The Molecular Machinery of Germ Line Specification

BEN EWEN-CAMPEN, EVELYN E. SCHWAGER, AND CASSANDRA G.M. EXTAVOUR*

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

SUMMARY

Germ cells occupy a unique position in animal reproduction, development, and evolution. In sexually reproducing animals, only they can produce gametes and contribute genetically to subsequent generations. Nonetheless, germ line specification during embryogenesis is conceptually the same as the specification of any somatic cell type: germ cells must activate a specific gene regulatory network in order to differentiate and go through gametogenesis. While many genes with critical roles in the germ line have been characterized with respect to expression pattern and genetic interactions, it is the molecular interactions of the relevant gene products that are ultimately responsible for germ cell differentiation. This review summarizes the current state of knowledge on the molecular functions and biochemical connections between germ line gene products. We find that homologous genes often interact physically with the same conserved molecular partners across the metazoans. We also point out cases of nonhomologous genes from different species whose gene products play analogous biological roles in the germ line. We suggest a preliminary molecular definition of an ancestral “pluripotency module” that could have been modified during metazoan evolution to become specific to the germ line.


INTRODUCTION

Across the plant and animal kingdoms, embryogenesis is that crucial developmental phase during which a single pluripotent cell, the fertilized ovum, must divide and differentiate to produce a plethora of differentiated, unipotent cell types. Sexually reproducing animals must ensure that one particularly important cell type is determined: the germ cells. These cells will be the sole progenitors of eggs and sperm in the sexually mature adult, and as such, their correct specification during embryonic development is critical for reproductive success and species survival. Germ cells and their embryonic origin have fascinated biologists for centuries, resulting in an enormous amount of primary literature on the subject (last comprehensively reviewed by Nieuwkoop and Sutasurya, 1979, 1981). The time and place when germ cells are first observed in embryogenesis, their histological and cytological characteristics, and the results of experimental manipulation of embryos on germ cell formation have been described in great detail for dozens of different species across the metazoa. All studies coincide in their observation of germ cell-specific cytoplasmic inclusions, visible under transmitted light and electron microscopy alike. Molecular studies from the last three decades have shown that this special cytoplasm, often called germ plasm, houses germ cell-specific gene products. Several excellent reviews have examined germ cell formation in specific animals (Saffman and Lasko, 1999; Raz, 2003; Strome, 2005; Hayashi et al., 2007; Saitou, 2009), the genetic mechanisms of specific germ line specification modes (Houston and King, 2000a; Strome and Lehmann, 2007), general molecular
characteristics of germ cells (Seydoux and Braun, 2006; Cinalli et al., 2008; Nakamura and Seydoux, 2008), or the function of germ cell-specific genes (Raz, 2000; Noce et al., 2001). As we strive to put biological processes into an evolutionary perspective, however, we now need to begin to consider the ancestral histories of not just germ cell-specific genes themselves, but also their molecular interactions and collective functions in the germ plasm. While it is clear that many of these genes are conserved across metazoas, it is less clear to what extent the specific molecular interactions of these mRNAs, proteins, and cellular organelles have changed or remained the same throughout evolution.

The recent molecular revisitation of classical comparative embryology, otherwise known as evolutionary developmental biology or “evo-devo,” has clarified a key paradigm that is relevant to the germ cell problem in this context. It is now possible, and moreover useful, to speak of molecular modules comprising gene regulatory networks (GRNs) (see e.g. Davidson et al., 2002). Such modules consist of a group of genes whose genetic interactions, or physical interactions of their gene products, are highly biochemically stable and thus highly conserved. The result of this genetic and molecular interaction stability is that the same batteries of genes, or modules, are found to operate in similar ways both in different organisms, and in different places and/or times during the development of a single organism (discussed in Wagner et al., 2007; Monteiro and Podila, 2009). The Notch–Delta pathway, for example, is a ligand/receptor-activated signal transduction pathway that ultimately regulates gene expression at the level of transcription. All members of this pathway are both highly conserved and operate together in all metazoans (reviewed by Kopan and Ilagan, 2009). Over the course of animal development, this module participates in a wide variety of developmental processes, including segmentation, neuroblast specification, and stem cell maintenance (reviewed by Artavanis-Tsakonas et al., 1999; Lai, 2004).

We can therefore ask, in the case of germ line-specific molecules, if it is possible to identify a group of genes that are not only highly conserved, but whose products also display conserved molecular interactions. If so, does this putative “module” also participate, like the Notch pathway, in a variety of developmental decisions, or is it confined to germ line specification? In this review, we will establish a framework for answering these questions by reviewing and summarizing recent data on the molecular functions and interactions of several genes that are critical for germ cell specification. Because examining all known genes involved in the process is beyond the scope of this review, several genes that are conserved in animal genomes, but whose role in germ line specification is either poorly understood or likely to be indirect, are indicated in Table 1, but not discussed further. Instead, we have focused on a subset of genes whose germ line specificity and critical roles in specification are well established. Some of these genes are highly conserved across the Metazoa, while for others, either their presence in the genome or their role in germ line specification, are lineage-specific.

### CONSERVED MOLECULAR COMPONENTS OF GERM LINE SPECIFICATION AND DIFFERENTIATION

#### Vasa

Products of the vasa gene family are the most widely used molecular germ cell markers for the Metazoa (discussed in Raz, 2000; Noce et al., 2001; Extavour and Akam, 2003). Vasa proteins are ATP-dependent RNA helicases of the DEAD box class, which was originally identified as a helicase family based on conservation of eight functional domains (Linder et al., 1989). DEAD box helicases are generally involved in RNA metabolism and can mediate both RNA–RNA and RNA–protein interactions (Rocak and Linder, 2004). A significant body of functional data for these helicases exists, based largely on studies of yeast DEAD box proteins (reviewed by Rocak and Linder, 2004). However, much less is known about the specific molecular function of Vasa, the germ cell-specific member of this class.

The vasa (vasa) locus was first identified in Drosophila in a screen for maternal effect genes involved in anterior–posterior axis formation (Schüpbach and Wieschaus, 1986). Drosophila Vasa protein localizes to cytoplasmic granules within pole plasm (Lasko and Ashburner, 1988; Hay et al., 1988a,b), and localization of the mRNA, protein, or both to germ plasm and germ cells at some stage of development is a universal characteristic of the vasa gene family (see e.g. Lasko and Ashburner, 1988, 1990; Hay et al., 1988a,b; Fujiwara et al., 1994; Komiya et al., 1994; Ikenishi and Tanaka, 1997; Yoon et al., 1997; Braat et al., 2000; Knaut et al., 2000; Tanaka et al., 2000; Toyooka et al., 2000; Ozhan-Kizil et al., 2009).

Localization of vasa gene products to germ plasm is consistent with its loss-of-function phenotypes in Drosophila, which are loss of or defective primordial germ cells (PGCs; also called pole cells in Drosophila) (Lasko and Ashburner, 1990), with additional oogenesis defects seen for null mutations (Styhler et al., 1998). Similarly, nematode (Gruidl et al., 1996; Kuznicki et al., 2000; Spike et al., 2008), frog (Ikenishi and Tanaka, 1997), flatworm (Ohashi et al., 2007), crustacean (Ozhan-Kizil et al., 2009), tunicate (Sunanaga et al., 2007), and mouse (Tanaka et al., 2000) vasa knockdowns or mutants show germ line defects at various stages of germ cell development, including gametogenesis. In zebrafish, however, morpholino-mediated protein knockdowns of Vasa affect neither germ cell number nor fertility (Braat et al., 2001). While vasa is almost always required for some stage of germ cell development, in no animal has it been shown to be sufficient (see, e.g., Ikenishi and Yamakita, 2003). However, a recent study (Lavial et al., 2009) has shown that experimentally induced vasa expression can reprogram chicken embryonic stem cells and direct them toward a germ cell fate. This suggests that vasa might be able to function as a germ cell determinant for cells that are already pluripotent.

Genetic interactions between vasa and other germ line genes have suggested a complex network of positive and negative regulation at multiple levels, including transcription, translation, and post-translational modification. In Caenorhabditis elegans and mice, various components of germ
cell-specific cytoplasmic aggregations such as P granules, chromatoid bodies, and nuage lose their localization in vasa mutants (Chuma et al., 2006; Hosokawa et al., 2007; Spike et al., 2008). Vasa’s identity as an RNA helicase suggests a role in translational regulation, and indeed, higher levels of some proteins in vasa mutants (Johnstone and Lasko, 2004) and a physical and genetic interaction with a translation initiation factor (Carrera et al., 2000) are both consistent with this hypothesis. However, very few direct molecular interactors have been identified for Vasa to date, and most of them effect or stabilize its localization (but see Carrera et al., 2000; Johnstone and Lasko, 2004). The SOCS-box/SPRY-domain gene gustavus was identified in a yeast two-hybrid screen that used Drosophila Vasa protein as bait (Styhler et al., 2002). Gustavus is a highly conserved protein whose zebrafish homolog localizes to germ plasm (Li et al., 2009b), suggesting an ancient origin for this protein interaction. Drosophila gustavus mutants fail to localize Vasa protein to the germ plasm, and other identified binding partners of Vasa protein also appear to play a role in localization, rather than function, of Vasa. The novel protein Oskar (discussed below) and the ubiquitin-specific protease Fat facets (Liu et al., 2003) interact physically with Vasa, and both are required for Vasa’s correct localization to germ plasm. Detailed studies of multiple vasa mutant alleles have shown that the RNA-binding domains of the Vasa protein are not necessary for its localization to the pole plasm, but are necessary for its germ cell function (Liang et al., 1994). While further work will be needed to identify the molecular partners and direct targets of Vasa’s regulatory function, it is clear that Vasa co-localizes to the germ plasm together with several other highly conserved germ cell gene products. Those for which the most functional data are available are discussed in the following sections.

**Nanos and Pumilio**

Orthologs of Nanos localize to germ cells of nearly all taxa studied (Extavour and Akam, 2003). The specific functions played by Nanos vary, but the phylogenetically widespread expression of these proteins in germ cells suggests that a germ line function of Nanos may have evolved very early in animals (Extavour and Akam, 2003; Extavour, 2007). Pumilio, which has orthologs in organisms as diverse as yeast and plants (Zamore et al., 1997), has been shown to physically interact with Nanos proteins in flies (Wharton, 1999). Pumilio orthologs vary in the germ cells of different organisms. In Drosophila and C. elegans, Pumilio directly regulates many RNAs in migrating PGCs to repress somatic identity (Kobayashi et al., 1996; Deshpande et al., 1999; Hayashi et al., 2004), halt the cell cycle (Asaoka-Taguchi et al., 1999), and prevent apoptosis (Hayashi et al., 2004; Sato et al., 2007). Similarly, in C. elegans, nos-1 and nos-2 are not necessary for the initial formation of PGCs, but rather for their maintenance and survival during embryogenesis (Subramaniam and Seydoux, 1999). In contrast, the C. elegans NOS-3/FBF complex is not involved in germ cell development, but rather in the sperm-to-oocyte transition in hermaphrodites (Kraemer et al., 1999). In zebras (Koprunner et al., 2001) and mice (Tsuda et al., 2003), nanos-related genes are required for PGC survival in both sexes, but specific targets of Nanos are largely unreported in vertebrates.

In both Drosophila and C. elegans, Nanos is also genetically implicated in the maintenance of a specific “chromatin architecture” that is associated with general transcriptional repression (Schaner et al., 2003). However, it has been pointed out that the cytoplasmic localization of Nanos protein, as well as our knowledge of its molecular function, implies that this function may be indirect (discussed in Seydoux and Braun, 2006).

Finally, there is also evidence for Nanos proteins functioning in the absence of known interactions with Pumilio.
proteins. As mentioned above, *C. elegans* NOS-3 physically interacts with the Puf protein FBF, but the two other nematode *nanos* orthologs, *nos-1* and *nos-2*, do not do so in a yeast two-hybrid assay (Kraemer et al., 1999). However, the *C. elegans* genome encodes eight Puf proteins, and knockdown of several of these proteins produces PGC defects indistinguishable from those observed in *nos-1* and *nos-2* knockdowns (Subramaniam and Seydoux, 1999), suggesting that these two *nos* orthologs could interact with other Puf proteins. In *Drosophila*, protein expression and detailed mutant analysis suggest that Nanos and Pumilio may have nonoverlapping roles in early oocyte development (Forbes and Lehmann, 1998). Therefore, while most of the studied roles of Nanos involve Pumilio-related proteins, it remains to be seen how the two may function in each other’s absence.

**Tudor**

The “grandchildless” phenotype of *tudor* mutants was first described in *Drosophila* by Boswell and Mahowald (1985). *tud* mutants do not maintain expression of germ granule components Oskar and Vasa (Thomson and Lasko, 2004), form abnormal germ granules, and ultimately fail to produce pole cells (Boswell and Mahowald, 1985; Thomson and Lasko, 2004). Proteins containing the so-called Tudor domains have since been found in organisms ranging from yeast to humans (Ponting, 1997), and Tudor proteins localize to germ granules of flies (Arkov et al., 2006), zebrafish (Mishima et al., 2006; Strasser et al., 2008), and male mice (Chuma et al., 2006; Hosokawa et al., 2007). In flies, Tudor protein also localizes between mitochondria and germ granules, and is required for transferring ribosomal RNAs from the mitochondria (Table 1) to germ granules, an essential process in germ cell specification (Amikura et al., 2001).

Insight into the molecular basis for Tudor function has come from studies of the protein and its interactors. Studies of Tudor domain proteins in humans revealed that these domains interact with methylated arginine and lysine residues of diverse protein partners, including Sm proteins of the spliceosome (Friesen and Dreyfuss, 2000; Brahms et al., 2001). Recent studies suggest that Tudor’s interactions with methylated proteins, as well as with proteins of the methylosome itself, may be required for the formation of germ granules. Localization of *Drosophila* Tudor to germ granules genetically requires the activity of the methylosome components *capsuleen* (also called *dart5*, the fly ortholog of human *dPRMT5*) and *Valois* (the fly ortholog of human *MEP50*) (Anne and Mechler, 2005; Gonsalvez et al., 2006). Further, Tudor can bind both Capsuleen and Valois in vitro, and these latter two proteins methylate Sm proteins, with which Tudor also physically interacts (Anne and Mechler, 2005). As Seydoux and Braun (2006) have pointed out, Sm proteins are common components of germ granules from vertebrates to *C. elegans*, and have been shown to be required for P granule localization and function in *C. elegans* (Barbee et al., 2002; Barbee and Evans, 2006), suggesting that the role of **Tudor** in assembling germ granules may involve its association with methylosome components and Sm proteins.

**A Link Between Tudor and PIWI-Family Proteins**

An additional role for Tudor was recently suggested by the finding that PIWI-family proteins in mice, frogs, and flies contain symmetrically methylated arginine residues of the type recognized by Tudor proteins (Kirino et al., 2009; Vagin et al., 2009). In the *Drosophila* ovary, PIWI-family proteins (discussed below) require *capsuleen*-dependent methylation in order to maintain their own expression and to maintain wild-type levels of piRNAs. In addition, *capsuleen* mutant ovaries accumulate abnormally high levels of retrotransposons that are normally silenced by PIWI-family proteins (Kirino et al., 2009). In mice, the three PIWI proteins were recently shown to directly interact with the methylosome complex of PRMT5 and WDR77/MEP50. This complex methylates arginine residues of Mili, Miwi and Miwi2, which in turn interact with Tudor domain-containing proteins. Additionally, specific PIWI and Tudor proteins also colocalize to nuage components (Vagin et al., 2009, Wang et al., 2009). Together, these results suggest a previously unrecognized connection between the interacting networks of proteins and RNAs in germ granules.

The PIWI family of proteins (called Piwi, Aubergine, and Ago3 in *Drosophila*; Zwi and Zili in zebrafish; Miwi, Miwi2, and Mili in mice; and Xiwi, Xili, and Xiwi2 in frogs) were named for a founding member (P-element-induced wimpy tests) uncovered in a screen for genes involved in maintaining germ line stem cells in the *Drosophila* ovary (Lin and Spradling, 1997). This protein family has since been intensively studied for its role in silencing retrotransposons in the germ line through interactions with a special class of small RNAs called piRNAs or rasiRNAs (reviewed by Hartig et al., 2007).

Specific roles for PIWI proteins in the specification and/or maintenance of germ cells have been suggested by mutant analyses. *Piwi* mutant flies have reduced numbers of pole cells (Megosh et al., 2006). Both Piwi and Aubergine localize to germ granules in nurse cells and pole cells, and Piwi physically interacts with Vasa until Piwi translocates to the nucleus, where it remains throughout germ cell migration and gametogenesis (Megosh et al., 2006). In zebrafish, *ziwi* RNA (Tan et al., 2002) and protein (Houwing et al., 2007) co-localize with Vasa to germ line-specific ribonucleoprotein complexes (RNP), and *ziwi* mutants are agametic owing to progressive apoptosis of germ cells (Houwing et al., 2007). In mice, mutants for *miwi*, *mili*, or *miwi2* fail to complete spermatogenesis, although these genes are not required for female germ line development (Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004; Carmell et al., 2007). Recently it has been shown that Miwi and Miwi2 also form a complex with Mvh, the mouse vasa homologue (Vagin et al., 2009).

Importantly, although the PIWI-family proteins are best known for their germ cell function, some members of the related Argonaute family localize not only to germ plasm, but also to somatic RNA processing organelles such as P bodies (see, e.g., Lin et al., 2006). P bodies and germ plasm granules may thus contain organelles with closely related roles in RNA processing in both germ line and soma. While it is clear that the germ line and somatic organelles are
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Functional gene product(s)</th>
<th>Molecular/genetic germline role demonstrated</th>
<th>Molecular function</th>
<th>Descriptive summary</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maelstrom</td>
<td>Protein</td>
<td>Dm, Mm</td>
<td>Putative protein binding</td>
<td>Nuage localization; piRNA pathway</td>
<td>Costa et al. (2006), Findley et al. (2003), Lim and Kai (2007), Soper et al. (2008), Zhang et al. (2008)</td>
</tr>
<tr>
<td>Par-1</td>
<td>Protein</td>
<td>Dm, Mm, Ce</td>
<td>Kinase</td>
<td>Cytoskeletal polarization; asymmetric cell division; Osk/PIE-1 stabilization</td>
<td>Cheecks et al. (2004), Doerflinger et al. (2003, 2006), Guo and Kemphues (1995), Reese et al. (2000), Riechmann et al. (2002), Shulman et al. (2000), Tian and Deng (2009), Vaccari and Ephrussi (2002), Zimyanin et al. (2007)</td>
</tr>
<tr>
<td>Germ cell less</td>
<td>Protein</td>
<td>Dm, Mm</td>
<td>Protein binding</td>
<td>Cell cycle and transcriptional regulation; PGC nuclear envelope localization</td>
<td>de la Luna et al. (1999), Jongens et al. (1992), Kimura et al. (1999, 2003), Leatherman et al. (2000), Li et al. (2006), Nili et al. (2001), Scholz et al. (2004)</td>
</tr>
<tr>
<td>Dead end</td>
<td>Protein</td>
<td>Mm, Dr, Xi, Gg</td>
<td>RNA binding</td>
<td>Protects germine RNAs from miRNA-based degradation; vertebrate-specific</td>
<td>Kedde et al. (2007)</td>
</tr>
<tr>
<td>Staufen</td>
<td>Protein</td>
<td>Dm, Mm, XI</td>
<td>dsRNA binding</td>
<td>osk localization; germ plasm component maintenance</td>
<td>Ephrussi et al. (1991), Irion et al. (2006), Ramasamy et al. (2006), St Johnston et al. (1991, 1992), Thomas et al. (2009), Yoon and Mowry (2004)</td>
</tr>
<tr>
<td>Dazl</td>
<td>Protein</td>
<td>Dm, Mm, Dr, Xi, Ce, Ol, Hs, Am</td>
<td>RNA binding</td>
<td>Germ plasm component</td>
<td>Anderson et al. (2007), Hashimoto et al. (2004), Houston and King (2000b), Houston et al. (1998), Johnson et al. (2001), Karashima et al. (2000), Kosaka et al. (2007), Moore et al. (2003), Reynolds et al. (2005, 2006), Saunders et al. (2003), Tung et al. (2006), Venables et al. (2001), Xu et al. (2007)</td>
</tr>
</tbody>
</table>

Am, *Ambystoma mexicanum*; Dm, *Drosophila melanogaster*; Mm, *Mus musculus*; Hm, *Homo sapiens*; Ol, *Oryzias latipes*; XI, *Xenopus laevis*; Ce, *Caenorhabditis elegans*; Gg, *Gallus gallus*; Dr, *Danio rerio*. Due to space limitations, we have focused here on those primary references that focus on the germline role of these gene products.
not identical, there is some overlap in the proteins and RNAs that they contain (Megosh et al., 2006; Gallo et al., 2008; Lykke-Andersen et al., 2008). These observations highlight the molecular similarities underpinning the functional analogies between RNA processing organelles in both germ cells and somatic cells, and are consistent with the hypothesis that RNPs are repressive regulatory organelles with an ancient eukaryotic history, predating the origin of the dedicated metazoan germline (see Eulalio et al., 2007 for a detailed review).

Silencing of transposable elements in the germ line is the most well-established role of PIWI-family proteins. In the above-listed PIWI-family mutations, germ cell failure is correlated with reduced piRNA levels and abnormal accumulation of transposable elements. This function is mediated through the interaction of PIWI-family proteins and additional factors with a special class of small RNAs that provide sequence specificity to a transcript-silencing complex (see Klattenhoff et al., 2007 for additional details). Two exciting studies have recently demonstrated a role for the piRNA pathway in silencing transposable elements in somatic cells of the gonad rather than in the germ cells themselves (Malone et al., 2009; Li et al., 2009a).

Among many other defects, PIWI family mutants also exhibit defects in maintaining the localization of essential germ granule components in Drosophila. For example, although piwi is not required for the initial expression of Oskar, Vasa, or Nanos, ectopic expression of Piwi protein is able to recruit these maternal factors and increase their expression levels (Megosh et al., 2006), suggesting that Piwi acts in a positive feedback loop with these factors. Recent studies have also shown that Vasa localization genetically requires aubergine and ago3, and that these two proteins require one another for their own localization (Li et al., 2009a). Additionally, aubergine mutant flies fail to properly localize the RNAi pathway members Krimper and Maelstrom (Lim and Kai, 2007). The mechanism by which PIWI-family proteins act to recruit and/or maintain localized expression of other factors to germ granules is unknown.

SYSTEM-SPECIFIC MOLECULAR COMPONENTS OF GERM LINE SPECIFICATION

Bmps

In contrast to organisms where germ cell determination relies on the inheritance of germ plasm (reviewed by Extavour and Akam, 2003), in the mouse this process requires inductive signals (Tam and Zhou, 1996) (Fig. 1). The first germ cell-inducing signal in mouse embryos was identified only a decade ago as Bmp4, a member of the Bone morphogenetic protein family. Prior to PGC induction in the proximal epiblast, Bmp4 is expressed in the tissue directly adjacent to the epiblast, the extraembryonic ectoderm (ExE). This expression is essential for PGC determination, as Bmp4 mutant mice do not form PGCs (Lawson et al., 1999).

**Figure 1.** Localization of germ line specification molecules throughout animal development. Developmental progression in time goes from top to bottom. Germ cell specification and the localization of molecules discussed in this review are schematized for the five genetic laboratory organisms that have come the most to our current understanding of the molecular mechanisms underlying germ cell specification. In all organisms except for mice (Mus musculus), germ cell-specific gene products (yellow), including mRNAs and/or proteins of vasa, nanos, pumilio, piwi, and tudor, are localized to the cytoplasm of germ cells either late in oogenesis or early during embryogenesis. The fruit fly (Drosophila melanogaster) protein Oskar and the zebrafish (Danio rerio) protein Bucky ball (dark blue star) can autonomously assemble many of these germ plasm components in oocytes and early embryos. The nematode (Caenorhabditis elegans) protein PIE-1 (light blue) plays an important role in regulating germ line gene expression (yellow). In mice, somatic signals (red) trigger the expression of Blimp1 (green) in PGCs, followed by the expression of conserved germ line genes (yellow).
Similar to Bmp4, Bmp8b is also expressed in the ExE, and Bmp8b mutants lack or show very reduced numbers of PGCs. However, the effects of Bmp4 and Bmp8b are not additive (Ying et al., 2000). Very recently it was shown that the role of Bmp8 is very different from that of Bmp4. Rather than being directly required for induction of PGCs, Bmp8b signaling from the ExE restrains growth of the anterior visceral endoderm (AVE), which sends still-unidentified signals to the epiblast, thereby inhibiting Bmp4 (Ohinata et al., 2009).

Unlike Bmp4 and Bmp8b, Bmp2 is predominantly expressed in the visceral endoderm (VE). Bmp2 mutants show a reduced number of PGCs, and Bmp2 and Bmp4 together have an additive effect on PGC development, while Bmp2 and Bmp8b together do not. Thus, Bmp signals from both the VE (Bmp2) and the ExE (Bmp4 and 8b) are required for PGC induction (Ying and Zhao, 2001).

Bmp ligands exert their function by binding and recruiting transmembrane type I and II Bmp serine/threonine kinase receptors on the cell surface. These receptors transmit the Bmp signal by phosphorylating Smad1, Smad5, or Smad8, which enter the nucleus as heterodimers with Smad4 and then serve as transcriptional regulators (Shi and Massagué, 2003). The precise downstream molecules that transduce Bmp signals and result in PGC formation are still largely unknown. There are only reports of three type I Bmp receptors possibly being involved in PGC determination, Alk2, Alk3, and Alk6 (de Sousa Lopes et al., 2004; Ohinata et al., 2009). Several Smads have been shown to be involved in PGC formation. Smad4, Smad1, and Smad5 mutants show reduced numbers or a complete lack of PGCs (Chang and Matzuk, 2001; Tremblay et al., 2001; Hayashi et al., 2002; Chu et al., 2004; Arnold et al., 2006). Furthermore, Smad1 and Smad5 are sufficient for PGC induction in combination with Bmp4 and Alk3 or Alk6 (Ohinata et al., 2009). However, a type II receptor involved in this process, as well as genes that are directly regulated by the Smads during PGC determination, remain to be identified.

Bmp signals have not yet been reported to be involved in germ line specification outside of mice, even though the epigenetic mode of germ cell formation has been hypothesized to be ancestral in metazoans (Extavour and Akam, 2003). More investigation will be necessary to provide functional evidence for or against this hypothesis, to determine whether the Bmp signal is conserved for germ cell induction, or whether different signaling pathways can be used by different animals to induce germ cells.

**Blimp1**

The earliest marker of mouse PGCs, Blimp1 (B lymphocyte induced maturation protein 1, also known as Prdm1) was discovered only recently (Ohinata et al., 2005). It was first described in the context of plasma cell differentiation (Turner et al., 1994). The histone methyl transferase Blimp1 contains a PR domain and a proline-rich region at the N-terminus, five C2H2 zinc fingers, and a C-terminal acidic domain (Turner et al., 1994; Tunyaplin et al., 2000). Blimp1 orthologs are found in many bilaterian animals (see e.g. de Souza et al., 1999; Tunyaplin et al., 2000; Hinman and Davidson, 2003; Ng et al., 2006; Arenas-Mena, 2008), but expression data do not suggest a role in germ cell specification outside of mammals. Among a variety of lethal defects, Blimp1 mutants exhibit only a very small number of PGC founder cells that fail to proliferate or migrate (Ohinata et al., 2005; Vincent et al., 2005; Robertson et al., 2007). These PGC-like cells do not show the wild-type pattern of Hox gene repression (Ohinata et al., 2005), and Blimp1 mutant cells fail to repress other somatic genes (Kurimoto et al., 2008). Consistent with this observation, in other cellular differentiation processes, Blimp1 has been observed to act as a transcriptional repressor that recruits a complex of Groucho-family proteins (Ren et al., 1999) as well as a histone deacetylase (Yurke et al., 2000). The molecular interactions of Blimp1 during germ cell specification are still unknown. Interestingly, Blimp1 co-immunoprecipitates with the mammalian ortholog of Drosophila Capsuleen/dart5 (Prmt5, discussed above), which is required for Tudor localization. This complex of Blimp1 and Prmt5 might play a role in germ line maintenance during PGC migration, as both proteins co-localize to the nucleus during this process. After migration they co-localize in the cytoplasm, which coincides with H2A/H4R3me2s downregulation in PGCs (Ancelin et al., 2006). As mentioned above, fly Capsuleen also binds Tudor and is necessary for its localization to the germ plasm in Drosophila. However, whether mouse Tudor-related proteins bind the Blimp1-Prmt5 in a functional complex during PGC specification remains an open question.

**Pie-1 and Polar Granule Component**

In *C. elegans*, one of the best-understood germ line determinants is the PIE-1 protein. The *C. elegans pie-1* mutant (pharyngeal and intestinal excess) was discovered in a screen as being required for the PGC precursor (Fig. 1) to follow germ line rather than somatic fate (Mello et al., 1992). The PIE-1 protein contains two CCCH zinc fingers (ZF1 and ZF2), which are separated by arginine–serine dipeptide repeats. PIE-1 is expressed maternally, is then asymmetrically distributed to the germ line blastomers, and continues to be expressed in the germ line throughout development (Mello et al., 1996).

The ZF2 domain of PIE-1 is required for the translation of NOS-2 protein from maternal nos-2 mRNA in the germ line (Tenenhaus et al., 2001). However, PIE-1’s principal role is to mediate transcriptional repression in the *C. elegans* germ line from P2 onwards (Seydoux et al., 1996). mRNA transcription requires phosphorylation of the C-terminal domain (CTD) of RNA polymerase II on Serine 5 (Ser5) for transcriptional initiation, and on Serine 2 (Ser2) for elongation (reviewed in Peterlin and Price, 2006; Saunders et al., 2006; Corden, 2007). The PIE-1 C-terminal region contains a motif that resembles the CTD, but has no phosphorylatable sites. This CTD-like motif, an HLX homology region, and additional C-terminal repeats, are sufficient for transcriptional repression (Batchelder et al., 1999). The PIE-1 CTD-like domain is thought to directly compete with the CTD for binding cyclin T (CycT), thereby inhibiting Ser2 phosphorylation of the CTD by the kinase CDK9 (Cyclin-dependent kinase 9). CycT and CDK9 together form the...
positive transcription elongation factor b (P-TEFb) (Zhang et al., 2003). The CTD-like motif of PIE-1 is essential for inhibiting Ser2 phosphorylation, but does not play a role in Ser5 phosphorylation; the latter activity is mediated by sequences around the CTD-like motif (Ghosh and Seydoux, 2008).

Pgc (polar granule component) accomplishes a similar transcriptional repression function by inhibition of RNA PolII via P-TEFb inhibition in Drosophila pole cells (Nakamura et al., 1996; Martinho et al., 2004). However, the 71-amino-acid Pgc protein does not bear any resemblance to C. elegans PIE-1, even though it interacts with P-TEFb and represses CTD Ser2 phosphorylation, thereby inhibiting its recruitment to transcription sites (Hanyu-Nakamura et al., 2008). Furthermore, pgc is essential only for germ cell migration and not for pole cell specification (Nakamura et al., 1996; Martinho et al., 2004). This is therefore an interesting case of independent evolution of two unrelated proteins that play an analogous role in the same molecular pathway. The PIE-1/Pgc relationship parallels that of two other proteins involved in germ line specification, Oskar and Bucky ball (discussed below).

oskar and bucky ball

All genes discussed thus far are either components of germ plasm (vasa, nanos, pumilio, piwi, tudor, pie-1, Blimp1), or molecules that accumulate the germ plasm components (BMPs). In the case of the germ plasm components described above, these genes are necessary but not sufficient for germ cell specification and function. The BMPs and their downstream effectors, in contrast, are both necessary and sufficient for germ cell specification, but are not themselves germ plasm components. There do exist, however, two molecules that are not only germ plasm components, but are also both necessary and sufficient for germ plasm formation. These genes, oskar and bucky ball, are lineage-restricted genes with independent, recent evolutionary histories, whose shared molecular function makes them of special interest in the context of this review.

oskar (osk) mRNA accumulates in the posterior cytoplasm during oogenesis in Drosophila (Ephrussi et al., 1991; Kim-Ha et al., 1991), and its translation is likewise confined to the posterior germ plasm (Kim-Ha et al., 1995). Loss-of-function mutants do not form germ cells (Lehmann and Nüsslein-Volhard, 1986). The sufficiency of osk for germ plasm assembly and germ cell formation was demonstrated in elegant experiments that drove osk expression in ectopic embryonic locations (Ephrussi et al., 1991). This showed that osk gene products can autonomously recruit germ plasm components, resulting in ectopic germ cells that are capable of functional gametogenesis (Ephrussi et al., 1991). In vas or tud mutants, however, ectopic osk does not lead to ectopic PGCs (Ephrussi and Lehmann, 1992), consistent with the hypothesis that the role of Osk is to recruit germ plasm components rather than to induce PGC fate directly.

osk mRNA localizes to the posterior pole during stages 8–10 of oogenesis, via a mechanism that requires Staufen (Table 1), microtubules, and the plus end-directed motor protein kinesin (Lehmann and Nüsslein-Volhard, 1986; Brendza et al., 2000; Zimyanin et al., 2008). osk translation is confined to the posterior cytoplasm both by positive regulation of localized transcripts and by negative regulation of unlocalized transcripts (Kim-Ha et al., 1995; Wilson et al., 1996; Micklem et al., 2000; Chekulaeva et al., 2006; Klattenhoff et al., 2007; Klattenhoff and Theurkauf, 2008). When osk mRNA is translated, alternative start codons in the osk message result in two isoforms of Osk protein, Short Osk and Long Osk, which have separable roles in germ plasm assembly (Markussen et al., 1995; Rongo et al., 1995; Breitwieser et al., 1996; Vanzo and Ephrussi, 2002). Both Oskar isoforms are phosphorylated by the Par-1 kinase (Table 1) (Riechmann et al., 2002), which is enriched at the posterior in an actin-dependent but microtubule-independent step during oogenesis (Doerflinger et al., 2006). Par-1-dependent phosphorylation is thought to stabilize Osk protein in the pole plasm (Riechmann et al., 2002), where it recruits Par-1 and thereby participates in a positive feedback loop that reinforces its posterior localization (Shulman et al., 2000; Zimyanin et al., 2007).

oskar’s highly upstream position in the Drosophila germ cell specification pathway stems from its ability to ectopically induce germ plasm assembly (Ephrussi et al., 1991; Ephrussi and Lehmann, 1992). Accordingly, Short Osk protein has been shown to directly interact with Staufen and Vasa proteins (Breitwieser et al., 1996), and recruits nanos mRNA (Ephrussi et al., 1991; Kim-Ha et al., 1995).

The recently described zebrafish gene bucky ball (buc) has biological properties that are remarkably similar to those of oskar. buc transcripts are localized to the vegetal pole during oogenesis, together with other germ cell-specific molecules, and Buc protein is subsequently localized to germ plasm in early cleavage stage embryos (Marlow and Mullins, 2008). Loss-of-function mutations in buc lead to defects in both anterior–posterior patterning (Marlow and Mullins, 2008) and germ cell formation, including a failure of vas, dazl, nos, and buc mRNAs, and other germ plasm organelles, to localize to germ cells (Bontems et al., 2009). Ectopic expression of buc in non-germ line cell lineages of early embryos results in supernumerary germ cells that are derived from the cells containing ectopic buc (Bontems et al., 2009). In summary, buc, like osk, appears to be both necessary and sufficient for germ plasm assembly and germ cell specification.

These two genes share another striking similarity: they are both very recently evolved and do not contain any recognizable functional domains. Both genes encode novel proteins, and while osk is restricted to the Diptera (two-winged flies), buc is restricted to the vertebrates (Bontems et al., 2009). Despite the presence of germ plasm and pole cells in several other insects, osk is not found in nonfly insect genomes (Extaivour, unpublished observations). The fact that the Diptera are the insect order furthest removed from the last common insect ancestor suggests that oskar may be a recent evolutionary innovation associated with germ cell segregation only in this derived lineage. Similarly, early determination of germ cells is observed in some nonvertebrate deuterostome lineages, but buc is not found in their genomes. Despite their evolutionary unrelatedness and
lineage restriction, their biological function is highly similar. This appears to be because their molecular interactors, all of which are conserved across metazoans, are the same in both cases: for example, both Os and Buc recruit gene products of the vasa and nanos loci to form germ plasm. More biochemical studies on the transcriptional and translational regulation of buc, and on its direct physical molecular interaction partners, will be necessary to determine the extent of the apparent similarity between the biological functions of buc and osk. Given that orthologs of buc exist across vertebrates, it will be fascinating to see whether this gene plays a role in germ cell specification of mice, whose germ cells are specified through inductive signalling rather than the cytoplasmic inheritance of germ plasm. It is conceivable, for example, that interactions between buc and germ line factors are conserved in the germ cells of mice, but that the expression of buc itself is induced through BMP signalling rather than through the localization of maternal buc.

**SUMMARY**

We have seen that many genes involved in germ cell specification are conserved across evolution, and expression studies have demonstrated some similarities in their modes of localization to germ cells (Fig. 1). Moreover, these molecules often also interact biochemically in similar ways in phylogenetically distant animals. The Nanos/Pumilio complex, the Tudor domain/PIWI family interaction, and the Tudor/spliosome component association may therefore represent evolutionarily ancient interactions. In other cases, proteins that are not homologous serve analogous functions in germ cells: PIE-1 in nematodes and Polar granule component in flies both regulate transcriptional elongation by inhibiting RNA polymerase II phosphorylation. Moreover, some of these molecules, and in some cases their molecular roles, are conserved not just in germ cells but are also found in pluripotent cells of many types, and in RNA-processing bodies of somatic cells (see e.g. Lin et al., 2006; Megosh et al., 2006; Gallo et al., 2008; Jud et al., 2008; Lykke-Andersen et al., 2008).

There is still too little biochemical information for us to be able to know how extensive a putative metazoan germ line GRN could be. However, the conserved protein–protein and protein–RNA interactions described above could represent components of an ancestral pluripotency module, which would have likely contained Tudor domain protein, PIWI family members, and a DEAD box helicase. In early multicellular animals where a pluripotent stem cell population produced gametes, those stem cells that entered into gametogenesis would have tailored this module by the addition of unique germ cell genes, such as nanos, vasa, and Aub/Ago3. With the advent of dedicated germ cells in metazoans, this specialized germ line module would have come under the control of cytoplasmic inheritance or inductive mechanisms that operated exclusively in the germ line, preventing somatic cells (including somatic stem cells) from producing gametes.

One of the predicted consequences of modularity in development is that modules themselves can remain highly conserved throughout evolution, while their upstream effectors and downstream targets can evolve independently. The germ line specification pathways fulfill this prediction: the robustness of the molecular interactions between the conserved germ line gene products links them together into a module that can be either induced by BMP signals (mouse), assembled autonomously in oocytes (nematode, frog), or possibly even nucleated by a single molecule (fruit fly, zebrafish). In fact, the mechanisms that localize germ line determinants to germ cells appear to be relatively flexible not only on an evolutionary time scale, but also even within developing individuals. A fascinating study has recently demonstrated that several germ line genes become ectopically expressed in somatic tissues of long-lived C. elegans mutants that lack insulin signalling (Curran et al., 2009). Although the restriction of germ line factors such as PIE-1 to the germ cells of C. elegans is normally achieved through asymmetric cell divisions, in this case the ectopic expression of PIE-1 is effected at the transcriptional level. Moreover, these somatically-expressed germ line factors appear to serve a crucial function, as knocking down any of these genes drastically reduces the longevity of these mutants. Such apparent flexibility in the spatial and temporal deployment of multiple functional germ line factors further supports the notion that such factors may operate as an interacting module, capable of being induced by a variety of upstream signals. Similarly, the downstream targets of germ line factors have evolved in lineage-specific ways.

Although we have been able to identify a few conserved molecular interaction motifs among germ line specification gene products, we still have far less biochemical data than we do genetic data on these mechanisms. We have knowledge of local interactions between a few pairs of molecules, but still lack information on how the entire suite of genes is linked together biochemically. To improve our definitions of the extent and limits of this modular network, many more biochemical studies, whose results are placed into evolutionary context, are needed.

In addition to molecular studies on the traditional laboratory model organisms that have provided us with the most data thus far, work on the physical interactions of germ line genes in “nonmodel” organisms will also be extremely informative and should be pursued in future. A thorough understanding of the genetic control of germ line development in any organism clearly requires adequate functional tools (Somm er, 2009). However, understanding the extent of the evolutionary conservation of biochemical interactions between germ line molecules is not dependant on a complete knowledge of the developmental genetics of germ line specification. Such studies therefore need not be confined to organisms for which functional genetic analysis tools have been established.

Finally, construction of a GRN for the germ line will require a somewhat different approach to those that have been undertaken thus far. Many of the powerful GRNs that have been constructed to model aspects of somatic differentiation rely largely on transcriptional regulation (see, e.g., Davidson et al., 2002; Loose and Patient, 2004; Koide et al., 2005; Imai et al., 2006). However, most of the molecules for which functional biochemical data are available appear to be...
involved in translational regulation and protein–protein interactions, suggesting that post-transcriptional gene regulation is particularly crucial in the germ line (reviewed in Cinalli et al., 2008; see e.g. Merritt et al., 2008). Moreover, it is becoming increasingly apparent that several other recently discovered mechanisms of gene regulation play a critical, albeit not yet well-defined, role in the germ line. These include piRNA-mediated transposon silencing, RNP formation to repress translation, and chromatin architecture-mediated gene regulation. We may therefore need new ways of building GRNs in order to create a framework for understanding the molecular network of the germ line.

ACKNOWLEDGMENTS

Our laboratory receives funding from the National Science Foundation (NSF; IOS-0817678). B.E.C. is supported by an NSF predoctoral award (1000081823). We apologize to colleagues whose work we have not been able to include due to space limitations.

REFERENCES


Anne J, Mechler BM. 2005. Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuléen. Development 132:2167–2177.


dent apoptosis to maintain the germ line in *Drosophila* embryos. Proc Natl Acad Sci USA 104:7455–7460.


Yoon YJ, Mowry KL. 2004. Xenopus Staufen is a component of a ribonucleoprotein complex containing Vg1 RNA and kinesin. Development 131:3035–3045.


