

## The MLL family of proteins in normal development and disease

Special Issue

Guest Editor: Thomas A. Milne

The story of the Mixed Lineage Leukemia (MLL) family of proteins can be presented from many different viewpoints. One possible place to start is from early work that identified the human chromosome site 11q23, now known to be the position of the *MLL* gene, as a common translocation breakpoint in aggressive human leukemias. The cloning of this breakpoint led to the discovery of the *MLL* gene (also called *ALL-1* or *HRX*, now called *Histone-lysine N-methyltransferase 2A* or *KMT2A*) and the recognition that it was a homolog of a gene in *Drosophila melanogaster* named *trithorax* (*trx*) [1-4].

Perhaps then, it might make more sense to start the story earlier, with the discovery of the *trx* gene in *Drosophila* [5]. The trithorax protein is the founding member of the trithorax group (trxG) of proteins, famous for their role in the maintenance of *Homeobox* (*Hox*) gene expression and their resultant importance in body patterning for multicellular organisms [6]. Finding that the *MLL* gene in humans was homologous to *trx* caused much excitement, as it linked together human disease with an important regulator of normal development.

Over time, multiple members of the MLL family were identified. To reflect this, MLL is sometimes also referred to as MLL1 (KMT2A), with the other members being MLL2 (KMT2B), MLL3 (KMT2C), MLL4 (KMT2D), SETD1A (KMT2F) and SETD1B (KMT2G). Of this group, MLL1 and MLL2 are the most closely related to *trx*, MLL3 and MLL4 are instead more closely related to the *Drosophila* trithorax-related (*trr*) protein, while SETD1A and SETD1B are more closely related to *Saccharomyces cerevisiae* Set1 and *Drosophila* dSet1 [7]. Part of what makes this group a family is that they all display a high degree of conservation in a region of the protein termed the SET domain.

The SET domain is also found in proteins outside the MLL family, some of which have opposing functional roles to MLL family proteins, such as members of the Polycomb Group (PcG) of proteins. A key insight into the function of this domain was the finding that some SET domains had highly specific lysine methyltransferase activity, but the MLL SET domain appeared to be functionally inert [8]. Subsequent work showed that both the yeast Set1 protein [11-13] as well as the MLL SET domain [9, 10] were able to specifically methylate lysine 4 (K4) on histone H3, a covalent modification correlated with gene activation. Additional work over the decades has revealed a wealth of other activities and developmental roles for this diverse set of proteins and the large multiprotein complexes they are members of, as well as

highlighting potentially wide-ranging implications for human disease. This collection of reviews reflects a small part of this incredible diversity, touching on a range of important topics from normal development, to disease, to the molecular and structural details and functions of the proteins themselves. The topics covered in this special issue are briefly summarized below.

It is often said that one cannot fully understand human disease without understanding normal development as well. From this perspective, Eric Antunes and Katrin Ottersbach provide a comprehensive overview of the entire MLL family and their roles in normal hematopoiesis (pages x-x). The differential roles of each family member during normal hematopoietic development may have important implications for their different contributions (or lack thereof) to the progression of leukemia. Although the most famous member to be involved in disease is *MLL*, Andrea Kranz and Konstantinos Anastassiadis provide a welcome focus on two other key family members: SETD1A and SETD1B (pages x-x). Expanding beyond a discussion of roles in hematopoiesis, they provide an overview of Set1 function in yeast, *C. elegans*, *Drosophila*, and mammals in multiple tissues. They also explore how these two proteins contribute to other human diseases including neurological disorders, and end with a discussion about why SETD1A and SETD1B are non-redundant and individually crucial for normal development.

Better understanding the distinct developmental role for each MLL family member requires a better understanding of the molecular details of how these protein complexes function. Liang Sha, Alex Ayoub, Uhn-Soo Cho and Yali Dou provide a comprehensive structural perspective on the MLL family proteins, including recent cryo-EM data highlighting distinctive structural and functional differences between the different family members (pages x-x). Complementary to this, Kihyun Park, Jung-Ae Kim and Jaehoon Kim focus on the details of how the different MLL family members directly influence transcriptional activation, with a particular focus on how each complex uses H3K4 methylation to accomplish this (pages x-x). Importantly, they also point out that many of the key transcriptional functions of these MLL protein complexes are separate from their methyltransferase activities, especially when observed in a tissue specific context. Expanding on this idea, Hao Jiang discusses the core subunits shared between MLL complexes, their function, and how they contribute to the “non-canonical” functions of the MLL family complexes (pages x-x). In addition, he also covers the key biological roles that many of these core factors have, especially with regards to normal stem cell function as well as in disease.

A key aspect of understanding how these MLL complexes regulate genes in a living cell is understanding how they are targeted to genes where their activity is required. Amy Hughes,

Jessica Kelley and Rob Klose provide a thoughtful piece on how something as seemingly generic as binding to CpG islands can provide a level of specificity for protein complex assembly when combined with other attributes of gene regulation (pages x-x). They also discuss a possible answer to the question about the function of H3K4me3. In helping maintain transcriptional consistency, H3K4me3 at promoters could reduce the probability of aberrant stochastic alterations of gene expression, especially in rapidly dividing cells.

As already mentioned, the most famous member to be involved in human disease is *MLL* itself. In contrast to the otherwise remarkable success story of treating childhood leukemias, Siobhan Rice and Anindita Roy focus on how *MLL* rearrangements even now still cause a very poor prognosis leukemia, especially in infants (pages x-x). This is possibly due to a combination of an incomplete understanding of the underlying biology of the leukemias, combined with a still incomplete understanding of the molecular mechanisms that lead to disease.

The most common *MLL* mutation is a translocation event in leukemia that fuses the *MLL* gene in frame with a wide range of different partner proteins creating novel fusion proteins. In a very insightful and personal review, Rolf Marschalek reminds us not to forget about the reciprocal translocations, as these may have key roles in promoting leukemogenesis as well (pages x-x). As he points out, a single translocation event that produces two functional oncogenes harkens back to the “2-hit rule”, where at least two mutations are thought to be required to initiate oncogenesis.

The molecular details of fusion proteins themselves are the subject of a comprehensive review by Satoshi Takahashi and Akihiko Yokoyama (pages x-x). They provide an exhaustive survey of different fusion proteins generated by *MLL* translocations and the range of different protein complexes assembled by them. They note several atypical *MLL* fusion protein complexes, but also provide a rationale for key features that unite the activity of these different fusion proteins. Debabrata Biswas, Subham Basu and Arijit Nandy provide a different angle on this topic by focusing on the activity of the endogenous non-mutated fusion partner complexes themselves (pages x-x). They discuss how these fusion partner complexes influence normal transcription, distinguishing between their potential roles in promoting transcription initiation as opposed to elongation. Robert Slany takes a different approach to this question by exploring a specific hypothesis of *MLL* fusion protein function (pages x-x). He uses two common members (*MLL-ENL* and *MLL-AF4*) to argue that their contrasting binding profiles at gene targets are reflective of their different complex component interactions. This is suggestive of fusion protein binding patterns being dependent on genes that are already active in the cells and an *MLL* fusion protein may only

be able to “capture” the gene expression profile of whatever target cell it initiates in. This has implications for treatment, as well as patient heterogeneity.

Juerg Schwaller covers the history of using mouse models of leukemia (pages x-x). As well as providing an interesting historical perspective on the development of these models, he explores what they can tell us about the biology of the human disease, as well as the relative importance of the target cell type. He also touches on how these models have been important in the development of novel therapies. Complementary to much of what else has been explored here, Nok-Hei Mickey Wong and Eric So provide a comprehensive view of how some of this molecular information is being used to produce new therapies, some of which is progressing through the clinics right now (pages x-x). This is an ever-evolving area, and it highlights how studying a rare disease can provide possible therapies for a wider range of diseases.

No set of reviews can be entirely comprehensive for such a diverse and rapidly expanding area, but the articles in this special issue providing many keen insights into this fascinating and important family of proteins.

## References

- [1] S. Ziemann-van der Poel, N.R. McCabe, H.J. Gill, R. Espinosa, 3rd, Y. Patel, A. Harden, P. Rubinelli, S.D. Smith, M.M. LeBeau, J.D. Rowley, et al., Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias, *Proc Natl Acad Sci U S A*, 88 (1991) 10735-10739.
- [2] M. Djabali, L. Selleri, P. Parry, M. Bower, B.D. Young, G.A. Evans, A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias, *Nat Genet*, 2 (1992) 113-118.
- [3] Y. Gu, T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, E. Canaani, The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila* trithorax, to the AF-4 gene, *Cell*, 71 (1992) 701-708.
- [4] D.C. Tkachuk, S. Kohler, M.L. Cleary, Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias, *Cell*, 71 (1992) 691-700.
- [5] P. Ingham, R. Whittle, Trithorax: A New Homoeotic Mutation of *Drosophila melanogaster* Causing Transformations of Abdominal and Thoracic Imaginal Segments, *Molecular and General Genetics*, 179 (1980) 607-614.
- [6] P.W. Ingham, trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective, *Int J Dev Biol*, 42 (1998) 423-429.
- [7] N.T. Crump, T.A. Milne, Why are so many MLL lysine methyltransferases required for normal mammalian development?, *Cell Mol Life Sci*, 76 (2019) 2885-2898.
- [8] S. Rea, F. Eisenhaber, D. O'Carroll, B.D. Strahl, Z.W. Sun, M. Schmid, S. Opravil, K. Mechtler, C.P. Ponting, C.D. Allis, T. Jenuwein, Regulation of chromatin structure by site-specific histone H3 methyltransferases, *Nature*, 406 (2000) 593-599.
- [9] T.A. Milne, S.D. Briggs, H.W. Brock, M.E. Martin, D. Gibbs, C.D. Allis, J.L. Hess, MLL targets SET domain methyltransferase activity to Hox gene promoters, *Mol Cell*, 10 (2002) 1107-1117.
- [10] T. Nakamura, T. Mori, S. Tada, W. Krajewski, T. Rozovskaia, R. Wassell, G. Dubois, A. Mazo, C.M. Croce, E. Canaani, ALL-1 is a histone methyltransferase that assembles a

supercomplex of proteins involved in transcriptional regulation, *Mol Cell*, 10 (2002) 1119-1128.

[11] S.D. Briggs, M. Bryk, B.D. Strahl, W.L. Cheung, J.K. Davie, S.Y. Dent, F. Winston, C.D. Allis, Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*, *Genes Dev*, 15 (2001) 3286-3295.

[12] A. Roguev, D. Schaft, A. Shevchenko, W.W. Pijnappel, M. Wilm, R. Aasland, A.F. Stewart, The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4, *Embo J*, 20 (2001) 7137-7148.

[13] P.L. Nagy, J. Griesenbeck, R.D. Kornberg, M.L. Cleary, A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3, *Proc Natl Acad Sci U S A*, 99 (2002) 90-94.