Published in final edited form as:

Birth Defects Res C Embryo Today. 2009 March ; 87(1): 43–63. doi:10.1002/bdrc.20147.

Genomic Landscape of Developing Male Germ Cells

Tin-Lap Lee,

Section on Developmental Genomics, Laboratory of Clinical Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

Alan Lap-Yin Pang,

Section on Developmental Genomics, Laboratory of Clinical Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

Owen M. Rennert, and

Section on Developmental Genomics, Laboratory of Clinical Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

Wai-Yee Chan*

Section on Developmental Genomics, Laboratory of Clinical Genomics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, Department of Pediatrics, Georgetown University College of Medicine, Washington, DC

Abstract

Spermatogenesis is a highly orchestrated developmental process by which spermatogonia develop into mature spermatozoa. This process involves many testis- or male germ cell-specific gene products whose expressions are strictly regulated. In the past decade the advent of high-throughput gene expression analytical techniques has made functional genomic studies of this process, particularly in model animals such as mice and rats, feasible and practical. These studies have just begun to reveal the complexity of the genomic landscape of the developing male germ cells. Over 50% of the mouse and rat genome are expressed during testicular development. Among transcripts present in germ cells, 40% – 60% are uncharacterized. A number of genes, and consequently their associated biological pathways, are differentially expressed at different stages of spermatogenesis. Developing male germ cells present a rich repertoire of genetic processes. Tissue-specific as well as spermatogenesis stage-specific alternative splicing of genes exemplifies the complexity of genome expression. In addition to this layer of control, discoveries of abundant presence of antisense transcripts, expressed pseudogenes, non-coding RNAs (ncRNA) including long ncRNAs, microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs), and retrogenes all point to the presence of multiple layers of expression and functional regulation in male germ cells. It is anticipated that application of systems biology approaches will further our understanding of the regulatory mechanism of spermatogenesis.†

Keywords

Transcriptome; alternative splicing; anti-sense; ncRNA; retrogene

†This article is a US Government work and, as such, is in the public domain in the United States of America.

*Correspondence to: Wai-Yee Chan, Section on Developmental Genomics, NICHD, NIH, 49 Convent Drive, MSC 4429, Bldg 49, Room 2A08, Bethesda, MD 20892-4429. chanwy@mail.nih.gov.

INTRODUCTION

Spermatogenesis is a highly orchestrated developmental process by which spermatogonia develop into mature spermatozoa. The whole process can be subdivided into three major phases, namely spermatogoniogenesis, meiosis, and spermiogenesis. Spermatogenesis begins shortly after birth, when the prospermatogonia (also known as gonocytes) cause type A spermatogonia. Type A spermatogonia can either self-renew or differentiate into type B spermatogonia, which then develop as spermatocytes and enter meiosis. During meiosis, two rounds of cell divisions, without intervening DNA replication, occur to reduce the chromosome number to haploid. The spermatids undergo extensive chromatin condensation and morphological changes to emerge as mature spermatozoa (Dym, 1994; Hecht, 1998). This complexly structured mechanism involves many testis- or male germ cell-specific gene products whose expressions are strictly regulated. On top of this, the proper development of male germ cells also depends on the testicular somatic cells such as Sertoli and Leydig cells, which provide growth factors and maintain the micro-environment necessary for the maturation and development of germ cells. As a result, knowing the change in gene expression and genetic programs in the developing germ cells would not only provide us the hints to understanding the regulatory mechanism of spermatogenesis at molecular level, but also to the identification of candidate genes for contraceptive development.

MALE GERM CELL TRANSCRIPTOME

Transcriptome Analyses of Male Germ Cells

With the advent of large-scale gene expression analytical techniques in the past decade, functional genomic studies of developmental processes such as spermatogenesis become feasible and practical. More than a dozen of gene expression profiling experiments of mammalian spermatogenesis, primarily in mice and rats, had been published within the past few years (Rockett et al., 2001; Wang et al., 2001; Fujii et al., 2002; Li et al., 2002; Sha et al., 2002; Tanaka et al., 2002; Anway et al., 2003; Pang et al., 2003; Schultz et al., 2003; Yu et al., 2003; Almstrup et al., 2004; Guo et al., 2004; Schlecht et al., 2004; Shima et al., 2004; Wu et al., 2004; Pang et al., 2006; Johnston et al., 2008). Only a limited number of human studies were reported (Sha et al., 2002; Chalmel et al., 2007). Among the rodent spermatogenic gene expression studies, the transcriptomes of whole testes (harvested from animals of different ages during the first wave of spermatogenesis) or specific types of germ cells and testicular cells isolated from the testes were examined to different extents with a diverse selection of gene expression profiling tools. For instance, differential display PCR technique had been used to study the overall gene expression changes during mouse spermatogenesis (Almstrup et al., 2004), and subtractive hybridization was adopted to enrich mouse transcripts that are preferentially expressed in adult over juvenile mouse testes (Fujii et al., 2002). A comparison of gene expression pattern between meiotic and post-meiotic male germ cells was performed using the NIA 15k cDNA microarray platform (Pang et al., 2006). In two separate studies (Schultz et al., 2003; Shima et al., 2004), the research groups utilized the same oligonucleotide gene chip array platform to carry out time-course analyses of the mouse testicular transcriptome. In the latter study, gene expression profiles of Sertoli cells, peritubular myoid cells, type A/B spermatogonia, pachytene spermatocytes and round spermatids were examined in parallel. Similar experiments were also performed by other research groups on rat testes and male germ cells (Schlecht et al., 2004; Johnston et al., 2008). Using sequencing-based Serial Analysis of Gene Expression (SAGE) technique, Wu and others reported a comprehensive analysis of the changes in transcriptomes among mouse type A spermatogonia, pachytene spermatocytes, and round spermatids (Wu et al., 2004; Lee et al., 2009).

It is understood that the different gene expression profiling platforms described above have their own advantages and shortcomings. Nevertheless, some remarkable conclusions can be

drawn from these studies which revealed the dynamic changes in the transcriptome of developing male germ cells.

Large portion of the genome is involved in spermatogenesis—One of the outcomes of these studies is the revelation of a large number of genes that are differentially expressed during spermatogenesis. It was estimated that ~50–58% of the mouse genome would be expressed during testis development from birth to adulthood (Schultz et al., 2003; Shima et al., 2004). A similar number (54%) was obtained from the study of rat spermatogenesis (Johnston et al., 2008). Meanwhile, around one-third of the mouse genome is differentially expressed during testicular development (Shima et al., 2004). Some of these genes are either male germ cell-specific or testis-predominant. Specifically, 2.3% of the rat testicular transcriptome is testis-specific (Johnston et al., 2008), and ~4% of the mouse genome is transcribed only in male germ cells (Schultz et al., 2003).

Many of the genes involved in spermatogenesis are uncharacterized—Despite the identification of the many genes that are involved in mammalian spermatogenesis, little is known about their function. This is exemplified by the percentage of uncharacterized/novel transcripts identified in the different studies. For instance, ~60% of the germ cell-enriched transcripts detected in the study by Schultz et al. (2003) represent uncharacterized transcripts. Half of the differentially expressed genes between meiotic and post-meiotic male germ cells are uncharacterized (Pang et al., 2006), and more than 40% of the germ cell transcripts identified by SAGE analysis (Wu et al., 2004) are uncharacterized.

Tri-phasic change in male germ cell transcriptome during spermatogenesis—Major peak expression of testicular transcripts was found to occur in three different phases which correspond to the mitotic phase (0–8 days after birth when spermatogonial proliferation predominates), the initiation of meiosis (14 days after birth when early pachytene spermatocytes appear) and the entry into spermiogenesis (20 days after birth when round spermatids first appear) (Wrobel and Primig, 2005). Interestingly, there is an elevation of transcript abundance in meiotic and post-meiotic male germ cells. The number of unique genes expressed in these cells is also significantly higher than that in spermatogonia (Shima et al., 2004). In contrast, many unique spermatogonial genes are down-regulated as male germ cells enter meiosis (Shima et al., 2004; Wu et al., 2004). Similar observation was made in a microarray validation experiment (Pang et al., 2006), in which the authors found that 80% of the genes showing differential expression between meiotic and post-meiotic male germ cells are absent or expressed at relatively low level in type A spermatogonia.

Transcript over-expression (as evaluated by the measurement of polyadenylated RNA level) in meiotic and post-meiotic male germ cells, with respect to other testicular cells, was first evident in an early study with rats (Morales and Hecht, 1994). Such phenomenon may be a bystander effect resulting from the open chromatin structure in the specified cell types, which leads to an overall activation of the transcriptional machinery (Kleene, 2001). On the other hand, it may be a mechanistic outcome to maintain transcript availability in response to the cessation of gene transcription due to chromatin condensation during spermiogenesis (Sassone-Corsi, 2002). The expression of increased number of unique genes may imply a concomitant increase in the demand of specific gene activities for the initiation and maintenance of meiosis-related events as well as the preparation for spermatozoon formation.

Another interesting phenomenon is that most of the transcripts that are first expressed during or after meiosis tend to be testis- or male germ cell-specific (Schultz et al., 2003; Chalmel et al., 2007). On the contrary, most of the genes active in spermatogonia (and Sertoli cells) are also expressed in non-reproductive tissues (Chalmel et al., 2007).

Altered prevalence of biological processes during spermatogenesis—The profiling of gene expression at genomic scale allows us to examine the biochemical characteristics of a specific type of cell or tissue at an unprecedented level. This can be achieved by applying gene ontology analyses on the preferentially expressed genes to identify the key biological processes they are associated with. Such biological process representation analysis had been adopted in several of the male germ cell transcriptome studies (Wrobel and Primig, 2005; Lee et al., 2006; Pang et al., 2006; Chalmel et al., 2007). It was noticed that certain categories of biological processes are distinctively associated with mitotic, meiotic and post-meiotic male germ cells, respectively. For instance, biological processes such as integrin signaling, ribosome biogenesis and assembly, carbohydrate metabolism, protein biosynthesis, RNA processing, cell cycle, DNA replication, chromosome organization and biogenesis, and germ cell development are preferentially associated with type A spermatogonia. Surprisingly, genes involved in embryonic development and gastrulation are also found to be prevalent in these cells. On the other hand, biological processes pertaining to reproduction and spermatogenesis are commonly seen in meiotic and post-meiotic male germ cells. Biological processes such as meiotic cell cycle, chromatin structure and dynamics, chromosome segregation, cytoskeleton, and protein degradation by ubiquitin cycle are over-represented in pachytene spermatocytes. Meanwhile, genes involved in protein turnover, signal transduction, energy metabolism, intracellular transport, ubiquitin cycle, proteolysis, peptidolysis, and fertilization are more prevalent in round spermatids.

Switch of preference of chromosomal linkage of genes from X to autosomes during spermatogenesis—It has been observed that genes displaying higher expression level in testis (notably genes related to male meiotic and post-meiotic functions) are often located in the autosomes (Khil et al., 2004). In contrast, genes expressed at earlier stages of spermatogenesis are abundant on the X chromosome (Wang et al., 2001; Khil et al., 2004). This observation was further verified by a recent cross-species male germ cell transcriptome study (Chalmel et al., 2007). The authors noticed an enrichment of loci in the rodent mitotic and somatic expression clusters on X chromosome, but no X-linkage was observed for mouse and rat genes that showed peak expression in male meiotic germ cells. The same is true for the human homologs of mouse meiotic genes: none of them are found on the X chromosome. A similar phenomenon is also observed in a particular subset of genes, the X chromosome-derived autosomal retrogenes and their X-linked progenitor genes. Although not all testis-specific autosomal genes are X-derived retrogenes or retrogenes at all, the absence of X-linkage in general is believed to be the consequence of the selective force imposed by meiotic sex chromosome inactivation (MSCI) which will be further discussed in the following section.

Conservation of male meiotic transcriptome between human and rodents—A cross-species comparison of testicular and male germ cell transcriptomes revealed hundreds of genes which show concordant meiotic and post-meiotic expression profiles among human, mouse and rat, implying the existence of “conserved” transcriptome of mammalian spermatogenesis (Wrobel and Primig, 2005; Chalmel et al., 2007). Such findings suggest that the use of rodent models in the study of spermatogenesis is a relevant means in the elucidation of human spermatogenic process and defects. The similar differential expression pattern of the testicular genes across species also suggests the presence of similar regulatory mechanism in the control of their transcription.

The analyses of mammalian testicular/male germ cell transcriptomes provide us an affluent source of information for the characterization of gene functions in male germ cell development. Besides the changes in expression patterns of protein-encoding genes, testis is also known to be a hub of non-coding transcripts such as antisense transcript as well as small and non-coding RNAs, which are implicated and recognized to play an important role in mammalian testis development (Amaral and Mattick, 2008; Hayashi et al., 2008). Also, lots of male germ cell-

specific transcripts undergo alternative splicing or are derived from sex-linked progenitor genes through retroposition to generate testis-specific isoforms of gene products to cope with the specific needs in the spermatogenic process. With the availability of the transcriptome data, the expression of, and the generation of, alternative products from the dynamic male germ cell transcriptome can now be better comprehended.

PROMINENT GENETIC PROCESSES IN MALE GERM CELLS

Alternative Splicing in Male Germ Cells

One of the surprising findings of the Human Genome Project is that the number of expressed transcripts is higher than the number of genes predicted based on the genomic sequence. For example, the estimated number of expressed transcripts in humans is 150,000 compared to 32,000 predicted based on the human genome sequence (Ben-Dov et al., 2008). This disparity suggests there may be underlying mechanisms leading to the production of gene complexity and diverse proteome. Such phenomena could be achieved by multiple transcription start site (Quelle et al., 1995), alternative pre-mRNA splicing (Ast, 2004), pre-mRNA editing (Keegan et al., 2001) and post-translational protein modifications (Banks et al., 2000). Among these mechanisms, alternative splicing is currently seen to be a major mechanism contributing to transcriptome diversification. This is achieved by splicing out introns in eukaryotic pre-mRNAs in various ways resulting in several different mRNAs and protein products from a single gene as illustrated in Figure 1. A recent genome-wide analysis shows that up to 60% of human genes have alternatively spliced forms, suggesting that alternative splicing is one of the most significant components of the functional complexity of the human genome (Maniatis and Tasic, 2002; Xu et al., 2002a; Boue et al., 2003).

The testis is a rich source for identifying gene regulation mechanisms because germ cell expansion and differentiation requires many cellular changes and regulatory steps (Yeo et al., 2004). Recent genome-wide analysis of expressed sequence tags (ESTs) showed that the testis has the greatest enrichment of tissue-specific splicing (Xu et al., 2002b). The multiple specialization events by a limited transcriptome during spermatogenesis provide a good model for studying alternative splicing and the identification of novel gene variants in germ cell development. In developing germ cells, the lengths of mRNA transcripts often vary as the cells mature (Eddy, 2002). Perhaps the best known example of alternative splicing in male cell development is cAMP-responsive element modulator (*CREM*) (Foulkes et al., 1992; Sanborn et al., 1997; Behr and Weinbauer, 2001; Blocher et al., 2005; Jaspers et al., 2007). The *CREM* regulatory pathway involved in the regulation of spermatogenesis illustrates the power and versatility of alternative splicing. *CREM* belongs to CREB (cAMP Responsive Element Binding protein) family of transcription factors. After binding to the promoter, *CREM* either activates or represses transcription (Foulkes et al., 1992; Delmas et al., 1993). When *CREM* acts as a transcriptional activator, it leads to gene transcription driven by cAMP-responsive elements in the promoter region. *CREM* is expressed primarily in the testes of males and the activity is controlled at the level of RNA processing. During spermatogenesis *CREM* pre-mRNA is alternatively processed by invoking different exon combinations and polyadenylation site selection.

The molecular basis of how alternative splice sites are selected is still largely unclear. Two general models for control of alternative splicing have been suggested. The first model envisions that alternative splicing events are controlled by specific alternative splicing factors. An example of this model include sex-lethal (*Sxl*) in *Drosophila melanogaster* and *NOVA-1*, a human neuron specific alternative splicing factor (Nagoshi et al., 1988; Jensen et al., 2000)). There are only small number of testis-specific splicing factors identified so far, such as cytotoxic granule-associated RNA binding protein-like 1 (*TIARI*), RNA binding motif protein, Y-linked family members (*RBMY*), fus-like protein (*FUS*). During spermatogenesis

different mechanisms have been proposed in the regulation of alternative splicing. For instance, *TIARI* is implicated in splice site control in spermatogonia and known to be required for germ cell survival (Forch et al., 2002), *RBMY* interacts with serine-arginine-rich protein (SR protein) to affect splicing and is specifically expressed in the nuclei of spermatogonia and spermatocytes (Venables et al., 2000); while male *FUS* knock-out mice are infertile and have defective synapsis in prophase of meiosis I (Kuroda et al., 2000). A partial list of transcripts known to be alternative spliced during male germ cell development is shown in Table 1.

Given that alternative RNA splicing is observed in a high proportion of genes and involved in the regulation of spermatogenesis, it is not surprising that alterations in RNA splicing can cause or be modified in diseases. It has been estimated that splicing defects account for nearly 15% of disease-causing mutations in humans, most of which can be attributed to point mutations in the 5' and 3' splice sites (Krawczak et al., 1992). Unusual splice isoforms have also been described in various cancers. Thus, characterization of specific splicing isoforms and the associated underlying regulations at different germ cell stages during spermatogenesis may provide new therapeutic targets and diagnostics for some male reproductive disorders.

Genome-wide identification of alternatively spliced variants—Despite the enormous potential to generate innumerable splice isoforms during germ cell development, only a small fraction of the variant have been observed due to the limitation of conventional experimental techniques. With the completion of the sequencing projects and full length cDNA libraries, global identification of alternative splicing becomes possible. Alignment of ESTs or full length cDNAs to genomic regions creates various splicing gene models by assessing splice-site consensus sequences and comparing transcripts originating from the same genomic location (Xu et al., 2002b; Carninci et al., 2005). The presence of genome sequences and large collections of transcripts provides a rich source for identifying alternative splicing events by computational methods (Kim and Lee, 2008). Tools have been recently developed that can automate much of this process (Noh et al., 2006; Birzele et al., 2008; Castrignano et al., 2008). Although computational prediction provides a systematic way to identify alternative splicing, it also has its limitations, such as quality of ESTs, insufficient numbers of transcript sequences that yields poor gene coverage and under-representation of alternatively spliced isoforms, expression biases that affect abundance, and the inability to distinguish cell type-specific alternative splicing from tissue-specific alternative splicing. Therefore, to validate and identify expressed alternative splicing on genome-wide scale reliably, high throughput expression experiment approaches should be applied. The use of exon-microarrays or sequencing-based method like SAGE allow unbiased interrogation of alternative splicing events without prior knowledge of the transcripts (Frey et al., 2005; Chan et al., 2006a; Cuperlovic-Culf et al., 2006; Calarco et al., 2007; Ruzanov et al., 2007). The identification of splice variants of *Crem* in the germ cell SAGE libraries demonstrates that SAGE offers an efficient approach to investigate the existence of alternatively processed transcripts (Wu et al., 2004). In this section, we will concentrate on the application of SAGE to identify alternative splicing events in male germ cell development.

SAGE provides a high-throughput means of generating tissue specific gene expression information that is not prone to potential orientation errors and sequence contaminations inherent in EST data. One of the compounding features is that SAGE shows extreme sensitivity for identifying 3' end transcript population heterogeneity. Alternative splicing is often regulated in a temporal or tissue-specific fashion giving rise to different protein isoforms in different tissues or at different developmental stages (Matter et al., 2002). Specific splicing of a gene at different stages of germ cell development can be identified by mapping and comparing the SAGE tags of a gene to the genomic sequence.

On the basis of the male germ cell SAGE libraries (Wu et al., 2004; Lee et al., 2009), a total of 29,654 SAGE tags were mapped to at least one germ cell stage. Two key steps were applied to find alternative splicing variants of a gene. The first step was to find tags that are not expressed in all stages. i.e., stage-specific. This was followed by tag-to-gene mapping by aligning the tag sequence to the genomic sequence for Unigene cluster retrieval and comparing the location of tags to different alternative splicing models. To screen potential tags representing alternative transcripts at different germ cell stages, a Venn diagram based on tags adjacent to the most 3' end of Nla III site followed by polyadenylation signal expression was first constructed to identify stage-specific tag expression (Sequence to Tag conversion). Stage-specific tags from the Venn diagram were then mapped to corresponding Unigene clusters based on SAGEmap available at NCBI (<http://www.ncbi.nlm.nih.gov/SAGE>), which provides tag-to-gene assignment information. To increase the reliability of subsequent mapping, tags with more or equal to three counts were considered. Genes mapped to any unknown or hypothetical transcripts, partial cDNA, image and EST clones were eliminated. Tags mapped to the same gene at different germ cell stages were then regrouped. Alternative splicing transcripts due to 3' end variations were predicted by aligning the tag sequences to the alternative splicing models based on ECgenes, which combines genome based EST clustering instead of pair-wise alignment of ESTs, and transcript assembly sub-clustering procedures (Kim et al., 2004; Lee et al., 2007). Figure 2 shows the flow of the process for the identification of alternative splice variants.

The number of genes with 3' end alternative splicing variants (3' alternative splicing) expressing in type A spermatogonia, pachytene spermatocytes and round spermatids is 74, 58, and 62 (Chan et al., 2006a). Two hundred and seven genes with 3' alternative splicing expressed in both type A spermatogonia and pachytene spermatocytes, 249 in both type A spermatogonia and round spermatids and 158 in both pachytene spermatocytes and round spermatids. There are 73 genes with different 3' alternative splicing in all three stages. Novel variants involved in developmental and transcriptional controls are identified, such as heat shock protein 4 (*Hspa4*), H3 histone, family 3B (*H3f3b*) and ubiquitin protein ligase E3A (*Ube3a*). Taken together, SAGE not only provides a rapid global survey of gene expression profile in the germ cell transcriptome, but also allows the identification of novel alternative splicing variants that contribute to the progression of spermatogenesis. Further functional study on the variants will provide new insights in germ cell development during spermatogenesis.

Antisense Transcription in Male Germ Cell Development

It is well-known that global inhibition of the initiation of mRNA translation occurs in pachytene spermatocytes and round spermatids (Kleene, 2001, 2003). In fact, translation repression is one of the characteristic features of mammalian spermatogenesis (Kleene, 2001). However, little is known about the mechanisms of translational regulation in spermatogenic cells. Antisense transcription may provide one explanation for this phenomenon.

In spite of the fact that many reports documented the occurrence of antisense transcripts (Knee et al., 1997; Korneev et al., 1999; Hastings et al., 2000; Nemes et al., 2000; Li et al., 2002; Kiyosawa et al., 2003; Runkel et al., 2003; Spiess et al., 2003; Vu et al., 2003; Hernandez et al., 2004; Ihalmio et al., 2004; Robb et al., 2004; Solda et al., 2005; Korneev et al., 2008; Orfanelli et al., 2008; Seim et al., 2008), the widespread occurrence of antisense transcripts in humans and mice has only been known in the last decade. Computational analyses estimate 8 to 20% of human and mouse genes form sense-antisense transcript pairs (Fahey et al., 2002; Lehner et al., 2002; Okazaki et al., 2002; Shendure and Church, 2002; Carmichael, 2003; Yelin et al., 2003; Chen et al., 2004; Kampa et al., 2004; Rosok and Sioud, 2004). A study of 10 human chromosomes indicated about 61% of surveyed loci had antisense transcripts (Cheng et al., 2005). A more recent study of five different human cell types found evidence for antisense

transcripts in 2900 to 6400 genes (He et al., 2008). Previous efforts to delineate mouse spermatogenic cell gene expression either only focused on a few genes or used the microarray platform which does not permit detection of antisense transcripts (Rockett et al., 2001; Wang et al., 2001; Fujii et al., 2002; Li et al., 2002; Sha et al., 2002; Tanaka et al., 2002; Anway et al., 2003; Pang et al., 2003; Schultz et al., 2003; Yu et al., 2003; Almstrup et al., 2004; Guo et al., 2004; Schlecht et al., 2004; Shima et al., 2004; Pang et al., 2006; Johnston et al., 2008). An antisense transcript, SPEER-2, was observed in late pachytene spermatocytes and early round spermatids in the mouse (Spiess et al., 2003). A systematic search for antisense transcripts in developing spermatogenic cells was only reported recently (Wu et al., 2004; Chan et al., 2006b; Chan et al., 2007).

Structural complexity of antisense transcripts identified in germ cells—Using orientation specific reverse transcription-polymerase chain reaction (RT-PCR), 19 of 64 (~31%) genes found to be differentially expressed at different stages of spermatogenesis were shown to have antisense transcripts (Wu et al., 2004; Chan et al., 2006b; Chan et al., 2007). The antisense amplicon of these 19 genes can be divided into three main groups based on the comparison with their sense genes (Chan et al., 2006b) (Fig. 3). In Group 1, the antisense amplicons were 100% complementary to the sense transcripts. This group could be divided into two subgroups; Subgroup 1A genes, which include a disintegrin and metalloprotease domain 5 (*Adam5*), diazepam binding inhibitor-like 5 (*DbiL5*), *DnaJ* (*Hsp40*) homolog, subfamily B, member 3 (*DnajB3*), four and a half LIM domains 4 (*Fhl4*), glucokinase activity related sequence 2 (*Gk-rs2*), phosducin-like 2 (*Pdcl2*), peptidylprolyl isomerase C (*Ppic*), isoform of protein phosphatase 1, catalytic subunit (*Ppp1cc*), protamine 2 (*Prm2*), sperm autoantigenic protein 17 (*Sap17*), associated molecule with the SH3 domain of STAM (*Sh3-Stam*), and testis specific gene 1 (*Tsg1*), cause antisense amplicons which are 100% complementary to a single exon of the sense gene. Subgroup 1B gene antisense amplicons are 100% complementary to the sense transcripts which are comprised of more than one spliced exon (two exons for t-complex-associated testis expressed 3 [*Tcte3*] and three exons for sperm specific lactate dehydrogenase 3, C chain [*Ldh3c*]). The antisense transcripts of these genes are short (< 300 bp) and cannot be extended beyond the exons identified in the antisense amplicons. Molecular cloning of the antisense transcripts of *Tcte3* yielded two transcripts with variable length, both transcripts lack a poly(A) tail. Protamine 1 (*Prm1*) is the sole member of Group 2; its antisense amplicon is complementary to the exons as well as intron of the sense gene. Molecular cloning identified three antisense transcripts spanning the neighboring *Prm1* and *Prm2* loci, with sequences overlapping with exonic, intronic, and intergenic sequences of the sense genes. Antisense amplicons of Group 3 genes, which include calmodulin 2 (*Calm-2*), heat shock 10 kD protein 1 (chaperonin 10) (*Ch10*), ubiquitin A-52 residue ribosomal protein fusion product 1 (*Uba52*), and ubiquitin B (*Ubb*), are not complementary to the sense transcripts. Instead, they are complementary to the pseudogenes on a different chromosome. Thus, in germ cells, anti-sense transcripts can be derived from a wide variety of origins, including processed sense transcripts, intronic and exonic sequences of a single gene or multiple genes, intergenic sequences, and pseudogenes.

Mechanisms of antisense transcription—Analyses of the spermatogenic antisense transcripts reveal a variety of potential mechanisms by which antisense transcripts are generated (Chan et al., 2006b, 2007). Total complementarity of the antisense amplicons in Group 1B to the mature mRNA suggests that they are synthesized after the sense transcripts were processed. Similar observations have been made previously with antisense transcript of cardiac troponin 1 (Podlowski et al., 2002; Bartsch et al., 2004), hemoglobin β (Volloch et al., 1996; Bonafoux et al., 2004), and rat urocortin (Shi et al., 2000). It has been postulated that these antisense transcripts are transcribed from the sense mRNA in the cytoplasm by RNA-dependent RNA polymerase (RdRP) activity (Bartsch et al., 2004; Rosok and Sioud, 2004;

Cheng et al., 2005). Even though a cytoplasmic, mitochondrial bound RdRP had been reported to be partially purified from rabbit reticulocyte lysates (Downey et al., 1973) no subsequent purification of the enzyme protein was reported. The presence of RdRP in eukaryotes remains an unconfirmed observation.

Group 2 antisense transcripts are apparently the products of post-transcription splicing of a larger antisense transcript which is processed by a splicing mechanism similar to that of the sense transcripts (Mount, 1982; Burset et al., 2000). Similar observation was reported in a more recent study that showed about ~1% of antisense fragments in two cell lines examined exhibit splicing (He et al., 2008). Several antisense transcripts have also been reported to arise in a similar manner previously (Knee et al., 1997; Hastings et al., 2000; Nemes et al., 2000; Li et al., 2002; Vu et al., 2003; Hernandez et al., 2004; Robb et al., 2004). The larger antisense transcripts spanning the *Prm1* and *Prm2* loci are comprised of different exonic, intronic, and intergenic sequences of sense genes. Similar to sense transcripts, these antisense transcripts appear to have arisen through alternative splicing, a phenomenon previously demonstrated in other antisense transcripts (Knee et al., 1997; Li and Murphy, 2000; Li et al., 2002; Alfano et al., 2005).

Pseudogenes can be another source of antisense transcripts as illustrated by the Group 3 antisense amplicons. The pseudogenes that cause antisense transcripts of the Group 3 genes are present in the intron of actively transcribed genes (the host genes) (Chan et al., 2007). Apparently, these pseudogenes were retrotransposed and inserted into the intron of the host genes. The direction of transcription of the functional “parent” gene is opposite to that of the host gene. The orientation of the pseudogene is the same as the parent gene. Anti-sense amplicons of the pseudogenes were apparently derived from spliced introns of the transcript of the host genes (see Fig. 4). The *Uba52* pseudogene resides in the intron of chromobox homolog 1 (*Cbx1*); *Calm2* pseudogene is in the intron of protein kinase, cAMP regulatory subunit beta (*Prkar2b*); *Ch10* pseudogene is in the intron of trans-activating transcription factor 3 (*Sp3*); and *Ubb* pseudogene resides in the intron of cation channel sperm associated 2 (*Catsper2*). The expression pattern of the parent gene and “host” gene during spermatogenesis are comparable, with the exception of *Prkar2b* and *Calm2*, where *Prkar2b* is only expressed in type A spermatogonia while *Calm2* is mainly expressed in pachytene spermatocytes. Different temporal expression of these two genes may imply no interaction between them. *Sp3*, the host gene of the *Ch10* pseudogene, is also only expressed in type A spermatogonia. Expression of *Cbx1* falls with differentiation while *Catsper2* is preferentially expressed in spermatocytes and spermatids (unpublished observations, Chan, W. Y. and Wu, S.M.). In contrast to the relative abundance of this type of antisense transcript in germ cells, there is only one reported example of pseudogene-derived antisense transcript in somatic cells. The antisense transcript of neural nitric oxide synthase (*nNOS*) was reported to be transcribed from a pseudogene and regulate the synthesis of nNOS (Korneev et al., 1999). Judging from the proportion of antisense transcripts derived from pseudogenes identified in the male germ cells (Chan et al., 2006b, 2007) and the number of pseudogenes identified in the mouse and human genome (Zhang et al., 2003; Zhang and Gerstein, 2004; Harrison et al., 2005; Khelifi et al., 2005), pseudogenes may be a rich source of antisense transcripts.

Expression of sense and antisense transcripts—Expression studies of antisense transcripts in male germ cells showed that the sense and the antisense transcripts can be regulated independently (Chan et al., 2006b, 2007). Testicular level of the sense transcripts is higher than that of the antisense transcripts in all cases while the relative expression in non-testicular tissues is variable implying tissue specific regulation of antisense transcription. Similar observations had been reported for other genes (Katayama et al., 2005; Nishida et al., 2005; Korneev et al., 2008), particularly imprinted genes (Li et al., 2002; Yamasaki et al., 2003; Lavorgna et al., 2004; Landers et al., 2005; Berteaux et al., 2008). Stage-specific

expression of antisense transcripts has also been observed. A recent report showed that 41 genes have spermatogonia-specific antisense transcripts, 29 genes have spermatocyte-specific antisense transcripts and 17 genes have spermatid-specific antisense transcripts (Chan et al., 2007). Thus, the regulation of anti-sense transcription may be dependent on the stage of development during spermatogenesis.

Biological functions of antisense transcripts—There is increasing evidence that antisense transcripts are intentional transcripts and not the products of leaky transcription of the non-coding strand (Dahary et al., 2005; He et al., 2008; Morris et al., 2008). A role for anti-sense transcription in disease processes has also been suggested (Tufarelli et al., 2003; Reis et al., 2004, 2005; Seitz et al., 2005; Orfanelli et al., 2008).

Several the antisense transcripts described in male germ cells, including those derived from pseudogenes, have open reading frames (Chan et al., 2006b). Even though whether they encode any functional proteins is not clear at present, one could not preclude the possibility that they encode functional proteins similar to protein-encoding antisense transcripts previously reported (Knee et al., 1997; Hastings et al., 2000; Runkel et al., 2003; Spiess et al., 2003). The significance of the majority of antisense transcripts is currently unknown though several biological functions, including transcriptional and translational regulation, have been proposed (Blin-Wakkach et al., 2001; Kiyosawa and Abe, 2002; Ogawa and Lee, 2002; Shendure and Church, 2002; Imamura et al., 2004; Lavorgna et al., 2004; Mattick, 2004; Shibata and Wutz, 2008). Among the proposed functions is the antisense transcript mediated transcriptional silencing through epigenetic modifications at the chromatin level (Weinberg et al., 2006; Morris et al., 2008; Pandey et al., 2008). Antisense transcripts have been suggested to play important roles in the regulation of monoallelic expression in X chromosome inactivation and genomic imprinting (Kiyosawa and Abe, 2002; Ogawa and Lee, 2002; Brown and Chow, 2003; Ogawa et al., 2008). Antisense transcripts derived from pseudogenes suggest a novel function for these otherwise functionless genes in the genome. Aside from potential protein coding capability of the antisense transcripts, they may serve to regulate translation of the sense mRNA or trigger RNAi via forming RNA duplex with its own sense transcript or with the sense transcript of the functional gene as illustrated in Figure 5. Identification of anti-sense transcripts derived from pseudogenes embedded in the intron of anti-parallel genes also suggests a novel function for introns. These introns may serve the role of messenger mediating the interaction between the anti-parallel gene pair.

Antisense transcription may play an important role in regulating gene expression during spermatogenesis. It is tempting to speculate that this vigorous control of gene expression is necessary to ensure the accuracy or to facilitate the functioning of the biological processes occurring during spermatogenesis such as genome-wide methylation–demethylation, genomic imprinting, monoallelic gene expression, etc. The complex nature of the antisense transcripts suggests that they are well suited to provide an additional layer of vigorous regulation of gene expression during male germ cell development.

Non-Coding RNA (ncRNAs) in Male Germ Cell Development

Past studies largely focused on protein-coding genes, which account for only around 5 to 10% of the human genome (Waterston et al., 2002; Lunter et al., 2006). However, recent large-scale transcriptome studies by whole genome tiling arrays (Johnson et al., 2005; Willingham and Gingeras, 2006), massively parallel signature sequencing (MPSS) (Jin et al., 2008), cap analysis of gene expression (CAGE) on the 5' end of transcripts (Yasuda and Hayashizaki, 2008), SAGE on the 3' end of polyadenylated transcripts (Chan et al., 2006a, b), and high-throughput cDNA sequencing by the FANTOM project (Bono et al., 2002) show that the number of expressed sequences is far more than annotated genes based on genomic sequences.

A large number of transcripts do not appear to encode for proteins. It is increasingly recognized that in addition to protein-encoding mRNA, genome transcription generates a large number of non-coding RNAs (ncRNAs) (Washietl et al., 2005). Years ago the non-coding sequences in the genome were referred to as “junk” and RNAs that do not encode protein are often considered as results of “leakage” of the transcription machinery. However, present research has highlighted that ncRNAs can have a wide range of functions and can be divided into different classes by sizes or functions (see Fig. 6). These include relatively well-known long members such as transfer RNA (tRNA) and ribosomal RNA (rRNA). The small members of ncRNAs are relatively recent discoveries. The family of ncRNAs has grown substantially and has witnessed a spectacular expansion in their roles in an array of biological processes including developmental timing, metabolism, cell cycle progression, gene silencing, and programmed cell death. Recent findings also suggested some ncRNAs participate in epigenetic regulation such as genomic imprinting, DNA methylation, and histone modification (Rinn et al., 2007; Wutz and Gribnau, 2007; Royo and Cavaille, 2008). The number of species and functions of ncRNA keeps expanding and therefore deciphering the roles of these non-coding RNA genes has emerged as one of the hottest topics in molecular biology today. This section will focus on the biological importance of different ncRNA species in male germ cell development.

Small ncRNA

MicroRNAs: MicroRNAs (miRNAs) are small ncRNAs. They are 21 to 23 nt in length, and are generated by a multi-step process in which Dicer, the RNaseIII-containing enzyme, catalyzes the precursors of miRNA to form mature miRNAs (Carmell and Hannon, 2004; Kim, 2005). It is then incorporated into the effector RNA-induced silencing complex (RISC) composed of Argonaute (Ago) proteins (Liu et al., 2004). Depending upon degree of complementarity with their RNA target, miRNAs silence gene expression either by repressing translation or by triggering mRNA degradation in RNA interference (RNAi) (Bartel, 2004). Although over 420 miRNAs have been identified in mouse so far, the biological role is still poorly understood. Despite their relatively recent discovery, it is already clear that miRNAs play important roles in male germ cell development. It has been shown that miR-17–92 and miR-290–295 clusters are highly expressed throughout development of primordial germ cells (PGCs)/spermatogonia and are required for cell cycle progression in PGC (Hayashi et al., 2008). The conditional germ-cell specific Dicer-knockout mice revealed that PGC and spermatogonia exhibit poor proliferation with suppressed retrotransposon activity in PGC but not in spermatogonia. Spermatogenesis was retarded at an early stage of proliferation and differentiation (Mallardo et al., 2008). Dicer is not only important in spermatogonial stem cell stage but also appears to be expressed in all stages of the seminiferous epithelial cycle. It interacts with a germ cell-specific chromatoid body component, the RNA helicase MVH (Kotaja et al., 2006). Deletion of *Dicer1* in PGC results in similar observations; the mutant animals are subfertile and exhibit spermatogenic defects (Maatouk et al., 2008). Though not directly demonstrated, it is tempting to speculate these effects of Dicer ablation are consequential of the disruption of miRNA processing. Occurrence of testis- or germ cell-specific mRNAs suggests potential presence of miRNA preferentially or specifically expressed in these cells. In fact, miRNA expression profiling assays revealed that 60% of the testis-expressed miRNAs are ubiquitously expressed and the remaining are either preferentially (35%) or exclusively (5%) expressed in the testis (Ro et al., 2007). Further evidence of the importance of miRNA in germ cell development is provided by the observation of aberrant expression of miRNA in germ cell tumor (GCT) (Gillis et al., 2007; Voorhoeve et al., 2007).

Piwi-interacting RNAs: Piwi-interacting RNAs (piRNAs) are small ncRNAs (26–31 nt in length, depending on subfamily members) that were identified independently in testes by several laboratories (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2006). piRNAs were cloned after co-purification of small RNAs with

a spermatogenesis specific Argonaute protein family member (including Mili, Miwi and Miwi2) and demonstrated distinctive localization pattern in the genome. They constitute the biggest small RNA family; 50,000 species have been identified and the total population is estimated to be around 200,000 (Betel et al., 2007). piRNAs associate with either Mili, which is found in earlier germ cells (spermatogonia to pachytene spermatocytes), or Miwi, which is present in more mature germ cells after the pachytene spermatocyte stage (Deng and Lin, 2002; Kuramochi-Miyagawa and others, 2001). Mili-associated piRNAs typically have 26–28 nucleotides while Miwi-associated piRNAs are larger, usually contain 29 to 31 nucleotides (Aravin et al., 2006). Recent studies suggested that piRNAs play a role in a number of biological processes including transcriptional suppression pathways (Aravin et al., 2007; Brennecke et al., 2007; Gunawardane et al., 2007; Houwing et al., 2007), histone modification (Yin and Lin, 2007), and silencing of repressor through RNAi-mediated (Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004; Pelisson et al., 2007).

Long ncRNA—Recent studies revealed the presence of long ncRNAs (>200 nt in length) (Kampa, 2004; Carninci et al., 2005). The expression of long ncRNA is prevalent throughout the mammalian transcriptome and the configuration is comparable to the protein coding RNAs (Carninci et al., 2005; Engström et al., 2006). The regulatory element, instead of the nucleotide sequence of the long ncRNA, is often conserved (Ponjavic et al., 2007). The lack of sequence conservation was interpreted by some to suggest they may simply be “transcription noise”. However, some long ncRNAs such as *Air* and *Xist*, even though poorly conserved (Nesterova et al., 2001), have well characterized functions (Brockdorff et al., 1992; Sleutels et al., 2002). In fact, long ncRNAs have been found to participate in a variety of cellular processes, including transcriptional regulations in normal and abnormal cellular development (Feng et al., 2006; Goodrich and Kugel, 2006; Pennacchio et al., 2006; Calin et al., 2007; Visel et al., 2008), post-transcriptional regulation including splicing (Beltran et al., 2008), translation (Wang et al., 2005), and siRNA processing (Wang et al., 2005; Golden et al., 2008) and epigenetic regulations (Sanchez-Elsner et al., 2006; Rinn et al., 2007). Examples of long ncRNA with developmental roles include *HOTAIR* (Rinn et al., 2007), *Xist* (Duret et al., 2006), *PINC*, *Ey/2* (Feng et al., 2006). Recent array studies also identified novel long ncRNAs actively involved in embryonic stem cell pluripotency and differentiation (Dinger et al., 2008), brain development (Mercer et al., 2008) and male germ cell cancer (Perez et al., 2008). Given the extensive functional coverage of long ncRNAs in cellular regulation and development, it is highly likely that long ncRNAs may also play a critical role in male germ cell development and provide insights on cellular transitions and maintenance in each key germ cell stage. One of the stumbling blocks with long ncRNAs research, however, is that there is no clue about their biological functions. Unlike protein coding RNAs, ncRNAs are comprised of a heterogeneous group of genes that fulfill diverse biological roles. Conventional prediction methods based on primary sequence or structural features cannot be applied. Given the vast number of ncRNAs, novel high-throughput screening approaches are needed to elucidate their biological functions.

Genomic Adaptation to Meiotic Sex Chromosome Inactivation—Evolution of Autosomal Retrogenes

Male germ cells enter meiosis after mitotic expansion. Shortly after the zygotene-pachytene transition, the X and Y chromosomes are condensed and compartmentalized into a cytological structure called “sex body” or “XY body” (McKee and Handel, 1993). Meiotic sex chromosome inactivation (MSCI) describes the process of transcriptional silencing of the sex chromosomes that occurs during the meiotic phase (pachytene stage) of spermatogenesis as a result of the condensation of sex chromosomes (Turner, 2007). Originally thought to be limited to male meiosis only, the repression of sex-linked gene transcription is now known to extend to the post-meiotic phase. As males carry only one copy of X chromosome, it is conceivable

that the loss of X-linked gene activities would jeopardize the survival of half of the species that possess differentiated sex chromosomes. To maintain the availability of sex-linked gene activities throughout and beyond male meiosis, two major mechanisms are believed to have evolved to overcome the transcriptional silencing event. The most acknowledged strategy is the generation of autosomal copies of X-linked genes (i.e., X-derived autosomal retrogenes) by retroposition which is believed to be important to the generation of functional gene duplicates in new genomic positions (Long et al., 2003). The second strategy, which was illustrated recently, is to increase X-linked gene copy numbers by gene amplification (Mueller et al., 2008). In this section, we will focus our discussion on the involvement of retrogenes in male germ cell development.

Identification of retrogenes and studies of their expression characteristics—

During retroposition, a mature transcript of the progenitor gene is reverse-transcribed into cDNA and integrated into the genome. Most retroposition events generate nonfunctional retroposed gene copies (processed pseudogenes) because they lack promoter elements for transcription and may have accumulated degenerative mutations (e.g., introduction of stop codons and frameshift mutations). Some of them, however, are transcribed and retain protein coding potential and are thus named retrogenes. Several genome-wide searches of retrogene movement among chromosomes have been published recently (Emerson et al., 2004; Vinckenbosch et al., 2006; Shiao et al., 2007; Potrzebowski et al., 2008). These studies predicted that at least ~100 progenitor-retrogene pairs, for which the retrogenes do not share chromosome linkage with the progenitors, exist in the genome of human, mouse, dog or opossum (Emerson et al., 2004; Potrzebowski et al., 2008). Further analyses on the chromosomal locations revealed a large excess of retrogenes to be originated from the X chromosome when comparing to autosomes. The excess X chromosome retroposons is not observed for processed pseudogenes, suggesting that functional retrogenes showing X-to-autosome movement are preserved by natural selection, which is believed to be driven by MSCI. Interestingly, the majority of X-derived retrogenes are specifically or predominantly expressed in the testis (particularly in meiotic and post-meiotic male germ cells), and they display functional roles in spermatogenesis after leaving the X chromosome when compared to autosome-derived retrogenes (Potrzebowski et al., 2008). In contrast, their X-linked progenitor genes demonstrate broad expression patterns in different somatic tissues but are down-regulated starting from male meiosis. Such reciprocal spermatogenic expression pattern is coherent with the compensation hypothesis (McCarrey and Thomas, 1987) that autosomal retrogenes are evolved to compensate for the loss of their X-linked progenitors during male meiosis. Since autosomes would not be subjected to MSCI effect, genes encoding products essential to male meiotic and post-meiotic activities would be fixed on autosomes on evolutionary scale. In this regard, any loss-of-function mutation to these retrogenes would be expected to incur spermatogenic defects, a phenomenon that has been demonstrated in several rodent studies (Bradley et al., 2004; Miki et al., 2004; Rohozinski and Bishop, 2004; Rohozinski et al., 2006).

Regulation of retrogene expression in male germ cells—The prevalence of X-derived autosomal retrogene expression in the testis leads to the interesting question of how their tissue-specificity is regulated. The majority of retroposition events are known to occur in the male germline (Khil et al., 2005). It has been predicted that most retrogenes, especially younger ones, would initially be transcribed in the testis because of the facilitated transcription in this tissue (Vinckenbosch et al., 2006; Shiao et al., 2007) and thus they would evolve a functional role in the testis first. Despite these observations, little is known about the mechanism of X-derived autosomal retrogene transcription and how exactly the testis-specificity of these genes is accomplished. In humans, a significant excess of retrogenes are found to locate close to other genes or within introns, suggesting retrogenes can exploit the

open chromatin of neighboring genes to drive their transcription. In this regard, genomic regions surrounding retrogenes are found to be transcriptionally more active than those surrounding transcriptionally silent retrocopies (Vinckenbosch et al., 2006). Retrogenes can also achieve competence in transcription by “hitchhiking” the regulatory elements of host genes (e.g., by promoter element acquisition or chimeric transcript formation) or acquiring untranslated exons *de novo* (Vinckenbosch et al., 2006; Shiao et al., 2007).

At the molecular level, there are examples that consistently point to the involvement of epigenetic factors, notably DNA methylation, in the regulation of tissue specificity of X-derived autosomal retrogenes. Hypermethylation of CpG dinucleotides has been observed in retrogenes such as *Pdha2* (Iannello et al., 1997), *Pgk2* (Geyer et al., 2004) and *Ard1b* (Pang et al., in press) in somatic cells/tissues that do not show endogenous expression of the respective genes. On the contrary, these genes are exclusively or predominantly expressed in the testis, with which a correlation of the extent of DNA hypomethylation and the upregulation of gene expression is observed (Pang et al., in press). It is not yet certain if DNA methylation is commonly employed to regulate X-derived autosomal retrogene transcription in the testis. A detailed examination of the genomic landscapes of all X-derived autosomal retrogenes may thus provide important clues on the involvement of epigenetic factors, including differentially methylated regions (Yagi et al., 2008) and histone modifications, (Berger, 2007) in the regulation of their expression.

SYSTEMS BIOLOGY OF MALE GERM CELLS—FINAL FRONTIER

The Emergence of Systems Biology

Systems biology examines the networks behind complex functionality. This requires a shift toward quantitative biology. The process of knowledge generation in systems biology involves several key steps as illustrated in Figure 7. Experimental paradigm in systems biology involves integration of experimental and various supporting evidence from the literature. This is an iterative, integrated process of data mining and collection. To facilitate data exchange and processing, the information has to be in digital format and able to be processed by the computer using different computational algorithms (*in silico*). Data mining and integration agglomerate sufficient detail for the generation of mathematical model, which allows the generation of various hypotheses about the characteristics and behavior of the system to be derived. These hypotheses can then be cross-validated *in silico* or in *in vitro* or *in vivo* experiments. The ultimate goal is to develop global representations of the biological system of the cell and the entire organism.

Data capture and mining—The first step of system biology research starts with data capture. Genomic studies involve large sets of genes or their products in the form of RNA expression (transcriptome), protein expression (proteome), protein–protein or protein–DNA interactions (interac-tome) by high-throughput genomic assays. It spawns new disciplines including various “omics” approaches to capture expression information. Methods such as microarray and SAGE discussed in the earlier sections are among the popular tools to survey the expression from the genomes. The generation of various datasets from omics approaches is the prerequisite for getting to the basic level of the biological system - the data level. Most of the biological information in the past decades is characterized in forms of qualitative and descriptive formats. Because of the complexity of biology and various type of information involved, creating a consolidated database is the essential step for the subsequent model generation. Development of high-throughput technologies in recent years facilitated the exponential growth in the size and the details of various databases which often reach petabyte scale (Cochrane et al., 2008). Specialized databases have been developed. The knowledge (data) base can be generally classified into primary database, theme-based database, and

pathway and interaction database. Some of the frequently used databases in systems biology are shown in Table 2.

Primary databases: Primary databases are usually the “hub” of basic biological information. These include microarray database Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI) and ArrayExpress at the European Bioinformatics Institute (EMBL-EBI), SAGE database (SAGEmap) (Lash et al., 2000), microRNA database (miR-Base) (Griffiths-Jones et al., 2008), mammalian non-protein-coding RNA database (RNAdb) (Pang et al., 2007), etc. They serve as repositories for expression datasets from high-throughput genomic assays.

Other than sequence context, functional context-based databases are evolved with the advent of omics data. A microarray experiment may result in hundreds of differentially expressed genes that are subject to interpretation and further analysis. As analyzing these lists gene-by-gene is tedious and error prone, the concept of Gene Ontology (GO) has emerged. The main objective of GO is to provide controlled vocabularies for the description of the molecular function, biological process, and cellular component of gene products across different species. The GO terms can then be associated to various datasets and provide insights in terms of unified biological vocabularies.

Theme-based databases: While most databases mentioned covers a wide variety of information in different organisms. Some databases are specialized in certain biological phenomenon or species. GermSAGE is a comprehensive transcriptome database for mouse male germ cells derived from SAGE experiments (Lee et al., 2009). A total of 452,095 tags derived from type A spermatogonia, pachytene spermatocytes and round spermatids were included. It allows browsing, comparing and searching male germ cell transcriptome data at different stages with customizable searching parameters. The data can be visualized in a tabulated format or further analyzed by aligning with various annotations available in the UCSC genome browser. This flexible platform is useful for gaining better understanding of the genetic networks that regulate spermatogonial cell renewal and differentiation, and will allow novel gene discovery. Another germ cell database, GermOnline, mainly stores microarray data related to the study of mitosis, meiosis, germline development and gametogenesis across species (Gattiker et al., 2007).

To extract information from the datasets, bioinformatics methods are used to identify trends and correlations of different genes from the prospective of expression level, sequence context or other intrinsic biological features, including potential protein domains and *cis* or *trans*-regulatory elements. Through the integration of informatics and results from different omics studies at different levels, data on *individual* components or functions are amassed.

Integration—To transit from the knowledge of individual component to the functional relationship between the components is the key to the understanding of an entire biological system. It is necessary to collect information on the interactions between the components in the system at different levels and on the dynamics of the interactions to obtain insights on the properties of complex systems such as regulation, control and adaptation. This integration step is facilitated by the availability of pathway and interaction databases.

Pathway and interaction databases: The explosion in the amount of biological data and information in recent years has made the construction of complex network of biological associations possible. Examples of public resources for finding associations among gene sets and known biological pathways include BioCarta which is useful for charting pathways, BIND which is a molecular interaction database, Reactome which provides information on core human pathways, etc. Some of the pathway and interaction databases are listed in Table 2.

Modeling and hypothesis generation—With sufficient biological knowledge, data can be assimilated for the next step - modeling and hypothesis generation. A model is the attempt to create an abstract representation from different knowledge or data resources, such as experimental observations on the structure and function of a particular biological element. Decoding the complexity of male germ cell development by SAGE data represents such an approach (Lee et al., 2006). Instead of resorting to descriptive analysis, the dynamics contained in the dataset was retrieved by applying clustering algorithm similar to microarray studies. The clustered gene sets demonstrating preferential expression for the three germ cell stages were exported for gene network analysis. Based on the unique, curated database of protein-protein and protein-DNA interactions, transcriptional factors, signaling, metabolism and bioactive molecules, a network specific for each germ cell stage was created, which allowed data visualization on a list of relevant biological networks (Fig. 8A–C). The figures not only link the gene members in the clusters, but also provide the functionality to identify co-regulated genes ranked by statistics and scores. This provides a foundation for the generation of novel biological hypotheses for studying spermatogenesis.

CONCLUSIONS

Spermatogenesis is essential for the perpetuation of the human race. Germ cells are the only cells in our body that undergo mitosis and meiosis. It is not surprising that germ cells express the largest number of genes. It is also not surprising that genetic processes that occur in different somatic cells are also operative in germ cells. The advances brought forth by the Human Genome Project give us tools to examine the complex genomic and genetic landscape of the germ cells. The discoveries of the biological activities of alternative splice variants, anti-sense transcripts, pseudogenes, retrogenes, miRNA, piRNA, long ncRNAs, etc., in male germ cells represent just the tip of the iceberg. It is almost certain that application of systems biology approaches to study male germ cells will bring more surprises in the foreseeable future.

Acknowledgments

Grant sponsor: NIH (Intramural Research Program), Eunice Kennedy Shriver National Institute of Child Health and Human Development.

References

- Alfano G, Vitiello C, Caccioppoli C, et al. Natural antisense transcripts associated with genes involved in eye development. *Hum Mol Genet* 2005;14:913–923. [PubMed: 15703187]
- Almstrup K, Nielsen JE, Hansen MA, et al. Analysis of cell-type-specific gene expression during mouse spermatogenesis. *Biol Reprod* 2004;70:1751–1761. [PubMed: 14960480]
- Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome* 2008;19:454–492. [PubMed: 18839252]
- Anway MD, Li Y, Ravindranath N, et al. Expression of testicular germ cell genes identified by differential display analysis. *J Androl* 2003;24:173–184. [PubMed: 12634303]
- Aravin A, Gaidatzis D, Pfeffer S, et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 2006;442:203–207. [PubMed: 16751777]
- Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 2007;318:761–764. [PubMed: 17975059]
- Aschim EL, Saether T, Wiger R, et al. Differential distribution of splice variants of estrogen receptor beta in human testicular cells suggests specific functions in spermatogenesis. *J Steroid Biochem* 2004;92:97–106.
- Ast G. How did alternative splicing evolve? *Nat Rev Genet* 2004;5:773–782. [PubMed: 15510168]
- Banks RE, Dunn MJ, Hochstrasser DF, et al. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 2000;356:1749–1756. [PubMed: 11095271]

- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–297. [PubMed: 14744438]
- Bartsch H, Voigtsberger S, Baumann G, et al. Detection of a novel sense-antisense RNA-hybrid structure by RACE experiments on endogenous troponin I antisense RNA. *RNA* 2004;10:1215–1224. [PubMed: 15272119]
- Behr R, Weinbauer GF. cAMP response element modulator (CREM): an essential factor for spermatogenesis in primates? *Int J Androl* 2001;24:126–135. [PubMed: 11380701]
- Belmonte SA, Romano PS, Sosa MA. Mannose-6-phosphate receptors as a molecular indicator of maturation of epididymal sperm. *Arch Androl* 2002;48:53–63. [PubMed: 11789684]
- Beltran M, Puig I, Pena C, et al. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 2008;22:756–769. [PubMed: 18347095]
- Ben-Dov C, Hartmann B, Lundgren J, Valcarcel J. Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem* 2008;283:1229–1233. [PubMed: 18024428]
- Berger SL. The complex language of chromatin regulation during transcription. *Nature* 2007;447:407–412. [PubMed: 17522673]
- Berteaux N, Aptel N, Cathala G, et al. A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Mol Cell Biol* 2008;28:6731–6745. [PubMed: 18794369]
- Betel D, Breitkreuz KE, Isserlin R, et al. Structure-templated predictions of novel protein interactions from sequence information. *PLoS Comput Biol* 2007;3:1783–1789. [PubMed: 17892321]
- Bharadwaj A, Ghosh I, Sengupta A, et al. Stage-specific expression of proprotein form of hyaluronan binding protein 1 (HABP1) during spermatogenesis in rat. *Mol Reprod Dev* 2002;62:223–232. [PubMed: 11984833]
- Birzele F, Kuffner R, Meier F, et al. ProSAS: a database for analyzing alternative splicing in the context of protein structures. *Nucleic Acids Res* 2008;36(Database issue):D63–D68. [PubMed: 17933774]
- Blin-Wakkach C, Lezot F, Ghoul-Mazgar S, et al. Endogenous Msx1 antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. *Proc Natl Acad Sci U S A* 2001;98:7336–7341. [PubMed: 11390985]
- Blocher S, Fink L, Bohle RM, et al. CREM activator and repressor isoform expression in human male germ cells. *Int J Androl* 2005;28:215–223. [PubMed: 16048633]
- Bonafoux B, Lejeune M, Piquemal D, et al. Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage. *Haematologica* 2004;89:1434–1438. [PubMed: 15590392]
- Bono H, Kasukawa T, Furuno M, et al. FANTOM DB: database of Functional Annotation of RIKEN Mouse cDNA Clones. *Nucleic Acids Res* 2002;30:116–118. [PubMed: 11752270]
- Boue S, Letunic I, Bork P. Alternative splicing and evolution. *Bioessays* 2003;25:1031–1034. [PubMed: 14579243]
- Bradley J, Baltus A, Skaletsky H, et al. An X-to-autosome retrogene is required for spermatogenesis in mice. *Nat Genet* 2004;36:872–876. [PubMed: 15258580]
- Brennecke J, Aravin AA, Stark A, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 2007;128:1089–1103. [PubMed: 17346786]
- Brockdorff N, Ashworth A, Kay GF, et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 1992;71:515–526. [PubMed: 1423610]
- Brown CJ, Chow JC. Beyond sense: the role of antisense RNA in controlling Xist expression. *Semin Cell Dev Biol* 2003;14:341–347. [PubMed: 15015741]
- Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 2000;28:4364–4375. [PubMed: 11058137]
- Calarco JA, Saltzman AL, Ip JY, Blencowe BJ. Technologies for the global discovery and analysis of alternative splicing. *Adv Exp Med Biol* 2007;623:64–84. [PubMed: 18380341]
- Calin GA, Liu CG, Ferracin M, et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007;12:215–229. [PubMed: 17785203]

- Carmell MA, Hannon GJ. RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol* 2004;11:214–218. [PubMed: 14983173]
- Carmichael GG. Antisense starts making more sense. *Nat Biotechnol* 2003;21:371–372. [PubMed: 12665819]
- Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. *Science* 2005;309:1559–1563. [PubMed: 16141072]
- Castrignano T, D'Antonio M, Anselmo A, et al. ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics* 2008;24:1300–1304. [PubMed: 18388144]
- Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, et al. The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A* 2007;104:8346–8351. [PubMed: 17483452]
- Chan WY, Lee TL, Wu SM, et al. Transcriptome analyses of male germ cells with serial analysis of gene expression (SAGE). *Mol Cell Endocrinol* 2006a;250:8–19. [PubMed: 16413108]
- Chan WY, Wu SM, Rusczyk L, et al. The complexity of antisense transcription revealed by the study of developing male germ cells. *Genomics* 2006b;87:681–692. [PubMed: 16458478]
- Chan, WY.; Wu, SM.; Rusczyk, L., et al. Antisense transcription in developing male germ cells. In: Lau, YFC.; Chan, WY., editors. *The Y Chromosome and Male Germ Cell Biology in Health and Diseases*. Hackensack: World Scientific; 2007. p. 201–220.
- Chen J, Sun M, Kent WJ, et al. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* 2004;32:4812–4820. [PubMed: 15356298]
- Cheng J, Kapranov P, Drenkow J, et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 2005;308:1149–1154. [PubMed: 15790807]
- Cochrane G, Akhtar R, Bonfield J, et al. Petabyte-scale innovations at the European Nucleotide Archive. *Nucl Acids Res* 2008;37(Database issue):D19–D25. [PubMed: 18978013]
- Cuperlovic-Culf M, Belacel N, Culf AS, Ouellette RJ. Microarray analysis of alternative splicing. *OMICS* 2006;10:344–357. [PubMed: 17069512]
- Dahary D, Elroy-Stein O, Sorek R. Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res* 2005;15:364–368. [PubMed: 15710751]
- Delmas V, Vanderhoorn F, Mellstrom B, et al. Induction of Crem Activator Proteins in Spermatids—Down-Stream Targets and Implications for Haploid Germ-Cell Differentiation. *Mol Endocrinol* 1993;7:1502–1514. [PubMed: 8114765]
- Deng W, Lin H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* 2002;2:819–830. [PubMed: 12062093]
- Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:P3. [PubMed: 12734009]
- Dinger ME, Amaral PP, Mercer TR, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 2008;18:1433–1445. [PubMed: 18562676]
- Downey KM, Byrnes JJ, Jurmark BS, So AG. Reticulocyte RNA-dependent RNA polymerase. *Proc Natl Acad Sci U S A* 1973;70:3400–3404. [PubMed: 4519633]
- Duret L, Chureau C, Samain S, et al. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 2006;312:1653–1655. [PubMed: 16778056]
- Dym M. Spermatogonial stem cells of the testis. *Proc Natl Acad Sci U S A* 1994;91:11287–11289. [PubMed: 7972051]
- Eddy, EM. Recent Prog Horm Res. Bethesda: Endocrine Soc; 2002. Male germ cell gene expression; p. 103–128.
- Emerson JJ, Kaessmann H, Betran E, Long M. Extensive gene traffic on the mammalian X chromosome. *Science* 2004;303:537–540. [PubMed: 14739461]
- Engström P, Suzuki H, Ninomiya N, et al. Complex loci in human and mouse genomes. *PLoS Genetics* 2006;2:e47. [PubMed: 16683030]
- Fahey ME, Moore TF, Higgins DG. Overlapping antisense transcription in the human genome. *Comp Funct Genomics* 2002;3:244–253. [PubMed: 18628857]
- Feki A, Jefford CE, Durand P, et al. BARD1 expression during spermatogenesis is associated with apoptosis and hormonally regulated. *Biol Reprod* 2004;71:1614–1624. [PubMed: 15240424]

- Feng J, Bi C, Clark BS, et al. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev* 2006;20:1470–1484. [PubMed: 16705037]
- Forch P, Puig O, Martinez C, et al. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J* 2002;21:6882–6892. [PubMed: 12486009]
- Foulkes NS, Mellstrom B, Benusiglio E, Sassonecorsi P. Developmental switch of CREM - function during spermatogenesis - from antagonist to activator. *Nature* 1992;355:80–84. [PubMed: 1370576]
- Frey BJ, Mohammad N, Morris QD, et al. Genome-wide analysis of mouse transcripts using exon microarrays and factor graphs. *Nat Genet* 2005;37:991–996. [PubMed: 16127451]
- Fujii T, Tamura K, Masai K, et al. Use of stepwise subtraction to comprehensively isolate mouse genes whose transcription is up-regulated during spermiogenesis. *EMBO Rep* 2002;3:367–372. [PubMed: 11943763]
- Furukawa K, Hotta Y. cDNA cloning of a germ-cell specific lamin-b3 from mouse spermatocytes and analysis of its function by ectopic expression in somatic-cells. *EMBO J* 1993;12:97–106. [PubMed: 8094052]
- Gattiker A, Niederhauser-Wiederkehr C, et al. The GermOnline cross-species systems browser provides comprehensive information on genes and gene products relevant for sexual reproduction. *Nucleic Acids Res* 2007;35(Database issue):D457–D462. [PubMed: 17145711]
- Geyer CB, Kiefer CM, Yang TP, McCarrey JR. Ontogeny of a demethylation domain and its relationship to activation of tissue-specific transcription. *Biol Reprod* 2004;71:837–844. [PubMed: 15140797]
- Gillis AJ, Stoop HJ, Hersmus R, et al. High-throughput micro-RNAome analysis in human germ cell tumours. *J Pathol* 2007;213:319–328. [PubMed: 17893849]
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 2006;442:199–202. [PubMed: 16751776]
- Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. *Mol Cell* 2008;31:309–312. [PubMed: 18691963]
- Goodrich JA, Kugel JF. Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol* 2006;7:612–616. [PubMed: 16723972]
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res* 2008;36(Database issue):D154–D158. [PubMed: 17991681]
- Grivna ST, Beyret E, Wang Z, Lin HF. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* 2006;20:1709–1714. [PubMed: 16766680]
- Gunawardane LS, Saito K, Nishida KM, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 2007;315:1587–1590. [PubMed: 17322028]
- Guo R, Yu Z, Guan J, et al. Stage-specific and tissue-specific expression characteristics of differentially expressed genes during mouse spermatogenesis. *Mol Reprod Dev* 2004;67:264–272. [PubMed: 14735487]
- Guo TB, Boros LG, Chan KC, et al. Spermatogenic expression of RNA-binding motif protein 7, a protein that interacts with splicing factors. *J Androl* 2003;24:204–214. [PubMed: 12634307]
- Harrison PM, Zheng D, Zhang Z, et al. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res* 2005;33:2374–2383. [PubMed: 15860774]
- Hastings ML, Ingle HA, Lazar MA, Munroe SH. Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring anti-sense RNA. *J Biol Chem* 2000;275:11507–11513. [PubMed: 10753970]
- Hayashi K, Chuva de Sousa Lopes SM, et al. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS ONE* 2008;3:e1738. [PubMed: 18320056]
- He Y, Vogelstein B, Velculescu VE, et al. The antisense transcriptomes of human cells. *Science* 2008;322:1855–1857. [PubMed: 19056939]
- Hecht NB. Molecular mechanisms of male germ cell differentiation. *Bioessays* 1998;20:555–561. [PubMed: 9723004]

- Hernandez A, Martinez ME, Croteau W, St Germain DL. Complex organization and structure of sense and antisense transcripts expressed from the DIO3 gene imprinted locus. *Genomics* 2004;83:413–424. [PubMed: 14962667]
- Houwing S, Kamminga LM, Berezikov E, et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 2007;129:69–82. [PubMed: 17418787]
- Iannello RC, Young J, Sumarsono S, et al. Regulation of Pdha-2 expression is mediated by proximal promoter sequences and CpG methylation. *Mol Cell Biol* 1997;17:612–619. [PubMed: 9001214]
- Ihalmo P, Rinta-Valkama J, Mai P, et al. Molecular cloning and characterization of an endogenous antisense transcript of Nphs1. *Genomics* 2004;83:1134–1140. [PubMed: 15177566]
- Imamura T, Yamamoto S, Ohgane J, et al. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem Biophys Res Commun* 2004;322:593–600. [PubMed: 15325271]
- Jaiswal BS, Conti M. Identification and functional analysis of splice variants of the germ cell soluble adenylyl cyclase. *J Biol Chem* 2001;276:31698–31708. [PubMed: 11423534]
- Jaspers S, Gellersen B, Kempf R, et al. Functional characterization of male germ cell-specific CREM isoforms. *J Androl* 2007;28:59–66. [PubMed: 16899810]
- Jensen KB, Dredge BK, Stefani G, et al. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* 2000;25:359–371. [PubMed: 10719891]
- Jin H, Vacic V, Girke T, et al. Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis. *BMC Mol Biol* 2008;9:6. [PubMed: 18194570]
- Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet* 2005;21:93–102. [PubMed: 15661355]
- Johnston DS, Wright WW, Dicandoloro P, et al. Stage-specific gene expression is a fundamental characteristic of rat spermatogenic cells and Sertoli cells. *Proc Natl Acad Sci U S A* 2008;105:8315–8320. [PubMed: 18544648]
- Kampa D. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 2004;14:331–342. [PubMed: 14993201]
- Kampa D, Cheng J, Kapranov P, et al. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 2004;14:331–342. [PubMed: 14993201]
- Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. *Science* 2005;309:1564–1566. [PubMed: 16141073]
- Keegan LP, Gallo A, O'Connell MA. The many roles of an RNA editor. *Nat Rev Genet* 2001;2:869–878. [PubMed: 11715042]
- Khelifi A, Duret L, Mouchiroud D. HOPPSIGEN: a database of human and mouse processed pseudogenes. *Nucleic Acids Res* 2005;33(Database issue):D59–D66. [PubMed: 15608268]
- Khil PP, Oliver B, Camerini-Otero RD. X for intersection: retrotransposition both on and off the X chromosome is more frequent. *Trends Genet* 2005;21:3–7. [PubMed: 15680505]
- Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet* 2004;36:642–646. [PubMed: 15156144]
- Kim N, Lee C. Bioinformatics detection of alternative splicing. *Methods Mol Biol* 2008;452:179–197. [PubMed: 18566765]
- Kim N, Shin S, Lee S. ASmodeler: gene modeling of alternative splicing from genomic alignment of mRNA, EST and protein sequences. *Nucleic Acids Res* 2004;32(Web Server issue):W181–W186. [PubMed: 15215376]
- Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 2005;6:376–385. [PubMed: 15852042]
- Kiyosawa H, Abe K. Speculations on the role of natural antisense transcripts in mammalian X chromosome evolution. *Cytogenet Genome Res* 2002;99:151–156. [PubMed: 12900558]
- Kiyosawa H, Yamanaka I, Osato N, et al. Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res* 2003;13(6B):1324–1334. [PubMed: 12819130]

- Kleene KC. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech Dev* 2001;106:3–23. [PubMed: 11472831]
- Kleene KC. Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenet Genome Res* 2003;103:217–224. [PubMed: 15051942]
- Knee R, Li AW, Murphy PR. Characterization and tissue-specific expression of the rat basic fibroblast growth factor antisense mRNA and protein. *Proc Natl Acad Sci U S A* 1997;94:4943–4947. [PubMed: 9144169]
- Korneev SA, Korneeva EI, Lagarkova MA, et al. Novel noncoding antisense RNA transcribed from human anti-NOS2A locus is differentially regulated during neuronal differentiation of embryonic stem cells. *RNA* 2008;14:2030–2037. [PubMed: 18820242]
- Korneev SA, Park JH, O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* 1999;19:7711–7720. [PubMed: 10479675]
- Kotaja N, Bhattacharyya SN, Jaskiewicz L, et al. The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc Natl Acad Sci U S A* 2006;103:2647–2652. [PubMed: 16477042]
- Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 1992;90:41–54. [PubMed: 1427786]
- Kuramochi-Miyagawa S, Kimura T, Yomogida K, et al. Two mouse piwi-related genes: miwi and mili. *Mech Dev* 2001;108:121–133. [PubMed: 11578866]
- Kuroda M, Sok J, Webb L, et al. Male sterility and enhanced radiation sensitivity in TLS(–/–) mice. *EMBO J* 2000;19:453–462. [PubMed: 10654943]
- Landers M, Calciano MA, Colosi D, et al. Maternal disruption of Ube3a leads to increased expression of Ube3a-ATS in trans. *Nucleic Acids Res* 2005;33:3976–3984. [PubMed: 16027444]
- Lash AE, Tolstoshev CM, Wagner L, et al. SAGEmap: a public gene expression resource. *Genome Res* 2000;10:1051–1060. [PubMed: 10899154]
- Lau NC, Seto AG, Kim J, et al. Characterization of the piRNA complex from rat testes. *Science* 2006;313:363–367. [PubMed: 16778019]
- Lavorgna G, Dahary D, Lehner B, et al. In search of antisense. *Trends Biochem Sci* 2004;29:88–94. [PubMed: 15102435]
- Lee TL, Alba D, Baxendale V, et al. Application of transcriptional and biological network analyses in mouse germ-cell transcriptomes. *Genomics* 2006;88:18–33. [PubMed: 16678385]
- Lee TL, Cheung HH, Claus J, et al. GermSAGE: a comprehensive SAGE database for transcript discovery on male germ cell development. *Nucleic Acids Res* 2009;37(Database issue):D891–D897. [PubMed: 18832368]
- Lee Y, Kim B, Shin Y, et al. ECgene: an alternative splicing database update. *Nucleic Acids Res* 2007;35(Database issue):D99–D103. [PubMed: 17132829]
- Lehner B, Williams G, Campbell RD, Sanderson CM. Antisense transcripts in the human genome. *Trends Genet* 2002;18:63–65. [PubMed: 11818131]
- Li AW, Murphy PR. Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation. *Mol Cell Endocrinol* 2000;170:233–242. [PubMed: 11162906]
- Li T, Vu TH, Lee KO, et al. An imprinted PEG1/MEST antisense expressed predominantly in human testis and in mature spermatozoa. *J Biol Chem* 2002;277:13518–13527. [PubMed: 11821432]
- Liu J, Carmell MA, Rivas FV, et al. Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science* 2004;305:1437–1441. [PubMed: 15284456]
- Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 2003;4:865–875. [PubMed: 14634634]
- Lunter G, Ponting CP, Hein J. Genome-wide identification of human functional dna using a neutral indel model. *PLoS Comput Biol* 2006;2:e5. [PubMed: 16410828]

- Lu BS, Bishop CE. Mouse GGN1 and GGN3, two germ cell-specific proteins from the single gene Ggn, interact with mouse POG and play a role in spermatogenesis. *J Biol Chem* 2003;278:16289–16296. [PubMed: 12574169]
- Maatouk DM, Loveland KL, McManus MT, et al. Dicer1 is required for differentiation of the mouse male germline. *Biol Reprod* 2008;79:696–703. [PubMed: 18633141]
- Mallardo M, Poltronieri P, D'Urso OF. Non-protein coding RNA bio-markers and differential expression in cancers: a review. *J Exp Clin Cancer Res* 2008;27:19. [PubMed: 18631387]
- Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 2002;418:236–243. [PubMed: 12110900]
- Matter N, Herrlich P, Konig H. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* 2002;420:691–695. [PubMed: 12478298]
- Matthews L, Gopinath G, Gillespie M, et al. Reactome knowledge-base of human biological pathways and processes. *Nucleic Acids Res* 2008;37(Database issue):D619–D22. [PubMed: 18981052]
- Mattick JS. RNA regulation: a new genetics? *Nat Rev Genet* 2004;5:316–323. [PubMed: 15131654]
- McCarrey JR, Thomas K. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* 1987;326:501–505. [PubMed: 3453121]
- McKee BD, Handel MA. Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* 1993;102:71–80. [PubMed: 8432196]
- Mercer TR, Dinger ME, Sunkin SM, et al. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A* 2008;105:716–721. [PubMed: 18184812]
- Miki K, Qu W, Goulding EH, et al. Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility. *Proc Natl Acad Sci U S A* 2004;101:16501–16506. [PubMed: 15546993]
- Morales CR, Hecht NB. Poly(A)+ ribonucleic acids are enriched in spermatocyte nuclei but not in chromatoid bodies in the rat testis. *Biol Reprod* 1994;50:309–319. [PubMed: 7511419]
- Morris KV, Santoso S, Turner AM, et al. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* 2008;4:e1000258. [PubMed: 19008947]
- Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 1982;10:459–472. [PubMed: 7063411]
- Mueller JL, Mahadevaiah SK, Park PJ, et al. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet* 2008;40:794–799. [PubMed: 18454149]
- Nagoshi RN, McKeown M, Burtis KC, et al. The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. *Cell* 1988;53:229–236. [PubMed: 3129196]
- Nemes JP, Benzow KA, Moseley ML, et al. The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). *Hum Mol Genet* 2000;9:1543–1551. [PubMed: 10888605]
- Nesterova TB, Barton SC, Surani MA, Brockdorff N. Loss of Xist imprinting in diploid parthenogenetic preimplantation embryos. *Dev Biol* 2001;235:343–350. [PubMed: 11437441]
- Nishida H, Tomaru Y, Oho Y, Hayashizaki Y. Naturally occurring anti-sense RNA of histone H2a in mouse cultured cell lines. *BMC Genet* 2005;6:23. [PubMed: 15892893]
- Noh SJ, Lee K, Paik H, Hur CG. TISA: tissue-specific alternative splicing in human and mouse genes. *DNA Res* 2006;13:229–243. [PubMed: 17107969]
- Ogawa Y, Lee JT. Antisense regulation in X inactivation and autosomal imprinting. *Cytogenet Genome Res* 2002;99:59–65. [PubMed: 12900546]
- Ogawa Y, Sun BK, Lee JT. Intersection of the RNA interference and X-inactivation pathways. *Science* 2008;320:1336–1341. [PubMed: 18535243]
- Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 2002;420:563–573. [PubMed: 12466851]
- Okuda S, Yamada T, Hamajima M, et al. KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res* 2008;36(Web Server issue):W423–W426. [PubMed: 18477636]
- Orfanelli U, Wenke AK, Doglioni C, et al. Identification of novel sense and antisense transcription at the TRPM2 locus in cancer. *Cell Res* 2008;18:1128–1140. [PubMed: 18957938]

- Pal-Bhadra M, Bhadra U, Birchler JA. RNAi related mechanisms affect both transcriptional and post-transcriptional transgene silencing in *Drosophila*. *Mol Cell* 2002;9:315–327. [PubMed: 11864605]
- Pal-Bhadra M, Leibovitch BA, Gandhi SG, et al. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* 2004;303:669–672. [PubMed: 14752161]
- Pandey RR, Mondal T, Mohammad F, et al. Kcnq1ot1 antisense non-coding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008;32:232–246. [PubMed: 18951091]
- Pang AL, Peacock S, Johnson W, et al. Testis-predominant and stage-specific expression of a novel acetyl-transferase retrogene involves DNA demethylation, 2009. *Biol Reprod*. February 25;2009 10.1095/biolreprod.108.073221
- Pang AL, Johnson W, Ravindranath N, et al. Expression profiling of purified male germ cells: stage-specific expression patterns related to meiosis and postmeiotic development. *Physiol Genomics* 2006;24:75–85. [PubMed: 16291737]
- Pang AL, Taylor HC, Johnson W, et al. Identification of differentially expressed genes in mouse spermatogenesis. *J Androl* 2003;24:899–911. [PubMed: 14581517]
- Pang KC, Stephen S, Dinger ME, et al. RNAdb 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res* 2007;35(Database issue):D178–D182. [PubMed: 17145715]
- Pelisson A, Sarot E, Payen-Groschene G, Bucheton A. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol* 2007;81:1951–1960. [PubMed: 17135323]
- Pennacchio LA, Ahituv N, Moses AM, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 2006;444:499–502. [PubMed: 17086198]
- Perez DS, Hoage TR, Pritchett JR, et al. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet* 2008;17:642–655. [PubMed: 18006640]
- Podlowski S, Bramlage P, Baumann G, et al. Cardiac troponin I sense-antisense RNA duplexes in the myocardium. *J Cell Biochem* 2002;85:198–207. [PubMed: 11891863]
- Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* 2007;17:556–565. [PubMed: 17387145]
- Potrzebowski L, Vinckenbosch N, Marques AC, et al. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol* 2008;6:e80. [PubMed: 18384235]
- Quelle DE, Zindy F, Ashmun RA, Sherr CJ. Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* 1995;83:993–1000. [PubMed: 8521522]
- Reis EM, Louro R, Nakaya HI, Verjovski-Almeida S. As antisense RNA gets intronic. *OMICS* 2005;9:2–12. [PubMed: 15805775]
- Reis EM, Nakaya HI, Louro R, et al. Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene* 2004;23:6684–6692. [PubMed: 15221013]
- Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007;129:1311–1323. [PubMed: 17604720]
- Ro S, Park C, Sanders KM, et al. Cloning and expression profiling of testis-expressed microRNAs. *Dev Biol* 2007;311:592–602. [PubMed: 17936267]
- Robb GB, Carson AR, Tai SC, et al. Post-transcriptional regulation of endothelial nitric-oxide synthase by an overlapping antisense mRNA transcript. *J Biol Chem* 2004;279:37982–37996. [PubMed: 15234981]
- Rockett JC, Christopher Luft J, et al. Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. *Genome Biol* 2001;2:Research0014. [PubMed: 11305942]
- Rohozinski J, Bishop CE. The mouse juvenile spermatogonial depletion (jsd) phenotype is due to a mutation in the X-derived retrogene, mUtp14b. *Proc Natl Acad Sci U S A* 2004;101:11695–11700. [PubMed: 15289605]
- Rohozinski J, Lamb DJ, Bishop CE. UTP14c is a recently acquired retrogene associated with spermatogenesis and fertility in man. *Biol Reprod* 2006;74:644–651. [PubMed: 16354793]

- Rosok O, Sioud M. Systematic identification of sense-antisense transcripts in mammalian cells. *Nat Biotechnol* 2004;22:104–108. [PubMed: 14704709]
- Royo H, Cavaillat J. Non-coding RNAs in imprinted gene clusters. *Biol Cell* 2008;100:149–166. [PubMed: 18271756]
- Runkel F, Michels M, Franz T. Fxyd3 and Lgi4 expression in the adult mouse: a case of endogenous antisense expression. *Mamm Genome* 2003;14:665–672. [PubMed: 14694902]
- Ruzanov P, Jones SJ, Riddle DL. Discovery of novel alternatively spliced *C. elegans* transcripts by computational analysis of SAGE data. *BMC Genomics* 2007;8:447. [PubMed: 18053145]
- Sanborn BM, Millan JL, Meistrich ML, Moore LC. Alternative splicing of CREB and CREM mRNAs in an immortalized germ cell line. *J Androl* 1997;18:62–70. [PubMed: 9089069]
- Sanchez-Elsner T, Gou D, Kremmer E, Sauer F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* 2006;311:1118–1123. [PubMed: 16497925]
- Sassone-Corsi P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* 2002;296:2176–2178. [PubMed: 12077401]
- Schaefer CF, Anthony K, Krupa S, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res* 2008;37(Database issue):D674–D679. [PubMed: 18832364]
- Schlecht U, Demougin P, Koch R, et al. Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. *Mol Biol Cell* 2004;15:1031–1043. [PubMed: 14718556]
- Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A* 2003;100:12201–12206. [PubMed: 14526100]
- Seim I, Carter SL, Herington AC, Chopin LK. Complex organisation and structure of the ghrelin antisense strand gene GHRLOS, a candidate non-coding RNA gene. *BMC Mol Biol* 2008;9:95. [PubMed: 18954468]
- Seitz A, Gourevitch D, Zhang XM, et al. Sense and antisense transcripts of the apolipoprotein E gene in normal and ApoE knockout mice, their expression after spinal cord injury and corresponding human transcripts. *Hum Mol Genet* 2005;14:2661–2670. [PubMed: 16091415]
- Sha J, Zhou Z, Li J, et al. Identification of testis development and spermatogenesis-related genes in human and mouse testes using cDNA arrays. *Mol Hum Reprod* 2002;8:511–517. [PubMed: 12029067]
- Shendure J, Church GM. Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol* 2002;3:RESEARCH0044. [PubMed: 12225583]
- Shi M, Yan X, Ryan DH, Harris RB. Identification of urocortin mRNA antisense transcripts in rat tissue. *Brain Res Bull* 2000;53:317–324. [PubMed: 11113586]
- Shiao MS, Khil P, Camerini-Otero RD, et al. Origins of new male germ-line functions from X-derived autosomal retrogenes in the mouse. *Mol Biol Evol* 2007;24:2242–2253. [PubMed: 17646254]
- Shibata S, Wutz A. Transcript versus transcription? *Epigenetics* 2008;3:246–249. [PubMed: 19013827]
- Shima JE, McLean DJ, McCarrey JR, Griswold MD. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol Reprod* 2004;71:319–330. [PubMed: 15028632]
- Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002;415:810–813. [PubMed: 11845212]
- Solda G, Boi S, Duga S, et al. In vivo RNA-RNA duplexes from human alpha3 and alpha5 nicotinic receptor subunit mRNAs. *Gene* 2005;345:155–164. [PubMed: 15716100]
- Spiess AN, Walther N, Muller N, et al. SPEER—a new family of testis-specific genes from the mouse. *Biol Reprod* 2003;68:2044–2054. [PubMed: 12606357]
- Tanaka K, Tamura H, Tanaka H, et al. Spermatogonia-dependent expression of testicular genes in mice. *Dev Biol* 2002;246:466–479. [PubMed: 12051830]
- Tufarelli C, Stanley JA, Garrick D, et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 2003;34:157–165. [PubMed: 12730694]
- Turner JM. Meiotic sex chromosome inactivation. *Development* 2007;134:1823–1831. [PubMed: 17329371]

- Venables JP, Elliott DJ, Makarova OV, et al. RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2-beta and affect splicing. *Hum Mol Genet* 2000;9:685–694. [PubMed: 10749975]
- Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A* 2006;103:3220–3225. [PubMed: 16492757]
- Visel A, Prabhakar S, Akiyama JA, et al. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* 2008;40:158–160. [PubMed: 18176564]
- Volloch V, Schweitzer B, Rits S. Antisense globin RNA in mouse erythroid tissues: structure, origin, and possible function. *Proc Natl Acad Sci U S A* 1996;93:2476–2481. [PubMed: 8637899]
- Voorhoeve PM, le Sage C, Schrier M, et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Adv Exp Med Biol* 2007;604:17–46. [PubMed: 17695719]
- Vu TH, Chuyen NV, Li T, Hoffman AR. Loss of imprinting of IGF2 sense and antisense transcripts in Wilms' tumor. *Cancer Res* 2003;63:1900–1905. [PubMed: 12702581]
- Wang H, Iacoangeli A, Lin D, et al. Dendritic BC1 RNA in translational control mechanisms. *J Cell Biol* 2005;171:811–821. [PubMed: 16330711]
- Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 2001;27:422–426. [PubMed: 11279525]
- Washietl S, Hofacker IL, Lukasser M, et al. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 2005;23:1383–1390. [PubMed: 16273071]
- Watanabe T, Takeda A, Tsukiyama T, Mise K, et al. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* 2006;20:1732–1743. [PubMed: 16766679]
- Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520–562. [PubMed: 12466850]
- Weinberg MS, Villeneuve LM, Ehsani A, et al. The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* 2006;12:256–262. [PubMed: 16373483]
- Willingham AT, Gingeras TR. TUF love for “junk” DNA. *Cell* 2006;125:1215–1220. [PubMed: 16814704]
- Willis RC, Hogue CW. Searching, viewing, and visualizing data in the Biomolecular Interaction Network Database (BIND). *Curr Protoc Bioinformatics Chapter 8:Unit* 2006;8:9.
- Wrobel G, Primig M. Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction. *Reproduction* 2005;129:1–7. [PubMed: 15615893]
- Wu SM, Baxendale V, Chen Y, et al. Analysis of mouse germ-cell transcriptome at different stages of spermatogenesis by SAGE: biological significance. *Genomics* 2004;84:971–981. [PubMed: 15533714]
- Wutz A, Gribnau J. X inactivation Xplained. *Curr Opin Genet Dev* 2007;17:387–393. [PubMed: 17869504]
- Xenarios I, Salwinski L, Duan XJ, et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30:303–305. [PubMed: 11752321]
- Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* 2002a;30:3754–3766. [PubMed: 12202761]
- Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucl Acids Res* 2002b;30:3754–3766. [PubMed: 12202761]
- Yagi S, Hirabayashi K, Sato S, et al. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res* 2008;18:1969–1978. [PubMed: 18971312]
- Yamasaki K, Joh K, Ohta T, et al. Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Hum Mol Genet* 2003;12:837–847. [PubMed: 12668607]
- Yasuda J, Hayashizaki Y. The RNA continent. *Adv Cancer Res* 2008;99:77–112. [PubMed: 18037407]

- Yelin R, Dahary D, Sorek R, et al. Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* 2003;21:379–386. [PubMed: 12640466]
- Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol* 2004;5:R74. [PubMed: 15461793]
- Yin H, Lin H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* 2007;450:304–308. [PubMed: 17952056]
- Yu Z, Guo R, Ge Y, et al. Gene expression profiles in different stages of mouse spermatogenic cells during spermatogenesis. *Biol Reprod* 2003;69:37–47. [PubMed: 12606389]
- Zhang Z, Gerstein M. Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev* 2004;14:328–335. [PubMed: 15261647]
- Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003;13:2541–2558. [PubMed: 14656962]

Figure 1.

Using SAGE to detect alternative spliced transcripts. Different splicing of pre-mRNA at the 3' end leads to generation of various gene products at different stages of spermatogenesis. Exons are represented by different colored bars. In this example, type A spermatogonia (Spga), pachytene spermatocytes (Spcy) and round spermatids (Sptd) express different gene isoforms due to differential usage of polyadenylation signals at the 3' untranslated region, which are detected by different SAGE tags mapped to the corresponding isoform. The isoforms were then validated by RT-PCR. The inclusion or exclusion of coding information after splicing may lead to different functional outcomes.

Figure 2.

Flow chart for the identification of stage-specific splicing variants by SAGE and bioinformatics approaches. SAGE tags with polyA evidence were aligned to the Unigene clusters by SAGEmap. Sense transcript with more than one tag mapped suggests the presence of alternative splicing variants at the 3' end of the transcript.

Figure 3.

Structural groups of antisense amplicons. Solid and empty bars represent antisense transcript. Hedged bars represent sense gene sequence. Homology is indicated by vertical lines. Thick dark line indicates how the exons are spliced together. Number of genes in each group is shown in bracket.

Figure 5.

Putative function of antisense transcripts derived from pseudogenes. Brown arrow represents antisense transcript; yellow and gray arrows represent sense transcripts. Yellow bar represents functional gene; gray bar represents pseudogene. Two potential modes of interaction between pseudogene antisense transcript and functional transcript are described. In Direct Interaction mode, antisense transcript from the pseudogene or functional gene forms RNA-duplex with the functional transcript. This may block the translation of the mRNA or results in RNA interference subsequent to the formation of double stranded small interfering RNA (siRNA). In Indirect Interaction mode, antisense and sense transcript of a pseudogene form RNA-duplex resulting in the formation of double stranded siRNA leading to RNA interference of the functional mRNA.

Figure 6.

Different classes of non-coding RNAs. Classification of ncRNA species by size. Small ncRNA is below 30 bp while long ncRNA is above 200 bp. miRNA: microRNA; piRNA: Piwi RNA; snRNA: small nuclear RNA; snoRNA: small nucleolar RNA; rRNA: ribosomal RNA; tRNA: transfer RNA; ncRNA: non-coding RNA.

Figure 7.
The overview of systems biology workflow.

Figure 8. Biological network in type A spermatogonia (A), pachytene spermatocytes (B) and round spermatids (C). SAGE data of developing mouse male germ cells were used to construct specific gene network at different stage of spermatogenesis. Input gene set are labeled in blue circle. Objects with known drug target are marked in gray, the name of compounds for the target are listed next to it.

TABLE 1**Genes Exhibiting Alternative Splicing in Male Germ Cells**

Gene	Suggested function	Expression pattern	References
Hyaluronan binding protein 1	Spermatogenic differentiation	Present in pachytene spermatocytes and round spermatids but not in elongated spermatids	Bharadwaj et al., 2002
Lamin B3 and C2	Reorganization of nuclear and chromosomal structures during meiotic division	B3 is generated by splicing from lamin B2 and present in spermatogenic cells, whereas C2 is present in spermatocytes only	Furukawa et al., 1993.
RNA-binding motif 7	Post-meiotic germ cell mRNA metabolism	RBM7 appears in the nucleus up to the mid pachytene stage and becomes cytoplasmic afterwards	Guo et al., 2003
Gametogenetin 1 and 3	Regulated POG for normal proliferation of PGCs	Both present in primary spermatocytes	Lu and Bishop, 2003.
Estrogen receptor beta	Unknown	Some isoforms present in spermatocytes	Aschim et al., 2004.
BRCA1 associated RING domain 1 beta	Induce apoptosis in meiotic spermatocytes and early round spermatids	Present in spermatocytes and round spermatids	Feki et al., 2004.
cAMP-responsive element modulator	Modulation expression of cAMP responsive genes	Present in spermatocytes and round spermatids	Foulkes et al., 1992
Type 1 hexokinase	Metabolize glucose in glycolytic pathway	Three germ cell-specific isoforms. Present in meiotic and post-meiotic germ cells	Jaiswal and Conti, 2001.
Mannose 6-phosphate receptor	Mediates transport of acid hydrolyases to lysosomes	Three isoforms: one in somatic cells and two in pachytene spermatocytes and round spermatids	Belmonte et al., 2002.

TABLE 2

Useful Databases in Systems Biology

Databases	Website address	Features
Primary		
NCBI	http://www.ncbi.nlm.nih.gov/	Genome and sequence information, expression data, tools
EMBL-EBI	http://www.ebi.ac.uk/ensembl/	Genome and sequence information, expression data, tools
GO	http://www.geneontology.org/	Gene ontology
miRBase	http://microrna.sanger.ac.uk/	MicroRNA data
RNAdb	http://research.imb.uq.edu.au/rnadb/	Non-coding RNA data
Pathway and interaction		
BioCarta	http://www.biocarta.com/	Charting pathways of life
KEGG	http://www.genome.jp/kegg/	Information on genomes, enzymatic pathways and biological chemicals
Reactome	http://www.reactome.org/	Information on core human pathways
PID	http://pid.nci.nih.gov/	Pathway interaction database contains human biomolecular interactions and cellular processes
BIND	http://bind.ca	Biomolecular Interaction Network Database is the largest manually-curated molecular interaction database
DIP	http://dip.doe-mbi.ucla.edu/	Database of Interacting Proteins contains information on experimentally validated protein interactions
Theme based		
GermSAGE	http://germsage.nichd.nih.gov/	SAGE data on male germ cell development
GermOnline	http://www.germonline.org/	Expression data related to germ cell research
MGI	http://www.informatics.jax.org/	Dedicated resources for mouse genome research
COSMIC	http://www.sanger.ac.uk/genetics/CGP/cosmic/	Catalogue Of Somatic Mutations In Cancer