measured at 207 nm to allow comparison with wt (6).

Supplementary Figure 3: Cross species conservation of mutated residues in Arnt PAS.A.

Sequences were retrieved from NCBI database with the following accession numbers: Human (NP_001659), Chimpanzee (XP_001170422), RhesusMonkey (XP_001106465), HouseMouse (NP_001032826), NorwayRat (NP_036912), GoldenHamster (BAF02595), GuineaPig (NP_001166569), Rabbit (NP_001075675), Dog (XP_540303), Cattle (NP_776418), Chicken (NP_989531), BaikalSeal (BAE16957), ClawedFrog (NP_001116925), RainbowTrout (NP_001118182), Zebrafish (NP_001007790), FruitFly (NP_731308), and aligned using ClustalW2 multiple sequence alignment server at EMBL-EBI. The mutations identified from RevB2H are shown on the alignment, colour coded according to the strength of disruption observed in the β-gal assay (Fig. 2a; red: < 60%, magenta: 61-70%, blue: 71-80%, and green: >80% of wt activity). The position of Hepa c4 mutation in mArnt (G341D) is shown in yellow. Black boxes indicate residues conserved across species, blue boxes indicate a conservative substitution.

Supplementary Figure 4: Specific activation of pML-6c reporter constructs through the CME or XRE response element by bHLH.PAS heterodimers.

293T cells were transiently co-transfected with puro6_HisMyc_Arnt wt together with puro6 constructs expressing the indicated partner proteins, and the pML-6c-wt (CME-driven), -X (XRE-driven) or -AM (mutant) firefly luciferase reporter plasmid (7) and phRL-CMV renilla luciferase plasmid for 24 h. Relative luciferase activity was normalized to activity with empty vector (puro6) to give fold activation. Data are mean ± s.d. of transfections performed in triplicate and are representative of 2 independent experiments.

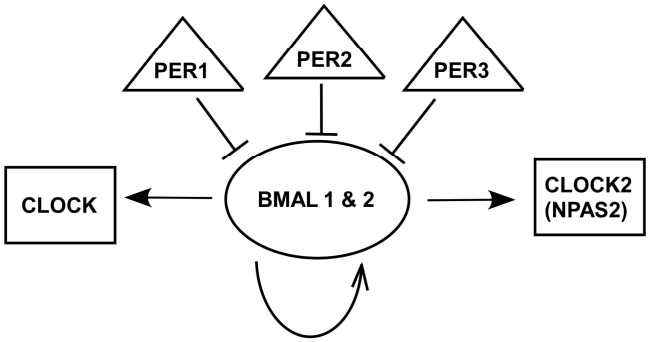
Supplementary Figure 5: Arnt E163K showed increased E-box driven activation relative to wt over a concentration range.

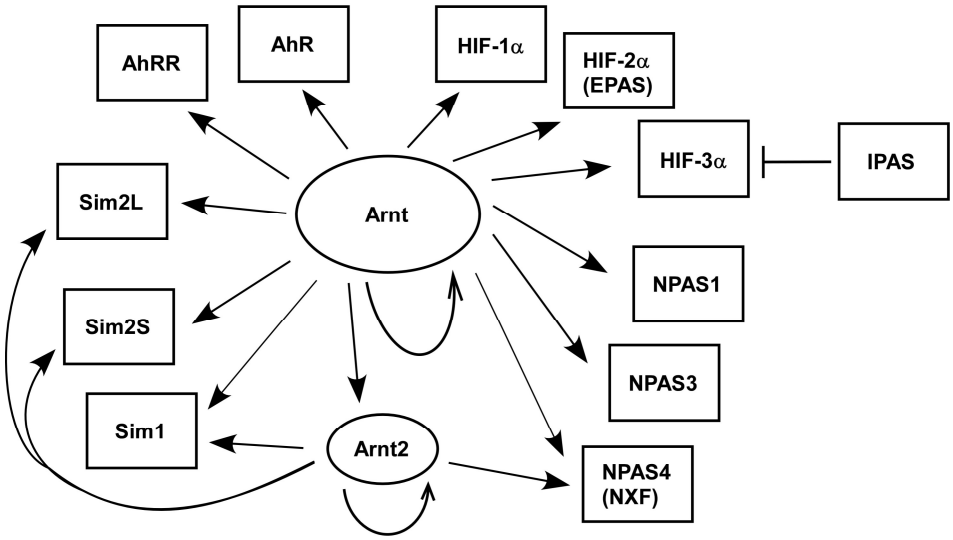
293T cells were transiently transfected with the indicated concentrations of puro6_HisMyc_Arnt wt or E163K, 100 ng 4(CACGTG)TKMPluc driven firefly luciferase reporter plasmid or TKMPluc empty vector, and phRL-CMV renilla luciferase plasmid for 48 h. Relative luciferase activity was determined as described in Methods. Data are mean ± s.d. of transfections performed in triplicate and are representative of 2 independent experiments.

Supplementary Figure 6: Cross species conservation of mutated residues in AhR PAS.A.

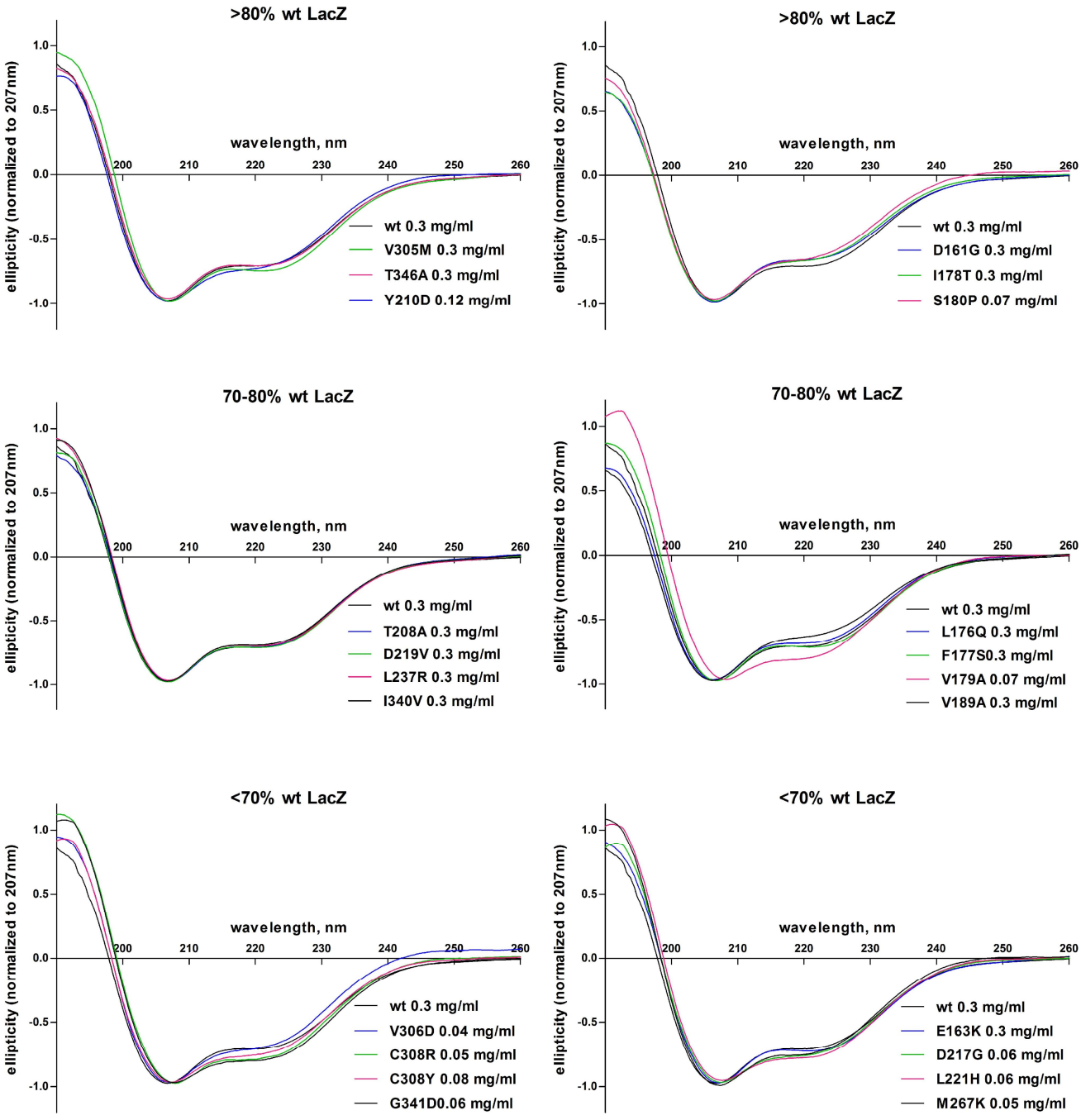
Sequences were retrieved from NCBI database with the following accession numbers: HouseMouse (NP_038492), Human (NP_001612), Chimpanzee (XR_024110, translated from partial mRNA sequence), RhesusMonkey (XP_001103903), NorwayRat (NP_037281), GoldenHamster (BAA99559), GuineaPig (NP_001166525), Rabbit (NP_0010756745), Dog (XP_532485), Cattle (XP_612996), Chicken (NP_989449), PekinDuck (AAF15279, partial sequence), CommonTern (AAF15281), BaikalSeal (BAB88683), BelugaWhale (AAL04031), Zebrafish_AhR2 (NP_571339), Zebrafish_AhR1b (NP_001019987), Zebrafish_AhR1a (NP_571103), FruitFly (NP_476748), and aligned using ClustalW2 multiple sequence alignment server at EMBL-EBI. The mutations identified from RevB2H are shown on the alignment, colour coded according to the strength of disruption observed in the β-gal assay (Fig. 5a; red: < 60%, magenta: 61-80%, green: >80% of wt activity). Black boxes indicate residues conserved across species, blue boxes indicate a conservative substitution.

Supplementary Figure 1

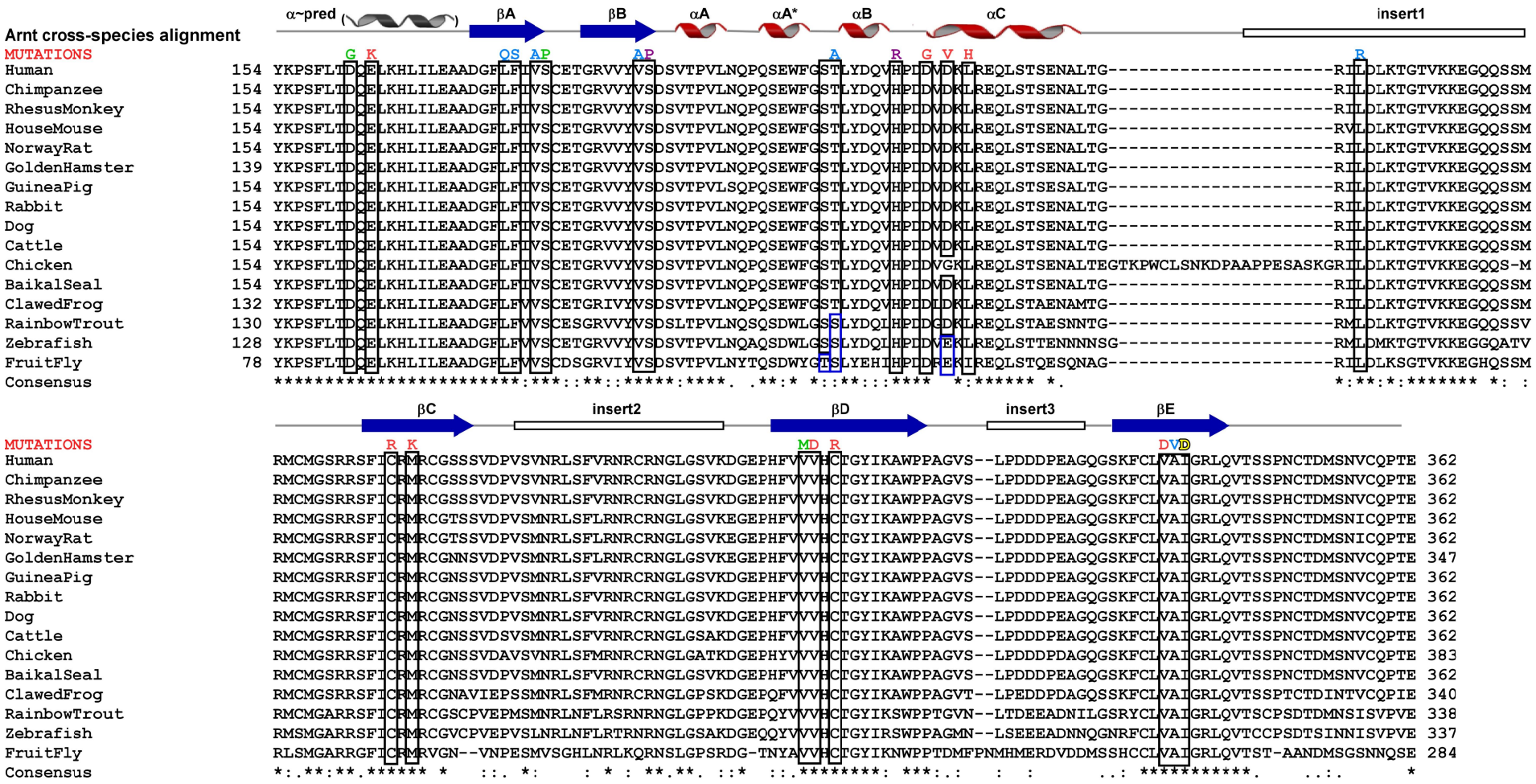




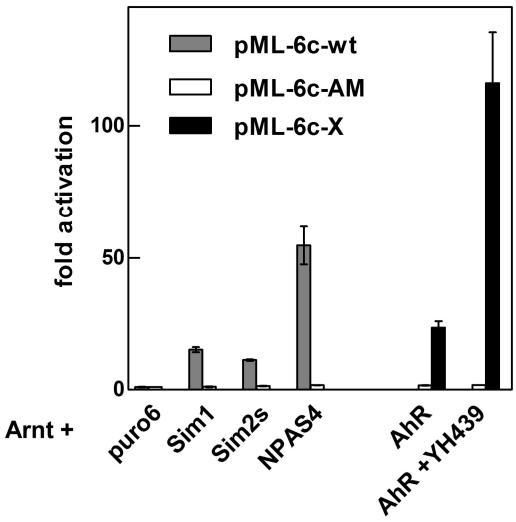
Supplementary Figure 2



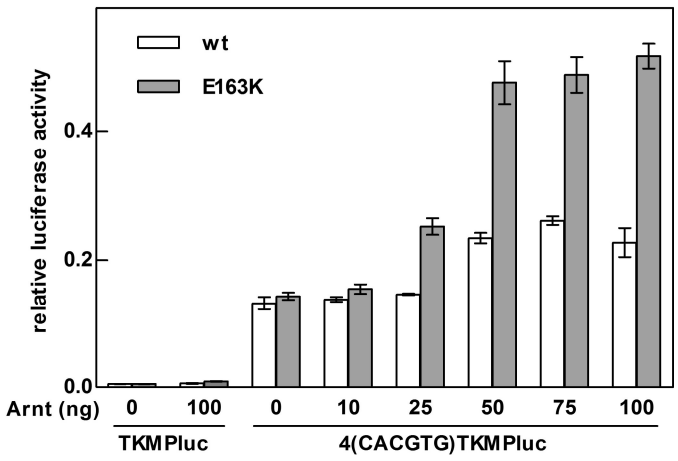
Supplementary Figure 3



Supplementary Figure 4

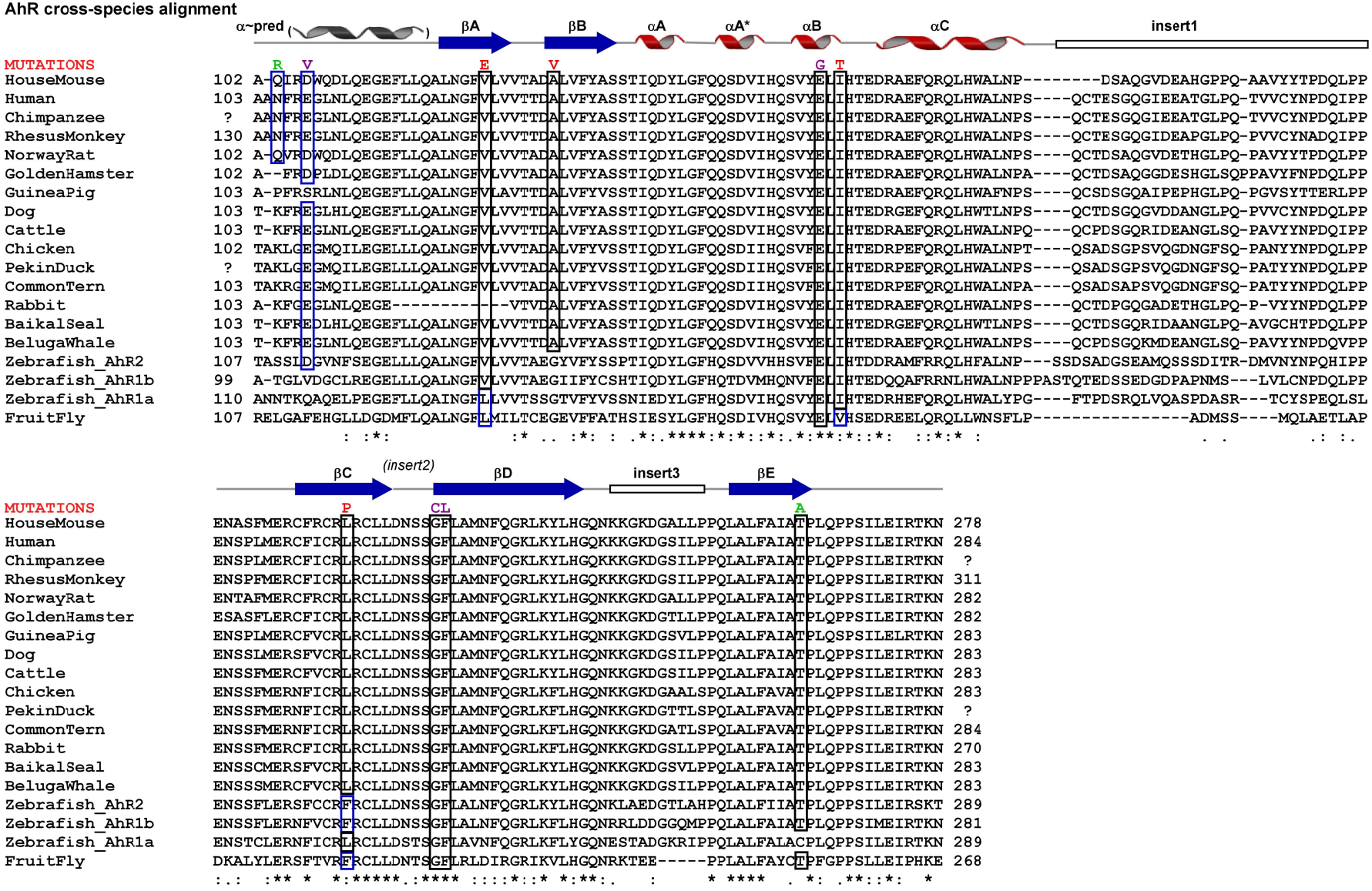


Supplementary Figure 5



Supplementary Figure 6

AhR cross-species alignment



		hArnt	%	mAhR	%
Total colonies screened by β-gal activity		**192**		**120**	
Isolates sequenced *		**91**	100	**24**	100
Mutations	single	**38**	42	**14**	58
	double	**18**	20	**1**	4
	triple	**8**	9	**0**	0
	> 3	**3**	3	**2**	8
	nonsense	**2**	2	**0**	0
	silent	**7**	8	**2**	8
	none	**15**	16	**4**	16
Single mutants reanalysed for β-gal activity		**26**	29	**14**	58
Confirmed reduction in β-gal activity		**22**	24	**10**	42

Supplementary Table 1. Summary of sequencing data for RevB2H selection and screening of Arnt and AhR PAS.A domains. This illustrates the efficiency of the 186cI CTD selection method for isolating loss of interaction mutations, as well as the utility of the additional alpha complementation approach in screening out nonsense and frame-shift mutations.

* Isolates sequenced had < 80% or < 90% wt β-gal activity for Arnt or AhR analyses, respectively, since the presence of LacZα as AhR C-terminal fusion reduced total activity for the wt interaction in the initial screening assay for AhR.

Supplementary Methods:

Plasmid contruction

lacOR2-62_186cI CTD selection plasmid

The pZS4lacOR2-62_186cI_CTD selection plasmid was made in three steps. First, an artificial lacOR2-62 promoter, harboring a λCI binding site (OR2) centered at -62 relative to the transcription start site, was generated by PCR amplification of the plac promoter with primers lac_OR2-62_AatII and lac_NheI (see below). This promoter is equivalent to that described by Dove and colleagues (8). The lacOR2-62 promoter was subcloned into plasmid pZE15_186cI_CTD_His (9) at AatII/NheI sites upstream of the *186cI CTD* gene. The lacOR2-62_186cI_CTD fragment was excised with AatII and AvrII, and re-ligated into AatII/AvrII sites of pUHS4/3-30/Pd plasmid (10) to give pZS4lacOR2-62_186cI_CTD.

λCI_bait and RNAPα_prey fusion protein bacterial expression plasmids

To make the RNAPα_AhR expression plasmid (pZE1_RNAPα_AhR278), cDNA encoding mouse AhR bHLH.PAS.A amino acids 1 to 278 was PCR amplified from pET32 _mAhR287 (11) with primers mAhR_1N_NotI and T7 terminator, and inserted into the NotI/XhoI site of pTRG (Agilent Technologies) downstream of the plpp/lacUV5 promoter. Next, the plpp/lacUV5 promoter-RNAPα_AhR278 fragment was PCR amplified with primers plpp/lacUV5_AatII and mAhR278C_AvrII, and shuttled into AatII/AvrII sites of pZE15_186cI_CTD_His, replacing the 186cI_CTD_His cassette.

Similarly, pZE1_RNAPα_Gal11p was made by PCR amplifying pTRG_Gal11p (Agilent Technologies) with primer pair plpp/lacUV5_AatII and Gal11_352C_AvrII, and inserting into the AatII/AvrII sites of pZE15_186cI_CTD_His. Modification of pZE1_RNAPα_AhR278 was also made to convert it into a host vector for screening purpose (pZE1_RNAPα_AhR278_Host_lacZα). To do this, a silent SalI site was introduced into AhR at amino acid 88 by PCR with primers mAhR_SalI_upper and mAhR_SalI _lower. Next, the SalI site present in RNAPα of pZE1_RNAP_AhR278 was removed by site directed mutagenesis using RNAPα_deltaSalI_upper and RNAPα_deltaSalI_lower primer set, making the SalI site in AhR a unique site in the plasmid. Finally, overlap extension PCR was performed using primers mAhR_17N, mAhR278_LacZ1_Upper, mAhR278_LacZ1_Lower, and LacZ_39C_AvrII, which allows in frame fusion of LacZα at the 3' end of AhR278, giving pZE1_RNAPα_AhR278_Host_lacZα.

To make the λCI_Arnt_lacZα expression plasmid (pBT_Arnt362_Host_LacZα), cDNA encoding human Arnt bHLH.PAS.A amino acids 1 to 362 was PCR amplified from pAC28_ Arnt362 (11) with primers hArnt_1N_NotI and hArnt_362C_XhoI, and inserted into NotI/XhoI sites of plasmid pBT (Agilent Technologies). A unique silent BamHI site was introduced at the

codons for amino acids 152/153 of Arnt by site directed mutagenesis with primers hArnt_BamHI_upper and hArnt_BamHI_lower. Finally, overlap extension PCR was performed using primers hArnt_68N, hArnt362_LacZ1_Upper, hArnt362_LacZ1_Lower, and LacZ_39C_XhoI, which allows in frame fusion of LacZα (amino acids 1-49) to the 3' end of Arnt362, giving pBT_Arnt362_Host_LacZα.

The Arnt bHLH domain (amino acids 1-142) was amplified by PCR with primers hArnt_1N_NotI and hArnt_142C_XhoI and cloned into NotI andf XhoI sites of pBT to create pBT_λCI_Arnt142. pZE1_RNAPα_AhR83 was constructed by PCR amplification of mAhR bHLH domain (amino acids 1-83) with primer pair plpp/lacUV5_AatII and mAhR_83C_AvrII, digested at HindIII and AvrII sites, and ligated into HindIII/AvrII sites of pZE1_RNAPα_AhR278.

The λCI_bait and RNAPα_prey fusion proteins were expressed from plpp/lacUV5 promoters upon induction with isopropyl β-D-1-thiogalactopyranoside (IPTG).

Mammalian expression plasmids

The construction of full length His_Myc tagged mAhR mammalian expression vector (pEFpuro/HisMyc_AhRfull) has been described (12). Mutations in AhR were introduced by swapping the wild type AhR PAS.A sequence in pEFpuro/HisMyc_AhR with mutated fragments from pZE1_RNAP_AhR278_Host_lacZα mutagenic plasmids using unique BlpI/BstBI sites. The full length His_Myc tagged Arnt mammalian expression vector (pEFpuro/HisMyc_Arnt_Host) was constructed by replacing the AhR coding sequence in pEFpuro/HisMyc_AhR with full length Arnt cDNA in two steps. First, an N-terminal Arnt fragment (amino acids 1-267) carrying the unique internal silent BamHI site was PCR amplified from pBT_Arnt362_Host with primers hArnt_1N_MluI and hArnt362C_AvrII, digested with MluI/BlpI, and inserted into MluI/BlpI sites of pEFpuro/HisMyc_AhR. Then, a C-terminal Arnt fragment (amino acids 268-789) was PCR amplified with primers Arnt_68N, and Arnt_789C_ClaI, digested with BlpI/ClaI and inserted into BlpI/ClaI sites of the plasmid. Mutations in Arnt were introduced by swapping the wild type Arnt PAS.A sequence in pEFpuro/HisMyc_Arnt_Host with mutated fragments from pBT_Arnt362_Host_LacZα mutagenic plasmids using the unique BamHI/BsrGI sites.

To construct lentiviral expression vectors pLV416neo_HisMyc_Arnt_Host, 6HisMyc tagged Arnt coding sequence was PCR amplified from pEFpuro/HisMyc_Arnt_Host plasmid with primers HisMyc_SalI and hArnt_789C_NotI, and ligated into SalI/NotI sites of pENTR1A (Invitrogen). Point mutations were introduced by swapping the wild type PAS.A sequence with mutated fragments using the unique BamHI/SpeI sites as described above. The wild type or mutant HisMyc_Arnt coding sequences were transferred into the lentiviral expression vector

pLV416neo (a kind gift from Dr. Simon Barry) by standard Gateway cloning according to the manufacturer's instructions (Invitrogen).

Error Prone PCR

The Arnt PAS.A mutagenic library was constructed in two steps. First, to make the template for error prone PCR, wt Arnt_PAS.A_LacZα was PCR amplified from pBT_Arnt362_Host_LacZα plasmid with high fidelity PfuTurbo DNA polymerase (Agilent Technologies) and primers hArnt_G152_BamHI and LacZ_39C_XhoI. Next, 12 ng of this template DNA was amplified for either 6 or 7 PCR cycles (i.e. 6 or 7 cycles library) using the same primer set, Taq DNA polymerase (New England Biolabs) and a final Mg^{2+} of 6 mM. The PCR product was digested with BamHI and XhoI, and ligated into pBT_Arnt362_Host_LacZα, replacing wt Arnt_PAS.A_LacZα.

AhR PAS.A mutagenic libraries were generated similarly. AhR PAS.A was first amplified from pZE1_RNAP_AhR278_Host_lacZα using mAhR_82N_SalI and lacZ_1C_BsrGI primers using PfuTurbo DNA polymerase, to serve as template for the second round of PCR. Following error prone PCR for 6 or 7 cycles, the resultant library was digested with SalI and BsrGI, and ligated back to pZE1_RNAP_AhR278_Host_lacZα, replacing wt AhR_PAS.A coding sequence.

Cloning Primers

Gal11_352C_AvrII	GTCATGCCTAGGTTATTACAAAGCTTGGATTTTTCTCAGG
hArnt_142C_XhoI	CAAGATCTCGAGTTACAAGGACTTCATGTGAG
hArnt_1N_MluI	AATGACACGCGTATGGCGGCGACTACTGCC
hArnt_1N_NotI	CATAAGCATGCGGCCGCAATGGCGGCGACTACTGC
hArnt_362C_XhoI	CAAGATCTCGAGTTACTCTGTTGGTTGACAAAC
hArnt_789C_ClaI	AATGACATCGATTTACTATTCTGAAAAGGGGGG
hArnt_789C_NotI	AATGACGCGGCCGCTTACTATTCTGAAAAGGGGGG
hArnt_BamHI_lower	GGAAAGACGGCTTATAGGATCCATCAGTGGATGTG
hArnt_BamHI_upper	CACATCCACTGATGGATCCTATAAGCCGTCTTTCC
hArnt_G152_BamHI	CACATCCACTGATGGATCC
hArnt362_LacZ1_Lower	GCCAGTGAATCCGTAATCATCTCTGTTGGTTGACAAACATT
hArnt362_LacZ1_Upper	AATGTTTGTCAACCAACAGAGATGATTACGGATTCACTGGC
HisMyc_SalI	ACTTCAGTCGACGCGACCATGGGGGGTCAC
lac_NheI	CAGATTGCTAGCAATTCCACACAACATACGAGCC
lac_OR2-62_AatII	TCGATGACGTCTAACACCGTGCGTGTTGCATTAGGCACCCCAGGC
lacZ_1C_BsrGI	GTGAATCCGTAATCATTGTACA
LacZ_39C_AvrII	GTCATGCCTAGGTTATTATTCGCTATTACGCCAGCTG
LacZ_39C_XhoI	TTAACTCTCGAGTTATTCGCTATTACGCCAGCTG
LacZ_39C_XhoI	TTAACTCTCGAGTTATTCGCTATTACGCCAGCTG
mAhR_17N	CCGGTGCAGAAAACAGTAAAG
mAhR_1N_NotI	CATAAGAATGCGGCCGCAATGGGCAGCGGCGCC
mAhR_82N_SalI	TTGATGTTGCATTAAAGTCGAC
mAhR_83C_AvrII	GTCATGCCTAGGATCAAAGAAGCTCTTGGCCC
mAhR_SalI _lower	CCATTTCTGTCAGCAGGGGTCGACTTTAATGCAACATCAAAG
mAhR_SalI_upper	CTTTGATGTTGCATTAAAGTCGACCCCTGCTGACAGAAATGG
mAhR278_LacZ1_Lower	GCCAGTGAATCCGTAATCATTGTACAGTTTTTGGTTCGAATTTCCAGG
mAhR278_LacZ1_Upper	CCTGGAAATTCGAACCAAAAACTGTACAATGATTACGGATTCACTGGC
mAhR278C_AvrII	GTCATGCCTAGGTTATTAGTTTTTGGTTCGAATTTC
plpp/lacUV5_AatII	TCGATGACGTCACATGAGAATTGAGCGAACG
RNAPα_deltaSalI_lower	CTGTAGCATGCGTCCACCAGCAGACGGC
RNAPα_deltaSalI_upper	GCCGTCTGCTGGTGGACGCATGCTACAG
T7 terminator	GCTAGTTATTGCTCAGCGGTGG

Diagnostic primers & Sequencing primers

hArnt_68N	GTCTAACGATAAGGAGCG
mAhR_40N	GACCGGCTGAACACAGAGTTAG
pBT_antisense	GGGTAGCCAGCAGCATCC
pBT_sense	TCCGTTGTGGGGAAAGTTATC
pTRG_antisense	GCTCACTCAAAGGCGGTAATACGGTT
pTRG_sense	TGGCTGAACAACTGGAAGC

qRT-PCR primers

mCyp1a1_F	ACCTCTTTGGAGCTGGGTTT
mCyp1a1_R	GATAGGGCAGCTGAGGTCTG
mHPRT_F	AGTCCCAGCGTCGTGATTAGC
mHPRT_R	CCAAATCCTCGGCATAATG
mVEGF_F	CTGCTGTACCTCCACCATGC
mVEGF_R	CGCTGGTAGACATCCATGAAC

Supplementary References:

1. Zhou, S., Degan, S., Potts, E.N., Foster, W.M. and Sunday, M.E. (2009) NPAS3 is a trachealess homolog critical for lung development and homeostasis. *Proc Natl Acad Sci U S A*, **106**, 11691-11696.
2. Brunskill, E.W., Ehrman, L.A., Williams, M.T., Klanke, J., Hammer, D., Schaefer, T.L., Sah, R., Dorn, G.W., 2nd, Potter, S.S. and Vorhees, C.V. (2005) Abnormal neurodevelopment, neurosignaling and behaviour in Npas3-deficient mice. *Eur J Neurosci*, **22**, 1265-1276.
3. Levesque, B.M., Zhou, S., Shan, L., Johnston, P., Kong, Y., Degan, S. and Sunday, M.E. (2007) NPAS1 regulates branching morphogenesis in embryonic lung. *Am J Respir Cell Mol Biol*, **36**, 427-434.
4. Erbel-Sieler, C., Dudley, C., Zhou, Y., Wu, X., Estill, S.J., Han, T., Diaz-Arrastia, R., Brunskill, E.W., Potter, S.S. and McKnight, S.L. (2004) Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc Natl Acad Sci U S A*, **101**, 13648-13653.
5. Kewley, R.J., Whitelaw, M.L. and Chapman-Smith, A. (2004) The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol*, **36**, 189-204.
6. Raussens, V., Ruysschaert, J.M. and Goormaghtigh, E. (2003) Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method. *Anal Biochem*, **319**, 114-121.
7. Moffett, P. and Pelletier, J. (2000) Different transcriptional properties of mSim-1 and mSim-2. *FEBS Lett*, **466**, 80-86.
8. Dove, S.L., Joung, J.K. and Hochschild, A. (1997) Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature*, **386**, 627-630.
9. Pinkett, H.W., Shearwin, K.E., Stayrook, S., Dodd, I.B., Burr, T., Hochschild, A., Egan, J.B. and Lewis, M. (2006) The structural basis of cooperative regulation at an alternate genetic switch. *Mol Cell*, **21**, 605-615.
10. Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, **25**, 1203-1210.
11. Chapman-Smith, A., Lutwyche, J.K. and Whitelaw, M.L. (2004) Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators. *J Biol Chem*, **279**, 5353-5362.
12. Whelan, F., Hao, N., Furness, S.G., Whitelaw, M.L. and Chapman-Smith, A. (2010) Amino acid substitutions in the aryl hydrocarbon receptor ligand binding domain reveal YH439 as an atypical AhR activator. *Mol Pharmacol*, **77**, 1037-1046.