

Update of human and mouse forkhead box (FOX) gene families

Brian C. Jackson,¹ Christopher Carpenter,¹ Daniel W. Nebert^{2*} and Vasilis Vasiliou^{1*}

¹Molecular Toxicology and Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Denver, Aurora, CO 80045, USA

²Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH 45267-0056, USA

*Correspondence to: Tel: +1 513 821 4664; Fax: +1 513 558 0925; E-mail: dan.nebert@uc.edu;

Tel: +1 303 724 3520; Fax: +1 303 724 7266; E-mail: vasilis.vasiliou@ucdenver.edu

Date received (in revised form): 29 April 2010

Abstract

The forkhead box (FOX) proteins are transcription factors that play complex and important roles in processes from development and organogenesis to regulation of metabolism and the immune system. There are 50 FOX genes in the human genome and 44 in the mouse, divided into 19 subfamilies. All human FOX genes have close mouse orthologues, with one exception: the mouse has a single *Foxd4*, whereas the human gene has undergone a recent duplication to a total of seven (*FOXD4* and *FOXD4L1* → *FOXD4L6*). Evolutionarily ancient family members can be found as far back as the fungi and metazoans. The DNA-binding domain, the forkhead domain, is an example of the winged-helix domain, and is very well conserved across the FOX family and across species, with a few notable exceptions in which divergence has created new functionality. Mutations in FOX genes have been implicated in at least four familial human diseases, and differential expression may play a role in a number of other pathologies — ranging from metabolic disorders to autoimmunity. Furthermore, FOX genes are differentially expressed in a large number of cancers; their role can be either as an oncogene or tumour suppressor, depending on the family member and cell type. Although some drugs that target FOX gene expression or activity, notably proteasome inhibitors, appear to work well, much more basic research is needed to unlock the complex interplay of upstream and downstream interactions with FOX family transcription factors.

Keywords: FOX, forkhead domain, winged-helix domain, transcription factors, gene family, evolution, cancer, metabolism, immunity

Introduction

Gene expression is controlled at multiple levels, including modulation of transcriptional activity, mRNA processing, and post-translational modification of proteins. The forkhead box (FOX) gene family encodes proteins that regulate the transcription of genes participating in a number of functions — including development of various organs, regulation of senescence or proliferation, and metabolic homeostasis. The first FOX gene to be discovered was *forkhead* (*flh*) in *Drosophila*, which, when mutated, gives the insect a fork-headed appearance.¹

Independently, another group characterised *FOXA1* in the rat.² In 1990, Weigel *et al.* discovered that these two proteins shared a similar DNA-binding domain, and named this domain the forkhead domain (also referred to as the winged-helix domain); this domain is well conserved among all FOX family members.³ At about 100 amino acids in length, the prototypical forkhead domain is monomeric and consists of three alpha-helices, three beta-sheets and two large loops ('wing' regions) that flank the third beta-sheet.⁴ In 1993, the crystal structure of the forkhead domain

bound to DNA was solved for FOXA1.⁵ Since then, several other structures have been solved, including the DNA-binding domains of FOXK1, FOXK2, FOXM1, FOXO1, FOXO3, FOXO4, FOXP1 and FOXP2 (Protein Data Bank search⁶).

Before 2000, *FOX* genes lacked a unified naming convention and were assigned a confusing array of names by the researchers who discovered them. The winged helix/forkhead nomenclature committee defined the FOX family as all genes/proteins having sequence homology to the canonical winged helix/forkhead DNA-binding domain; subclasses FOXA to FOXO were defined based on a phylogenetic analysis of the forkhead domain (other domains were highly divergent, and alignment was unclear between subclasses).⁷ Since then, the family has been expanded to include subclasses FOXP to FOXS.

The FOX gene family

The *FOX* family consists of 50 members in the human genome (plus two known pseudogenes, *FOXO1B* and *FOXO3B*) and 44 members in the mouse genome (Table 1). Ancient and diverse, the FOX family plays a role in a wide array of developmental, general and tissue-specific functions, and thousands of papers and hundreds of reviews have been published on the topic (PubMed search, April 2010). In 2007, Tuteja and Kaestner published a snapshot of human *FOX* genes, including a table of potential regulatory partners, cellular and developmental roles, known mouse mutant phenotypes and roles in human disease.^{8,9}

FOXO subfamily

A great deal of attention has been given to the FOXO subfamily, which plays a role in the regulation of metabolism, oxidative stress resistance and cell-cycle arrest. Under fasting conditions, FOXO transcriptionally activates insulin-responsive genes, which include genes encoding enzymes responsible for gluconeogenesis (including glucose-6-phosphatase and phosphoenolpyruvate carboxykinase) in the liver. If nutrients are plentiful, however, insulin activates the

phosphatidylinositol 3-kinase (PI3K) pathway, which causes AKT/protein kinase B to phosphorylate FOXO, excluding it from the nucleus and turning off insulin-responsive genes.¹⁰ FOXO1 appears to be the primary protein in this pathway, but there is overlap in many FOXO functions. In pancreatic beta cells, FOXO1 has been shown to protect against glucotoxic and lipotoxic oxidative stress by upregulating manganese superoxide dismutase and catalase.¹¹ FOXO1 has also been shown to be a cell-cycle inhibitor in beta cells¹² and in lipocytes with a correlated increased expression of p2.¹³ The post-transcriptional regulation of the FOXO family allows for complex and sensitive actions from each family member. The variety of possible phosphorylation, acetylation, methylation and O-linked glycosylation has been dubbed the FOXO code.¹⁴

FOXA subfamily

FOXA proteins have been shown to act as 'pioneer' factors — proteins that can open tightly compacted chromatin without the involvement of the switching-sucrose non-fermenting (SWI-SNF) chromatin remodelling complex.¹⁵ This is accomplished by interaction of the C-terminal domain of FOXA proteins with histones H3 and H4. This helps FOXA factors to regulate the development of multiple organ systems — including liver, pancreas, lung, prostate and kidney.¹⁶ The 'pioneer' function of FOXA has also been shown to facilitate the binding of nuclear hormone receptors, including the glucocorticoid receptor and oestrogen receptor (ER). In the adult, FOXA proteins have been shown to play a role in metabolism — for example, in the expression of gluconeogenic enzymes in the liver in response to fasting, and energy utilisation by adipose tissue in response to excess caloric intake.

FOXP subfamily

The FOXP subfamily plays a role in immune response; specifically, constitutively expressed FOXP3 is considered a critical biomarker for thymus-derived natural T_{reg} cells. FOXP3 expression is required for self-tolerance and

Table 1. List of the FOX genes in humans and mice, chromosomal locations and percentage sequence similarity between the human and mouse orthologous proteins.

Human		Mouse		% Similarity
Protein	Chromosome	Orthologue	Chromosome	
FOXA1	14q12-q13	FOXA1A	12	96.2
FOXA2	20p11	FOXA2	2	97.4
FOXA3	19q13.2-q13.4	FOXA3	7	90.4
FOXB1	15q22	FOXB1	9	99.7
FOXB2	9q21.2	FOXB2	19	92.6
FOXC1	6p25	FOXC1	13	95.1
FOXC2	16q24.1	FOXC2	8	92.6
FOXD1	5q12-q13	FOXD1	13	85.8
FOXD2	1p34-p32	FOXD2	4	94.3
FOXD3	1p31.3	FOXD3	4	90.8
FOXD4	9p24.3			
FOXD4L1	2q14.1	FOXD4	19	63.7
FOXD4L2	9p12			
FOXD4L3	9q13			
FOXD4L4	9q13			
FOXD4L5	9q13			
FOXD4L6	9q12			
FOXE1	9q22	FOXE1	4	91.2
FOXE3	1p32	FOXE3	4	79.3
FOXF1	16q24	FOXF1	8	97.4
FOXF2	6p25.3	FOXF2	13	94.6
FOXG1	14q11-q13	FOXG1	12	96.3
FOXH1	8q24	FOXH1	15	76.3
FOXI1	5q34	FOXI1	11	89.9
FOXI2	10q26.2	FOXI2	7	78.6
FOXI3	2p11.2	FOXI3	6	84.5
FOXJ1	17q25.1	FOXJ1	11	94.5
FOXJ2	12p13.31	FOXJ2	6	93.2
FOXJ3	1p34.2	FOXJ3	4	96.5

Continued

Table 1. Continued

Human		Mouse		% Similarity
Protein	Chromosome	Orthologue	Chromosome	
FO XK1	7p22	FO XK1	5	92
FO XK2	17q25	FO XK2	11	95
FO XL1	16q24	FO XL1	8	73
FO XL2	3q23	FO XL2	9	96.8
FO XM1	12p13	FO XM1	6	88.6
FO XN1	17q11-q12	FO XN1	11	90.9
FO XN2	2p22-p16	FO XN2	17	48
FO XN3	14q24.3-q31	FO XN3	12	90
FO XN4	12q24.12	FO XN4	5	86.2
FO XO1	13q14.1	FO XO1	3	95.4
FO XO3	6q21	FO XO3	10	96.1
FO XO4	Xq13.1	FO XO4	X	93.3
FO XO6	1p34.2	FO XO6	4	86
FO XP1	3p14.1	FO XP1	6	94.6
FO XP2	7q31	FO XP2	6	99.7
FO XP3	Xp11.23	FO XP3	X	91.4
FO XP4	6p21.1	FO XP4	17	80.5
FO XQ1	6p25	FO XQ1	13	93.6
FO XR1	11q23.3	FO XR1	9	77.7
FO XR2	Xp11	FO XR2	X	73
FO XS1	20q11.1-q11.2	FO XS1	2	83.3

immune homeostasis.¹⁷ The forkhead domain of FOXP members is different from that in other FOX family members: wing 1 is truncated and wing 2 forms a helix rather than a loop; in addition, the forkhead domain is located near the C-terminus, rather than the N-terminus, as in most subclasses. FOXP members are also unusual in that they can form dimers by domain swapping (two monomers interact by exchanging helix H3 and strands S2 and S3).¹⁸ The orientation of the dimers requires the protein to bind opposing (non-

adjacent) DNA sites; the result is that FOXP proteins may participate in the regulation of higher-order protein–DNA complexes. FOXP2 is involved in the development of speech, and a mutation in this gene has been linked to speech and language disorders.¹⁹

Evolution of the FOX genes

As mentioned previously, the 100-amino acid forkhead-binding domain is well conserved across

species and families. This domain is often exclusively used for phylogenetic analysis. Figure 1 shows a neighbour-joining dendrogram of human and mouse forkhead domains. Nineteen subfamilies, denoted by different letters (A, B, C, etc.) can be distinguished based on evolutionary divergence. Note that FOXN proteins are split into two distinct subgroups, at the top and bottom of the dendrogram, and that the FOXL proteins are also split into different branches. Figure 2 shows the same alignment using the full FOX protein sequence. Due to high divergence on either side of the forkhead domains, the full sequence is difficult to align between subfamilies, but has been helpful in defining each subfamily.⁷ In Figure 2, however, one again can see the splitting of the FOXN proteins and the FOXL proteins into distinctly separate branches. This global alignment also divides the FOXJ members into far different branches. One can conclude that alignment of the forkhead domain only provides a better assessment of evolutionary divergence.

Where possible, the nomenclature committee gave the same name to orthologues across species. Based on analysis of full sequences, mice have orthologues of all human FOX genes with high sequence similarity, with one exception. The murine FOXD4 protein clusters together with seven human proteins — FOXD4, and the FOXD4-like FOXD4L1 to FOXD4L6, and shares the most identity with FOXD4L1. The duplications that gave rise to the FOXD4 group appear to be relatively recent — that is, during hominid evolution. Very little information is available about these proteins, but researchers have shown that at least two *FOXD4L* genes are transcriptionally active; furthermore, evidence of purifying selection in the forkhead domain of these proteins suggests that they may play a physiological role.²⁴

Many authors have performed detailed phylogenetic analyses, but the first analysis and naming scheme has been generally upheld,⁴ with a few criticisms illustrated in Figures 1 and 2. In analysis of the forkhead domain, FOXR falls within the FOXN subclass and some authors have proposed combining these two groups. In analysis of the full

sequence, however, FOXR1 and FOXR2 are associated more closely with FOXN1 and FOXN4, but not with FOXN2 and FOXN3. Thus, some researchers have proposed splitting these into three subclasses. Finally, in many analyses, FOXL1 and FOXL2 do not cluster together.

In an analysis of the origin and expansion of early transcription factors, it was found that FOXJ1 was probably the oldest family member, present in the opisthokont last common ancestor of the fungi and metazoans.²⁵ Expansion occurred early, with bilaterians having 19 FOX genes and most mammals more than 40.

FOXs in disease

Immune dysregulation/polyendocrinopathy/enteropathy/X-linked syndrome (IPEX) is a human disease caused by a mutation in *FOXP3*. The disorder is characterised by a wide range of autoimmune symptoms, including type 1 diabetes, eczema, food allergy, thyroid disorders and inflammatory bowel disease.²⁶ The link between *FOXP3* and autoimmune disease has been ascribed to the function of T_{reg} cells; replacement or stimulation of these cells has been suggested as an aetiology for a number of autoimmune disorders.¹⁷

Given their key role in the expression of many genes that affect cell proliferation and survival, FOX family members have been suggested as possible cancer therapeutic targets. FOX family members have been shown to be up- or downregulated in many cancers. *FOXP1* has been suggested as a tumour promoter and an oncogene, depending on the cell type, although these observations are primarily based on correlations between mRNA levels and clinical outcomes.²⁷ *FOXA1* is a pioneer factor, which is required for the expression of many oestrogen-responsive genes. Nakshatri and Badve²⁸ suggest that maintenance of *FOXA1* expression may force ER- α -positive breast cancers to remain oestrogen dependent, increasing their responsiveness to anti-oestrogens. There is a correlation between retinoic acid (a *FOXA1* inducer) and growth inhibition in cells, and insulin (which inhibits *FOXA1*) and anti-oestrogen resistance.¹⁴

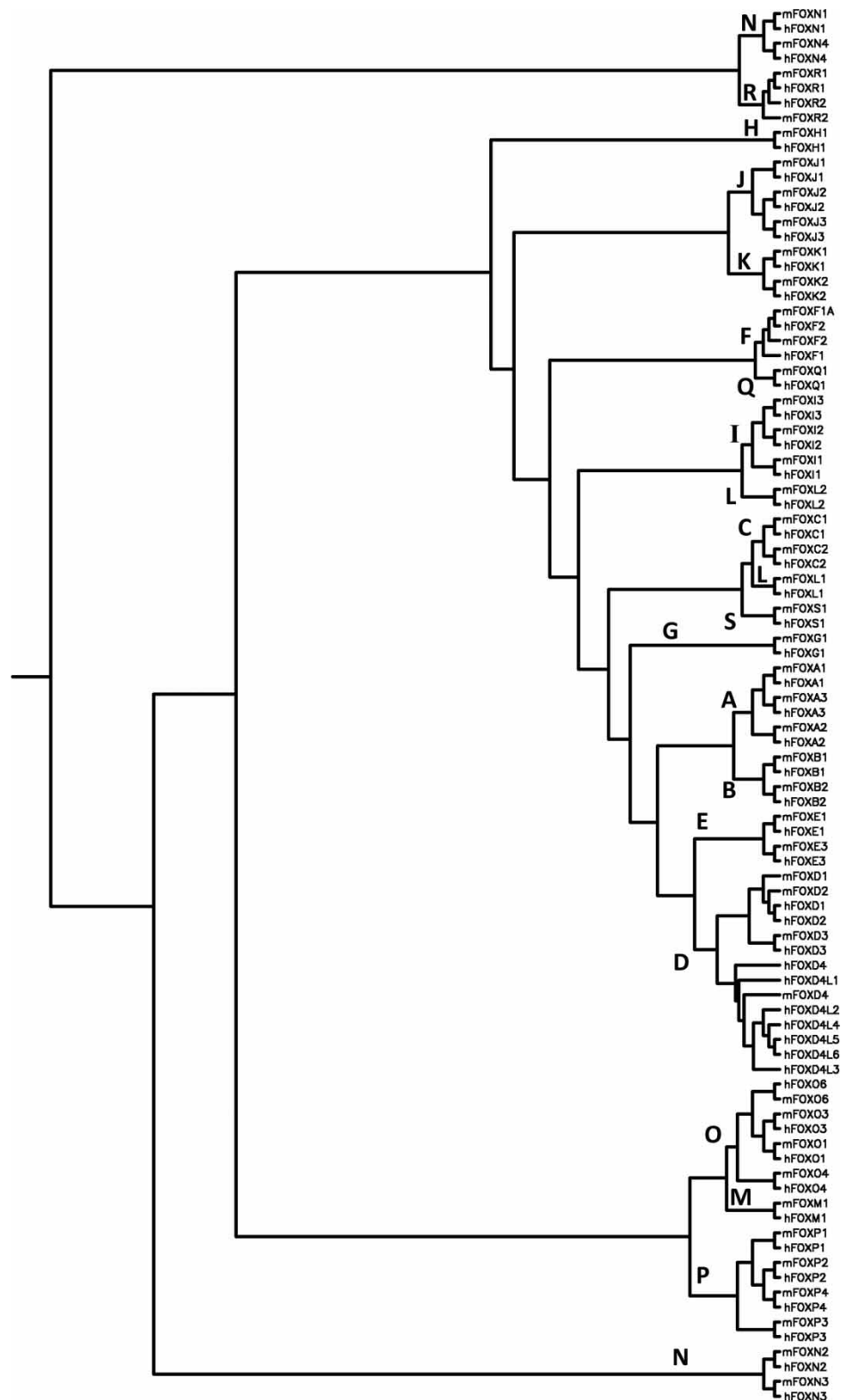


Figure 1. Neighbour-joining dendrogram of the forkhead domains of all human and mouse FOX proteins. Protein sequences were aligned using ClustalW. Labels at branches indicate subclasses. Sequence information was downloaded from UniProt²⁰ through the Hugo Gene Nomenclature Committee (HGNC) website at <http://www.genenames.org>²¹ for human, and Mouse Genome Informatics (MGI) at <http://www.informatics.jax.org>²² for mouse. Alignment and trees were created using CLUSTALW²³ through the GenomeNet server at <http://align.genome.jp>.

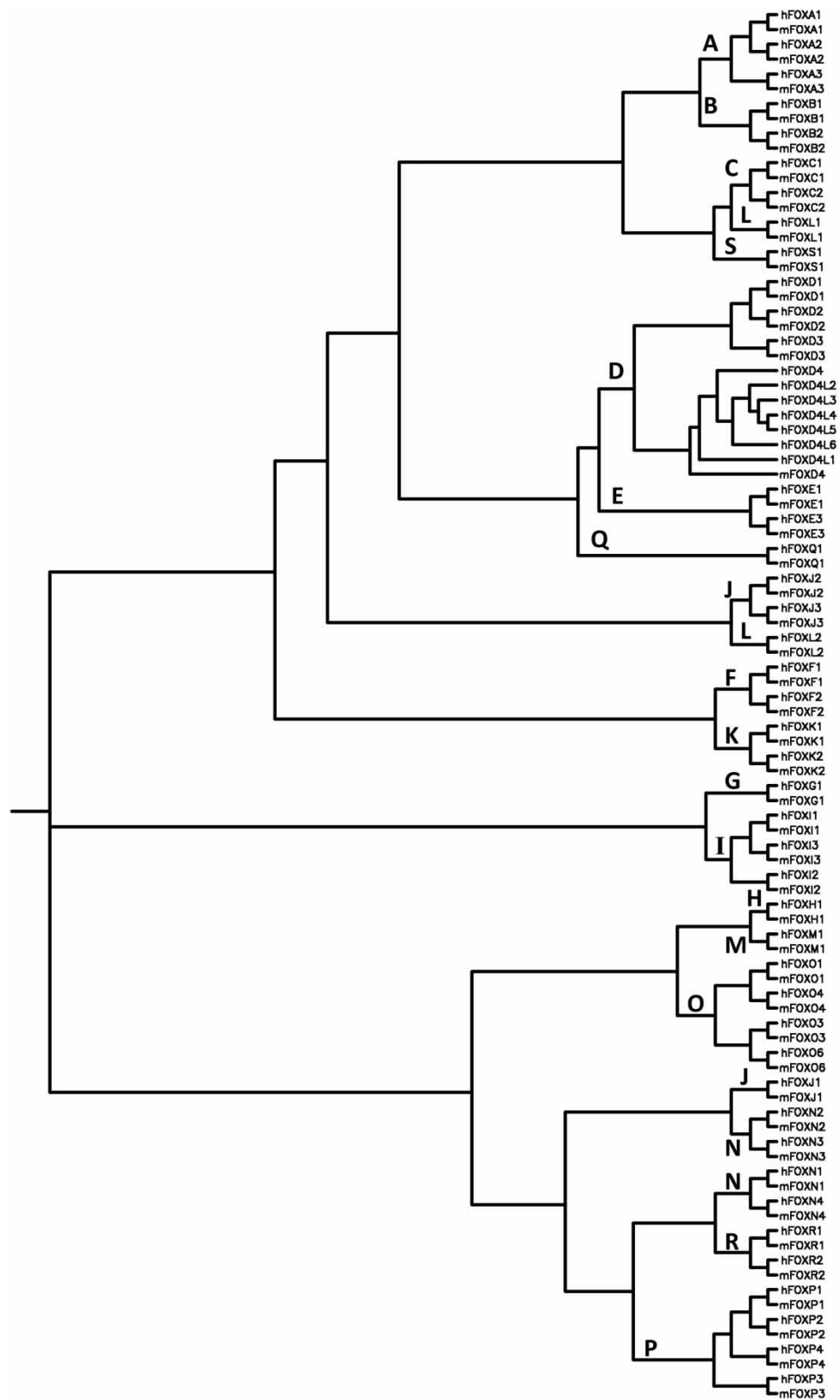


Figure 2. Neighbour-joining dendrogram of the full sequences of all human and mouse FOX proteins. Sequences were aligned using ClustalW. Labels at branches indicate subfamilies.

FOXM1 is an oncogene that is highly expressed in most carcinomas, but expressed in low amounts in normal cells, and has been recently identified as a key target of both well known and new classes of proteasome inhibitors.²⁹

The Online Mendelian Inheritance in Man (OMIM) database lists four known FOX genes that cause human diseases: *FOXC1* mutations elicit dominantly inherited glaucoma phenotypes, *FOXC2* mutations lead to lymphoedema-distichiasis syndrome, *FOXP2* loss of function leads to language acquisition defects and *FOXP3* mutations are associated with severe autoimmune disorders such as IPEX.⁴

Conclusions

It is clear that the FOX family is an important and complex family of proteins that is a tantalising therapeutic target in many disease states — especially given their role in the regulation of metabolism (metabolic syndromes), the immune system (autoimmunity) and proliferation (cancer). The family includes 50 genes in human and 44 in mouse. Each of these targets requires a protein- and tissue-specific approach. Thus, more basic research will be required to understand their regulation and activity by identifying upstream and downstream protein partners, mechanisms of action, spatiotemporal expression and post-translational modifications.

References

- Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. *et al.* (1989), 'The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo', *Cell* Vol. 57, pp. 645–658.
- Lai, E., Prezioso, V.R., Smith, E., Litvin, O. *et al.* (1990), 'HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally', *Genes Dev.* Vol. 4, pp. 1427–1436.
- Weigel, D. and Jackle, H. (1990), 'The fork head domain: A novel DNA binding motif of eukaryotic transcription factors?', *Cell* Vol. 63, pp. 455–456.
- Hannenhalli, S. and Kaestner, K.H. (2009), 'The evolution of Fox genes and their role in development and disease', *Nat. Rev. Genet.* Vol. 10, pp. 233–240.
- Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993), 'Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5', *Nature* Vol. 364, pp. 412–420.
- Berman, H., Henrick, K. and Nakamura, H. (2003), 'Announcing the worldwide Protein Data Bank', *Nat. Struct. Biol.* Vol. 10, p. 980.
- Kaestner, K.H., Knochel, W. and Martinez, D.E. (2000), 'Unified nomenclature for the winged helix/forkhead transcription factors', *Genes Dev.* Vol. 14, pp. 142–146.
- Tuteja, G. and Kaestner, K.H. (2007), 'Forkhead transcription factors II', *Cell* Vol. 131, p. 192.
- Tuteja, G. and Kaestner, K.H. (2007), 'SnapShot: forkhead transcription factors I', *Cell* Vol. 130, p. 1160.
- Gross, D.N., Wan, M. and Birnbaum, M.J. (2009), 'The role of FOXO in the regulation of metabolism', *Curr. Diab. Rep.* Vol. 9, pp. 208–214.
- Salih, D.A. and Brunet, A. (2008), 'FoxO transcription factors in the maintenance of cellular homeostasis during aging', *Curr. Opin. Cell Biol.* Vol. 20, pp. 126–136.
- Gross, D.N., van den Heuvel, A.P. and Birnbaum, M.J. (2008), 'The role of FoxO in the regulation of metabolism', *Oncogene* Vol. 27, pp. 2320–2336.
- Nakae, J., Kitamura, T., Kitamura, Y. *et al.* (2003), 'The forkhead transcription factor Foxo1 regulates adipocyte differentiation', *Dev. Cell* Vol. 4, pp. 119–129.
- Calnan, D.R. and Brunet, A. (2008), 'The FoxO code', *Oncogene* Vol. 27, pp. 2276–2288.
- Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D. *et al.* (2002), 'Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4', *Mol. Cell* Vol. 9, pp. 279–289.
- Friedman, J.R. and Kaestner, K.H. (2006), 'The Foxa family of transcription factors in development and metabolism', *Cell Mol. Life Sci.* Vol. 63, pp. 2317–2328.
- Miyara, M., Wing, K. and Sakaguchi, S. (2009), 'Therapeutic approaches to allergy and autoimmunity based on FoxP3+ regulatory T-cell activation and expansion', *J. Allergy Clin. Immunol.* Vol. 123, pp. 749–755.
- Stroud, J.C., Wu, Y., Bates, D.L., Han, A. *et al.* (2006), 'Structure of the forkhead domain of FOXP2 bound to DNA', *Structure* Vol. 14, pp. 159–166.
- Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. *et al.* (2001), 'A forkhead-domain gene is mutated in a severe speech and language disorder', *Nature* Vol. 413, pp. 519–523.
- UniProt Consortium. (2010), 'The Universal Protein Resource (UniProt) in 2010', *Nucleic Acids Res.* Vol. 38 (Database issue), pp. D142–D148.
- Bruford, E.A., Lush, M.J., Wright, M.W., Sneddon, T.P. *et al.* (2008), 'The HGNC Database in 2008: A resource for the human genome', *Nucleic Acids Res.* Vol. 36 (Database issue), pp. D445–D448.
- Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E. *et al.* (2008), 'The Mouse Genome Database (MGD): Mouse biology and model systems', *Nucleic Acids Res.* Vol. 36 (Database issue), pp. D724–D728.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R. *et al.* (2007), 'Clustal W and Clustal X version 2.0', *Bioinformatics* Vol. 23, pp. 2947–2948.
- Fan, Y., Newman, T., Linardopoulou, E. and Trask, B.J. (2002), 'Gene content and function of the ancestral chromosome fusion site in human chromosome 2q13-2q14.1 and paralogous regions', *Genome Res.* Vol. 12, pp. 1663–1672.
- Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S. *et al.* (2008), 'Genesis and expansion of metazoan transcription factor gene classes', *Mol. Biol. Evol.* Vol. 25, pp. 980–996.
- Gambineri, E., Torgerson, T.R. and Ochs, H.D. (2003), 'Immune dysregulation, polyendocrinopathy, enteropathy, and X-linked inheritance (IPEX), a syndrome of systemic autoimmunity caused by mutations of FOXP3, a critical regulator of T-cell homeostasis', *Curr. Opin. Rheumatol.* Vol. 15, pp. 430–435.
- Koon, H.B., Ippolito, G.C., Banham, A.H. and Tucker, P.W. (2007), 'FOXp1: A potential therapeutic target in cancer', *Expert Opin. Ther. Targets* Vol. 11, pp. 955–965.
- Nakshatri, H. and Badve, S. (2007), 'FOXa1 as a therapeutic target for breast cancer', *Expert Opin. Ther. Targets* Vol. 11, pp. 507–514.
- Gartel, A.L. (2010), 'A new target for proteasome inhibitors: FoxM1', *Expert Opin. Investig. Drugs* Vol. 19, pp. 235–242.