Review

Testicular postgenomics: targeting the regulation of spermatogenesis

Pierre Calvel[†], Antoine D. Rolland[†], Bernard Jégou
and Charles Pineau*

Inserm, U625, IFR 140, University of Rennes I, Campus de Beaulieu, Rennes 35042, France

Sperm are, arguably, the most differentiated cells produced within the body of any given species. This is owing to the fact that spermatogenesis is an intricate and highly specialized process evolved to suit the individual particularities of each sexual species. Despite a vast diversity in method, the aim of spermatogenesis is always the same, the idealized transmission of genetic patrimony. Towards this goal certain requirements must always be met, such as a relative twofold reduction in ploidy, repackaging of the chromatin for transport and specialized enhancements for cell motility, recognition and fusion. In the past 20 years, the study of molecular networks coordinating male germ cell development, particularly in mammals, has become more and more facilitated thanks to large-scale analyses of genome expression. Such postgenomic endeavors have generated landscapes of data for both fundamental and clinical reproductive biology. Continuous, large-scale integration analyses of these datasets are undertaken which provide access to very precise information on a myriad of biomolecules. This review presents commonly used transcriptomic and proteomic workflows applied to various testicular germ cell studies. We will also provide a general overview of the technical possibilities available to reproductive genomic biologists, noting the advantages and drawbacks of each technique.

Keywords: testis; spermatogenesis; transcriptome; proteome

1. INTRODUCTION

Among the various cell types that comprise a mammalian organism, sperm, or spermatozoa, are the subjects of a unique and, from an evolutionary standpoint, seminal destiny. To the parental body they are entirely dispensable and, indeed, are freely dispensed to serve their role of genomic transmission intercorporeally. The importance of that role is such that the genes which contribute to spermatogenesis are under particular selective pressure. This permits certain extraordinary features of spermatogenesis, such as high degrees of germ cell-specific gene expression and a low degree of correlation between gene and protein expression to be maintained. Spermatogenesis features the germ cell-specific events associated with meiosis, as well as certain unique events related to chromatin remodelling, repackaging and transcriptional reprogramming. Moreover, spermatozoa bear hyperselected traits associated with shape and energy production, as well as enhanced flagellum development. Control of this unique differentiation incorporates juxtacrine, paracrine and endocrine factor information, and is conditioned by the successive activation and/or repression of thousands of genes and

proteins, including numerous testis-specific isoforms. All these features contribute to make the testis, together with brain, one of the most complex tissues in the body. Unravelling aspects of this complexity using cell culturing approaches has been problematic, owing to difficulties in culturing more differentiated male germ cell-types. The absence of this ‘testing ground’ places greater emphasis on quality genomic analysis in advance of *in vivo* testing of hypotheses.

Advances in molecular biology and genomics have improved our knowledge of spermatogenesis by identifying numerous genes essential for the development of functional male gametes (for reviews, see Matzuk & Lamb 2002; de Rooij & de Boer 2003). Indeed, significant progress has been made in the large-scale analysis of testicular function, enabling a more profound insight into normal and pathological spermatogenesis. Several laboratories have built on rapid progress in genome sequencing and microarray development, carrying out genome-wide expression studies leading to the identification of hundreds of genes spatially and temporally regulated during the ontogenesis of the testis (for review see Wrobel & Primig 2005). On the other hand, the development of tools for high-throughput protein identification has allowed a few laboratories to perform differential and/or systematic analysis of testicular proteomes from various species, either on the entire organ (Huang *et al.* 2005) or on isolated cells (Com *et al.* 2003; Essader *et al.* 2005; Rolland *et al.* 2007).

* Author for correspondence (charles.pineau@rennes.inserm.fr).

[†] These authors contributed equally to this work.

One contribution of 17 to a Theme Issue ‘The biology and regulation of spermatogenesis’.

Figure 1. (Caption opposite.)

This chapter will review the current state of large-scale gene and protein expression analysis of spermatogenesis from germ cells development during sex determination up to sperm cells maturation and capacitation. The advantages and limitations of transcriptomics and proteomics in the context of studies on the testicular germ cells expression programme will be reviewed. Finally, the concept of systems biology, which involves integrative ‘omics’ (i.e. combining

genomics, transcriptomics, proteomics) as well as bioinformatics and modelling, are discussed.

2. GENE EXPRESSION PROFILING TECHNOLOGIES: THE UNDERLYING DIFFERENCES

Although various technologies are now available to study gene expression at the mRNA and protein

Figure 1. (*Opposite.*) The basics of the transcriptomic and proteomic technical workflows. (a) Transcriptomics: the high-scale analysis of gene expression can be operated through two distinct strategies. In sequencing-based approaches such as SAGE and CAGE, small DNA fragments (10–14 bp) are generated from cDNAs and subcloned in a DNA library, in which the abundance of each tag is correlated to the level of expression of the corresponding gene. The high-throughput sequencing of libraries corresponding to various samples gives access to quantitative data on genes expression. On the other hand, oligonucleotide microarrays are based on the ability of complementary DNA or RNA fragments to hybridize together. Two classes of arrays are frequently used: in spotted DNA microarrays, micro droplets of long complementary DNA fragments are spotted on a chip. The cDNAs prepared from two samples are labelled with fluorescent dyes and hybridized simultaneously on the chip. The ratio of the two fluorescences emitted by each spot will determine the relative expression of the corresponding gene in the two samples. With the *in situ* synthesized microarrays, two sets of 11 25 bp-probes are directly synthesized for each gene on the chip. For each sample, biotinylated cRNAs are produced from cDNAs and hybridized individually on the chip. The RNAs' abundance is revealed by a peroxidase assay. (b) Proteomics: the first strategy dealing with the high-throughput analysis of proteins aims to systematically identifying the protein content of a biological sample. Classically, proteins extracted from a whole organ or isolated cells are separated by two-dimensional-gel electrophoresis. Each detectable spot is then excised from the gel and submitted to a trypsin digestion. The resulting peptides are listed by MALDI-MS mass spectrometry and the corresponding protein is identified thanks to the peptide mass fingerprint (PMF) technique. Recently, the progress made with highly sensitive mass spectrometry enabled scientists to avoid the complex two-dimensional separation process. A whole trypsin-digested proteic sample can thus be separated by liquid chromatography and directly injected in an ESI-MS, where the peptides will be identified by tandem mass spectrometry (MS/MS). An intermediate way is to prefractionate the samples through one-dimensional-gel electrophoresis. Bands containing hundreds of proteins are then sliced off the gel and trypsin-treated before their identification by LC-ESI-MS/MS sequencing. Quantification can also be considered with proteomics, either by electrophoresis-based or MS-based methods. Following technical principles similar to those used in spotted oligonucleotide microarrays, the two-dimensional difference in gel electrophoresis (2D-DIGE) relies on the simultaneous analysis of two differentially cyanine-labelled samples. Protein spot exhibiting asymmetrical fluorescence can be identified by PMF. Differential analysis can also be performed with quantitative mass spectrometry. In the case of isotope-coded protein label technique (ICPL), samples can be differentially labelled with light or heavy isotopes and digested together prior to MS analysis. The determination of the ratio between the intensity of a light and heavy labelled peak on the mass spectra followed by MS/MS protein identification allows the determination of the protein relative expression level within the two samples. Other quantification techniques based on MS analysis are available (ICAT, ITRAQ), but they require post-trypsin digestion labelling procedures.

levels, fundamental differences can be pointed out and will be summarized below (figure 1).

(a) *Transcriptome and transcriptomics*

The transcriptome can be defined as the complete set of ribonucleic acid (RNA) transcripts resulting from the specific expression of a genome in a given cell type, following the influence of spatial, temporal and environmental parameters. Transcriptomics describe the tools and techniques applied to the global analysis of genome expression. With the underlying idea that genes which are simultaneously expressed or repressed can be co-regulated and functionally related, this global approach aims to depict precisely the dynamic of genes regulation in the cell. The widespread technologies for genome-wide analysis of gene expression include either spotted and oligonucleotides microarrays, developed with sequences *a priori*, or de novo sequencing-based approaches such as SAGE and CAGE. These various transcriptomic technologies now allow scientists to study tens of thousands of genes at once instead of working on a gene-by-gene basis (figure 1a).

(b) *Proteome and proteomics*

The word 'proteome' was first coined by Marc Wilkins at the 4th Sienna meeting for two-dimensional gel electrophoresis in 1994 (Wilkins et al. 1996) and is a portmanteau of 'protein' and 'genome'. The term encapsulates the complex and dynamic nature of protein expression at the points of reference that can span from cell to organism. Whereas genomes are essentially invariant in different cells in an organism, proteomes, as well as transcriptomes, vary from cell

to cell, according to time, environmental stimuli and/or stress.

Proteomics is the research area revealing the temporal dynamics of proteins expressed in a given biological compartment at a given time. The definition has until recently covered proteins as gene products. Recently, there has been an alteration of the proteomics definition to include not only gene products expression, but also the structural alteration and the chemical modifications of these gene products occurring in cellular metabolisms and turnover (i.e. post-translational modifications; Aebersold & Mann 2003). The science of proteomics, one of the most important areas of research in the post-genomic era, is not new in terms of experimental foundations. It has, nonetheless, benefited from unprecedented advances in genome sequencing, bioinformatics and the development of robust, sensitive, reliable and reproducible analytical techniques (figure 1b).

(c) *Transcriptomics versus proteomics, when your heart balances*

The recent completion of the first high-quality drafts of the mouse (Waterston et al. 2002) and human (Lander et al. 2001; Venter et al. 2001) genome, has provided scientists with access to a most valuable range of relevant sequence information necessary to the functional characterization of gene products in a systematic and comprehensive manner. A surprising finding of these genome projects is that there are far fewer protein-coding genes in mammalian genomes than expected. Approximately 22 000 genes have indeed been found in both genomes (Ensembl

release 51; Hubbard *et al.* 2009) which would correspond to 22 000 functional products when considering the original dogma of molecular biology. However by alternative splicing, 22 000 genes can routinely encode for 100 000 proteins (Modrek *et al.* 2001; Johnson *et al.* 2003). Adding post-translational modifications (e.g. phosphorylation, glycosylation and proteolysis; Kettman *et al.* 2002; Mann & Jensen 2003), the estimated number of protein species in humans can reach as many as one million (Humphery-Smith 2004; Mueller *et al.* 2007) each having different functions. This complexity of the multi-layered gene expression mechanisms is partially responsible for the discrepancy often observed between abundances of mRNA and proteins (Gygi *et al.* 1999; Chen *et al.* 2002; Conrads *et al.* 2005). Although transcriptomics most probably still defeats proteomics in terms of throughput capacities, it is nevertheless quite clear that protein diversity cannot be fully characterized by gene expression analyses alone. Much care must then be taken before undertaking such large-scale experiments depending on whether one's biological question is more closely related to transcriptional and/or splicing mechanisms underlying a particular process or to those proteins and their subtle different isoforms directly involved in this same process.

(d) *Expression profiling of the embryonic germ cells*

The embryonic gonads arise in mammals within the intermediate mesoderm as undifferentiated and bipotential structures. The transient expression of a single gene named *Sry* (sex-determining region, chromosome Y) in the somatic supporting cells of the genital ridges triggers the sexual differentiation of the gonads into testis. Under the influence of *Sry*-expressing pre-Sertoli cells, the primordial germ cells (PGCs) then develop into prospermatogonia, the precursor cells of the male germline (for review, see Wilhelm *et al.* 2007).

If the sexual determination of the somatic cells of the embryonic gonads has been thoroughly studied by oligonucleotide microarrays (Nef *et al.* 2005; Small *et al.* 2005; Beverdam & Koopman 2006; Bouma *et al.* 2007) or proteomic-based approaches (Wilhelm *et al.* 2006), little has been achieved so far to decipher the biology of the PGCs at the post-genomic level, owing to their restricted number in the embryo. Han *et al.* (2005) performed a proteomic analysis of cultured PGCs from chicken which revealed the expression of 50 proteins including growth factors and developmentally regulated proteins. Nevertheless, no such proteomic study of PGCs has been carried out so far in mammals. Recently, the use of single-cell microarrays-based strategies has bypassed the problem of PGCs' limited number and allowed the analysis of the developing mouse PGCs prior to their migration in the genital ridges (Yabuta *et al.* 2006; Kurimoto *et al.* 2008a,b). These genome-wide expression analyses have provided consistent clues about the genetic programmes that drive the emergence of the PGCs from the epiblast, and have identified *Blimp1* (B-lymphocyte-induced

maturation protein-1) and *Prdm14* as two determinant factors for the initiation of this process. Mise *et al.* (2008) also carried out a genome-wide comparison of gene expression profiles between male and female 11.5–13.5 dpc mouse PGCs, embryonic stem cells (ESCs), embryonic germ cells (EGCs) and germline stem cells (GS). They were able to give a subset of 97 genes that constitute the molecular signature of PGCs compared with other stem cells type and showed that PGCs are unique cells, in which pluripotency markers such as *Oct 3/4* and *Nanog*, and specific germ cell lineage markers important for germ cells development (*Dazl*, *Fkbp6* and *Asz1*) coexist. They also found that genes involved in DNA methylation such as *Dnmt3b* and *Dnmt3l* were underrepresented within 11.5 and 13.5 dpc PGCs, suggesting a particular epigenetic status for these cells compared with other stem cells. In contrast to this latter study, Lefevre & Mann (2008) showed that these genes encoding for DNA-methyltransferases and strongly repressed at 11.5 and 13.5 dpc were specifically overexpressed in 15.5 dpc mouse prospermatogonia, together with several histone-demethylases. This suggests that the transition between the pluripotency and the differentiated state of the male PGCs implied a rapid epigenetic reprogramming both at the DNA and histone levels.

(e) *The spermatogonial stem cell niche:*

reinvestigation of an old concept with new tools

The spermatogonial stem cells (SSCs) of the adult testis persist during the whole life of the male. These cells, which derived from the PGCs, are a striking example of stem cells pluripotency, as illustrated by their ability to differentiate through a series of mitotic divisions into pre-meiotic spermatogonia and to maintain their number by a subtle balance between self-renewing processes and apoptosis events (de Rooij 2001; Oatley & Brinster 2008). Despite their fundamental role in the initiation of spermatogenesis, the molecular mechanisms underlying the maintenance of their diverse functions remained largely unknown until the end of the last century. The emerging state-of-the-art technologies dedicated to the global analysis of gene and protein expression has recently allowed scientists to better understand the molecular signature of spermatogonia and of its cellular environment, called the SSC niche.

The development of culture systems in rodents has made it possible to carry out gene expression profiling on spermatogonial cells in various developmental states (figure 2). Hamra *et al.* (2004) used enriched preparations of rat SSCs cultured on different feeder cell lines to identify genes associated with the maintenance (in MSC-1 cells) or loss (in STO cells) of stem cell activity. As many as 248 genes were found to be downregulated during the loss of stem cell activity, their level of transcription remaining stable while this activity was maintained. These genes are therefore probably involved in self-renewal rather than differentiation. Interestingly, very few of the corresponding gene products were found to be enriched in neural, hematopoietic or ESCs, suggesting that the

mechanisms implied in the maintenance of SSC activity might be different. The authors then focused on a subset of 115 genes for which mouse *homologues* also displayed downregulation during germ cell differentiation *in vivo*, including *Bcl6b* (see below Oatley *et al.* 2006). Mean expression levels for these genes, referred to as the 'stem cell index', were strongly correlated with SSC activity, and with the expression of individual genes, such as *Erg3*, expressed only in undifferentiated stem cells.

Oligonucleotide microarrays were also extensively used to study specific pathways related to single factors involved in SSCs' self-renewal and spermatogonial differentiation. *Zfp145*-null mice lack a transcriptional repressor specifically expressed in spermatogonia in the testis (promyelocytic leukemia zing finger, PLZF) and show a progressive loss of spermatogonia with age. Microarray gene expression analysis of isolated spermatogonia ($\alpha 6$ -integrin-positive cells) from one-week-old *Zfp145*-null mice identified more than 230 genes as differentially expressed with respect to the wild-type (Costoya *et al.* 2004). Many of these genes are known to be involved in the control of cell cycle or the differentiation of spermatogonia, suggesting that absence of PLZF in these cells may alter the tight balance between self-renewal and differentiation, thus resulting in increased apoptosis in spermatogonia. Chen *et al.* (2005) also investigated the impact of the sertolien factor Ets-related molecule (ERM) on the SSC niche maintenance. Invalidation of *Erm* gene in the mouse results in a progressive loss of germ cells during the first wave of spermatogenesis and to a Sertoli-cell-only syndrome. Using total testis from four-week-old mutant mice, an age at which no obvious phenotype was detected, they show that the expression of many spermatogonial genes was altered, suggesting a strong influence of ERM on the SSC/spermatogonial expression programme. They also showed that a large number of genes were severely repressed in isolated Sertoli cells in the absence of ERM, some of them encoding for chemokines (CXCL-12, CXCL-5 and CCL7) previously described as regulators of the haematopoietic stem cell niche. The SSC self-renewal and differentiation pathways mediated by one such factor, the glial cell line-derived neurotrophic factor (GDNF), have also been analysed by gene expression profiling experiments (Hofmann *et al.* 2005; Oatley *et al.* 2006). Hofmann *et al.* sorted GRF α -1 (the GDNF coreceptor)-positive germ cells from 6-day-old mice after isolation by sedimentation under gravity and differential plating. They monitored the gene expression profiles of these cells cultured in the presence or absence of GDNF for 10 h (Hofmann *et al.* 2005). They identified 378 genes overexpressed in the presence of GDNF including numerous cell cycle and cell proliferation regulators. The authors then focused on one of these genes, encoding fibroblast growth factor receptor 2 (FGFR2), whose upregulation indicates that GDNF could render germ-line stem cells responsive to bFGF. Indeed, addition of bFGF into the culture medium greatly amplifies the effect of GDNF alone on the gene expression profile of spermatogonia, suggesting a cooperative effect of these two factors on

spermatogonial self-renewal or proliferation. With the aim of identifying new factors contributing to GDNF action in SSCs, the same team used an identical strategy to compare gene expression between GRF α -1-positive and GRF α -1-negative SSCs (Kokkinaki *et al.* 2009). Among the 99 transcripts showing a twofold increase in GRF α -1⁺ cells, a transcript encoding the receptor for macrophage colony-stimulating factor (Csf1r) showed the highest level of overexpression. When incubated with CSF1, spermatogonia underwent proliferation, suggesting a cooperative effect between GDNF and CSF-1. This hypothesis was strengthened by an *in silico* pathway analysis which showed that both GRF α -1 and CSF1-R could trigger the MAP kinases pathway that leads to cell proliferation. In order to refine the precise timetable of GDNF-regulated genes expression in spermatogonia, Oatley *et al.* (2006) cultured Thy1-positive cells on STO cells (which do not maintain SSC activity) and analysed by oligonucleotide microarrays the effect of a 18 h-GDNF/GRF α -1 withdrawal on the biology of SSCs. One hundred and ninety-nine genes were identified as downregulated by a GDNF/GRF α -1 deprivation ('self-renewal'-associated genes) whereas 79 genes were upregulated ('differentiation'-associated genes). Thereafter, they monitored a time-course analysis of SSC genes expression following the reintroduction of GDNF/GRF α -1 in the medium (2, 4 and 8 h after reintroduction). Interestingly, 67 per cent of the 193 transcripts that were upregulated by GDNF/GRF α -1 reintroduction were seen at 4 h, whereas 83 per cent of the 63 downregulated transcripts were seen 8 h after this treatment. It should be noted that GDNF withdrawal experiments driven on cultured rat SSC showed similar effects on the SSC gene expression programme, indicating strong correspondences between species (Schmidt *et al.* 2009). The authors then focused on the transcriptional repressor *Bcl6b*, one of the six genes that were both downregulated after GDNF/GRF α -1 withdrawal and upregulated following GDNF reintroduction. *In vitro* knockdown of *Bcl6b* by siRNA experiments showed a strong decrease in the maintenance of SSC activity and an increased apoptosis. These results were consistent with the dramatic impairment of spermatogenesis observed in *Bcl6b*-null mouse testis where many seminiferous tubules presented Sertoli-cell-only syndrome (Oatley *et al.* 2006).

Whereas the above studies have evidenced numerous factors involved in SSCs and spermatogonia signalling pathways, they also highlighted the fact that such factors are often important to other stem cell lineages. This suggests that the expression of a specific combination of factors with its own specific regulation scheme rather than the expression of germ-line specific factors accounts for the identity and unique capacities of SSC to promote spermatogenesis.

The progress of mature spermatogonia towards meiotic division has been also recently studied by Rossi *et al.* (2008), who focused on the effect of Kit ligand (KL) on the mouse spermatogonia differentiation. Using DNA microarrays, they demonstrated that a significant number of genes encoding for known spermatogonial markers, such as *Egr2/3* or *Uty*, were strongly repressed in 7-day-old cultured

Figure 2. Pathways for the self-renewal and differentiation of spermatogonial stem cells (SSC). The daily production of millions of spermatozoa in mammals is ensured by the presence within the male germ line of stem cells able to maintain their own stock and to differentiate and continuously initiate new waves of spermatogenesis. The balance between the maintenance of spermatogonial stem cells, their proliferation and their differentiation into mature spermatogonia is triggered by intrinsic factors, such as PLZF, and by signals from the spermatogonial stem cell niche, including Sertolian signals, such as GDNF, ERM or Kit ligand. GDNF: glial cell line-derived neurotrophic factor; ERM: Ets-related molecule; PLZF: promyelocytic leukemia zinc finger. References: 1, Chen *et al.* (2005); 2, Hofmann *et al.* (2005); 3, Oatley *et al.* (2006); 4, Kokkinaki *et al.* (2009); 5, Schmidt *et al.* (2009); 6, Hamra *et al.* (2004); 7, Costoya *et al.* (2004); 8, Guillaume *et al.* (2000); 9, Com *et al.* (2003); 10, Rossi *et al.* (2008); 11, Zhou *et al.* (2008).

spermatogonia incubated 24 h with KL, whereas many transcripts, some corresponding to early-meiotic markers such as *lhx8*, were upregulated. These results clearly confirm that the interaction between Sertoli cell-secreted KL and c-kit promotes the meiotic differentiation of spermatogonia. A microarray profiling of Thy1-positive gonocytes prepared from Retinoic acid (RA)-treated 2 dpp mouse testes also showed a marked induction of genes involved in RA metabolism, some being known as early meiotic markers such as *stra8* (Zhou *et al.* 2008). Proteomic technologies have also strongly impacted our knowledge of the developing spermatogonia. With the view of gaining insights into the biology of spermatogonial cells, our group established reference proteome maps of mature spermatogonia freshly isolated from rat 9-dpp testes (Guillaume *et al.* 2000). Another study provided information about the low-copy number proteins expressed in these cells through the prefractionation of protein

cell extracts on two-dimensional gels with a narrow pH range (Com *et al.* 2003). These two studies led to the first time identification in spermatogonia of several proteins whose roles during the spermatogenic process were further investigated by our group afterwards (Guillaume *et al.* 2001a,b; Com *et al.* 2006). Global approaches of this type should pave the way to the comprehension of the complex protein networks that drive the spermatogonia's biology. Indeed, the availability of genome sequence data has generated an urgent need for systematic protein identification for elucidation of the encoded protein networks governing cellular function. Large-scale protein–protein interaction maps have generally been based on results obtained with the yeast two-hybrid system, which detects only binary interactions (for review, see Ito *et al.* 2001). However, the advent of highly sensitive protein identification methods based on mass spectrometry has made it feasible to identify protein

complexes directly, at the proteome-wide scale (for review, see Charbonnier *et al.* 2008).

(f) *Post-natal development of the male germ cells*

Among the many components implied in the enhancement of male fertility, the way spermatozoa are produced throughout spermatogenesis remains the most studied part of male reproduction through large-scale experiments. Indeed, spermatozoa development proceeds from a succession of cellular events of which the precise timetable is regulated through the expression/repression of many molecular actors. Thus, applying transcriptomics and proteomics to the study of spermatogenesis makes sense since these techniques can produce very precise snapshots of the molecular networks operating at each step of the process.

Many groups have adopted basic strategies to address this issue, by carrying out systematic characterization of genes and proteins expressed in isolated germ cells at a given time of their development or in total testis (figure 3). These studies include systematic identification of germ cells chromatin-associated proteins in *Caenorhabditis elegans* (Chu *et al.* 2006), insoluble chromatin-associated proteins in mouse elongated spermatids (Govin *et al.* 2006), testicular proteins in the fruitfly (Takemori & Yamamoto 2009), the pig (Huang *et al.* 2005) or the mouse (Zhu *et al.* 2006) as well as expression analyses of the germline genes in *C. elegans* (Reinke *et al.* 2000), and in the testis of the mouse (Yao *et al.* 2004; Divina *et al.* 2005) or human (Fox *et al.* 2003). Large datasets including genes and proteins known to be important for testicular function and male fertility re-emerged from these studies. However, the absence of reliable information about their spatial and temporal regulation during spermatogenesis constitutes the main drawback of these approaches. More sophisticated strategies were then used to get deeper snapshots of spermatogenesis, especially by carrying differential expression analysis of genes and proteins. The first strategy adopted by several groups aimed at comparing the expression of genes or proteins between different categories of purified germ cells. For example, the proteome analysis of rat spermatogenesis by a differential approach using the two-dimensional difference in-gel electrophoresis (2D-DIGE) technique was carried out by our group (Rolland *et al.* 2007). This study aimed to identify proteins that were specifically expressed in spermatogonia, pachytene spermatocytes or early spermatids. Among the 123 proteins displaying very high differential expression (more than 2.5), several were identified for the first time in the male germline and further characterized by targeted studies, providing a more precise picture of spermatogenesis. As an example, the spermatid-specific CLPH (Casein-like phosphoprotein) was evidenced following this study as a calcium-binding protein phosphorylated by casein kinase 2, one of the major contributor of sperm head development (Calvel *et al.* 2009). Comparative studies involving purified germ cells were also carried by SAGE experiment in the mouse (Wu *et al.* 2004) or GeneChip microarray experiments in the mouse

(Shima *et al.* 2004; Namekawa *et al.* 2006; Chalmel *et al.* 2007b) and in the rat (Schlecht *et al.* 2004; Chalmel *et al.* 2007b; figure 3). The second strategy, which aims to compare total testis samples from animals of various ages during the first wave of spermatogenesis, was conducted in the mouse through a differential display analysis (Almstrup *et al.* 2004), spotted PCR microarray experiments (Ellis *et al.* 2004; Clemente *et al.* 2006) and GeneChip microarray experiments (Schultz *et al.* 2003; Shima *et al.* 2004; figure 3). As both strategies bring about advantages and drawbacks, great care should be taken by the experimenter when defining both the starting biological materials and the method of analysis: the use of isolated populations of germ cells make the identification of low-copy number transcripts or proteins more efficient, especially in the case of proteomic studies where no amplification step can be performed. However, the cell purification protocols are generally time-consuming and stressful, thus contributing probably to slight alterations in the general expression profile of the cells. On the other hand, working on total testis at different periods of its development allows the analysis of multiple time points, and can also avoid the generation of artifactual expression signals due to sample preparation. Nevertheless, this method faces the impossibility of establishing for sure the cellular origin of a given expression signal. Indeed, during the post-natal development of the testis and the progressive apparition of new germ cells types, not only germ cells are responsible for the expression changes observed in such experiments, but also the different somatic cells, which exhibit developmental patterns of expression during this period (O'Shaughnessy *et al.* 2003; Ge *et al.* 2005). The potentialities of these two different methodologies were simultaneously addressed by Shima *et al.* (2004), who first analysed gene expression at 11 time points of the ontological development of the mouse testis (from birth to 56 days post partum) and secondly identified transcripts specifically expressed in freshly isolated mouse testicular cells (peritubular cells, Sertoli cells, type A and B spermatogonia, pachytene spermatocytes and round spermatids). They found that the expression profiles arising from their developmental time course was consistent with the cellular origin of most of the transcripts. For example, genes specifically expressed in the isolated somatic cells of the testis showed a decreasing signal during testis development, due to the progressive appearance of germ cells that overcomes the global population of somatic cells and generates a 'dilution' effect. Consistently, the pool of transcripts that showed an increasing signal towards adulthood was mainly expressed in meiotic or post-meiotic cells. This work cleverly demonstrated the relevance of both strategies to perform high-scale expression analysis of the testis, considering that neither the alteration of gene or protein expression during isolation procedures nor the diversity of cell types in total testis samples appeared as a hindrance during the monitoring of spermatogenesis.

Correlation between expression profiling experiments has also been reported, not only between purified germ cells and total testis samples, but

among species. With the aim of highlighting evolutionarily-conserved and testis-specific genes between human and rodents, we undertook the comparison of mouse, rat and human spermatogenesis using oligonucleotide microarrays (Chalmel *et al.* 2007b). We identified several thousands of genes as being differentially expressed among the different categories of purified germ cells, out of which one thousand orthologues could be identified. On the basis of a high correlation between rodent profiles, 888 orthologous genes were shown to display similar expression profile along spermatogenesis in the three species, thus constituting a part of the core expression programme common to human and rodents. Considering that the high reproducibility of commercial GeneChips allows the implementation of external data to one's own analysis, we compared our data with those available for 17 mouse somatic tissues. We found that very few testis-specific genes were actually found in spermatogonia or Sertoli cells, strongly suggesting that the functional identity of these cells preferentially relies on the coordinated action of widely expressed factors rather than the expression of very specific genes. These results illustrate the potentialities of such *trans*-species genome-wide expression analysis to identify genes whose function is determinant to achieve mammalian spermatogenesis.

Apart from these analyses of testis and germ cell expression patterns, several studies have investigated the existing correlations between transcriptomic and proteomic data trying to take translation regulation into account (figure 3). Cagney *et al.* (2005) developed a human tissue-profiling experiment of nuclear protein-enriched extracts using multidimensional protein identification technology (MudPIT). This combination of cation-exchange and reverse-phase chromatographic separation of tryptic digests before mass spectrometric identification allowed them to identify as many as 1713 non-redundant proteins across the eight tissues studied. The expression profiles of 683 of them were then unambiguously compared with those obtained in microarray experiments (Cagney *et al.* 2005). Although a difference of sensitivity between the methods might bias the comparison of data, the authors found that out of the eight tissues studied, the testis exhibited the lowest correlation coefficient between the transcriptome and the proteome (i.e. 0.138), as compared with that of the liver (i.e. 0.432). Such a low correlation could be the consequence of peculiar properties of gene/transcript regulation occurring during spermatogenesis, such as mRNA storage in translationally repressed free ribonucleoprotein (RNP) particles or delays between transcription and translation. This issue was specifically addressed in a microarray experiment aimed at monitoring the movement of mRNAs between RNPs and polysomes during meiotic and post-meiotic mouse testis development (Iguchi *et al.* 2006). The authors identified over 700 translationally regulated transcripts (e.g. exhibiting a redistribution of at least 20 per cent of mRNAs between the free RNPs and the polysomal fractions). These transcripts included the vast majority of mRNAs being translationally upregulated

during late spermiogenesis, a common regulation to compensate transcription silencing from mid-spermiogenesis onwards, as well as a small cluster of meiotic mRNAs that were translated only latter in post-meiotic cells. The fact that the genes identified in this study were translationally regulated does not necessarily mean that their RNA and protein expression patterns will differ significantly. Indeed several genes identified in the latter study as being translationally regulated were found to have similar mRNA and protein expression patterns when comparing our rat germline transcriptome (Chalmel *et al.* 2007b) and proteome (Rolland *et al.* 2007) data. To the contrary, some other genes which exhibit different RNA and protein patterns were not found to be translationally regulated (Iguchi *et al.* 2006). These apparent discrepancies could probably be the result of additional mechanisms influencing the transcript *versus* protein ratio. As a matter of fact, small RNAs and especially miRNAs have recently emerged as alternative ways of regulating gene expression in testicular cells at the post-transcriptional and translational levels (for review, see He *et al.* 2009). Indeed, the specific disruption of *Dicer1*, one of the enzymes responsible for miRNAs synthesis, either in germ cells (Hayashi *et al.* 2008; Maatouk *et al.* 2008) or Sertoli cells (Papaioannou *et al.* 2009) resulted in an impaired spermatogenesis and infertility phenotypes, confirming the importance of miRNA processing for the completion of mammalian spermatogenesis. This has prompted several laboratories to develop high-throughput methods dedicated to the identification and quantification of small RNAs in the testis (figure 3). Recently, cloning and sequencing methods combined to PCR-based approaches were used to detect and compare the expression of miRNAs in mouse immature and mature testis (Ro *et al.* 2007) or between mouse testis and ovary (Mishima *et al.* 2008). Although these studies showed a great efficiency in the listing and screening of novel miRNAs in testicular cells, PCR-based expression profiling appeared to be heavy and time-consuming. As an alternative, microarrays were chosen by several groups to study miRNA expression profiles. The first microarrays-based study to focus on testicular miRNAs was carried out by Yan *et al.* (2007) who simultaneously compared the expression of 892 miRNAs between immature (one-week-old) and mature (seven-week-old) mouse testes. Among the 19 miRNAs that were differentially expressed during testis post-natal development, 14 were overexpressed in immature testis. *In silico* sequence analysis identified multiple putative target mRNAs for these 14 miRNAs, such as *Rbpl1*, *Sox 5* or *Nr6a1*. Interestingly, these factors have been previously shown as being specifically expressed and biologically active in the meiotic and haploid germ cells, suggesting that their translational regulation during the spermatogenic process could be due to small RNAs silencing. An identical study driven by Yan and co-workers in rhesus monkey (*Macaca mulatta*) and human, led to consistent results, indicating that these regulatory processes could be conserved throughout mammalian spermatogenesis (Yan *et al.* 2009).

Paradoxically, the expanding list of testicular miRNA did not really result in a better understanding of their function, as little experimental information has been provided so far about their molecular targets. A recent study tried to tackle this issue and provided meaningful results on the molecular action of the miRNA Mir-709 in the regulation of DNA methylation and imprinting. Using miRNAs microarrays, the authors evaluated the effects of X-ray radiation exposure on the mouse testis microRNAome (Tamminga *et al.* 2008). Among the 20 miRNAs whose expression varied during irradiation, they found Mir-709 to be strongly overexpressed in the germ cells of X-ray-exposed testis. By carrying out *in silico* prediction, luciferase assay and DNA methylation measurement, Tamminga and coworkers demonstrated that Mir-709 specifically targets and degrades BORIS (Brother of the regulator of imprinted sites) mRNA, an important regulator of DNA methylation in the germ cells. This degradation resulted in a decreased expression of BORIS together with a global genome hypomethylation. This type of work well illustrates that large-scale analyses of biological processes such as spermatogenesis largely benefit from such targeted functional validation.

In addition to small RNA processing in germ cells, the class of non-coding transcripts also includes antisense RNAs and long non-coding RNAs. Considering that hybridization-based approaches are experimentally designed to investigate the coding part of the genome, very few have been done so far to get a more global insight on the expression of antisense and long non-coding RNAs. Interestingly, several indications now justify tackling some functional roles for these uncommon transcripts in the regulation of spermatogenesis (for review, see Lee *et al.* 2009). The development of whole genome screening techniques, such as tiling arrays (Mockler *et al.* 2005) or deep RNA sequencing, but also of more specific approaches such as ASSAGE (He *et al.* 2008), will make it possible to describe expression profiles at the entire genome scale. RNA sequencing, for example, is a very promising technique that is expected to revolutionize the way complex transcriptomes are studied (Wang *et al.* 2009). With the aim of overcoming technical limitations inherent to hybridization-based approaches (*a priori* approaches on sequences, high background, limited dynamic range of detection), this technique relies on the high-throughput sequencing of cDNA library fragments and subsequent alignment of the obtained sequences on a genome of reference. This straightforward approach produces both qualitative and quantitative results. Indeed, it offers precise sequence information such as SNP or exon connections while providing high accuracy expression quantification data that cover the entire ribonucleic material present in cells or tissues. This approach is still immature and will have to face important challenges, in methodology and bioinformatics, to generate reproducible and easily interpretable results. However, it should enable researchers to round off and refine the datasets obtained so far, but also to perform the complete profiling of male germ cells'

transcriptional activity by including the non-coding counterpart of the genome.

(g) *The spermatozoon in the post-genomic era: revealing the secrets of a silent cell*

Transcriptomics usually generates larger datasets than proteomics, but spermatozoa are a special case, for which proteome-based studies have probably generated more relevant data. Indeed, due to its terminally differentiated and autonomous state, the male gamete is barely transcriptionally inactive. Thus, the spermatozoon will not respond to environmental *stimuli* at the genome level but rather by the activation, modification or subcellular relocalization of its protein content in order to achieve the capacitation and fertilization processes. Nevertheless, several laboratories have investigated the human sperm transcriptome, through both microarray-based experiments (Ostermeier *et al.* 2002, 2005; Wang *et al.* 2004; Nguyen *et al.* 2008) and SAGE library construction (Zhao *et al.* 2006; figure 4). These studies highlighted an unexpected diversity of mRNA species, raising the question of their functional relevance. Whereas it has been demonstrated by a proteomic approach that protein synthesis is slightly active in the sperm mitochondria (Zhao *et al.* 2009), some of these mRNAs could also be delivered to the oocyte and play further roles in early embryogenesis (Dadoune 2009). From another aspect, evaluating the spermatozoa RNA content using DNA microarrays provided meaningful results to assess sperm quality and/or to elucidate some cases of male sterility (Platts *et al.* 2007; Garrido *et al.* 2009).

The proteomic profiling of spermatozoa has become a major issue in the study of human fertility and was facilitated by the possibility of recovering large amounts of pure spermatozoa from healthy or case donors. As a result, whereas transcriptomic and proteomic analyses of male germ cells preferentially relied on rodents or non-mammalian models, the deciphering of sperm constituents and biological parameters such as morphology, motility or fertilization ability was mainly performed on human samples. Among the different strategies employed to characterize the human sperm proteome, systematic mapping achieved by two-dimensional electrophoresis (2DE) and MALDI-TOF mass spectrometry led to the resolution of over 1000 protein-spots and the unambiguous identification of 98 proteins (Martinez-Heredia *et al.* 2006). A much deeper unravelling of the human sperm proteome following a differential extraction of proteins and a subsequent nano-LC-MS/MS analysis was also reported (Johnston *et al.* 2005). Thus, 1760 proteins (out of the 2300 the authors predict to be present in human sperm) were identified, leading to the largest catalogue of proteins potentially involved in or important for fertilization and to a myriad of potential contraceptive targets. A similar strategy was also recently employed for the analysis of Triton X-100 soluble and insoluble sperm proteins (Baker *et al.* 2007). This study listed 1053 proteins, among which 8 per cent were characterized for the first time in sperm cells. Major advances have also been made in the proteomic global analysis of

Figure 3. Deciphering spermatogenesis through expression analyses. Spermatogenesis is a complex, coordinated and continuous process by which spermatogonia give rise to mature spermatozoa. In rodents, the first wave of this process starts a few days after birth, when gonocytes (prespermatogonia) resume proliferation and become undifferentiated spermatogonia or spermatogonial stem cells. Once SSC are engaged in differentiation process (i.e. they become differentiated spermatogonia) they undergo six successive mitotic divisions, giving rise to meiotic spermatocytes, which go through two consecutive divisions, with a single round of DNA replication, to give rise to haploid spermatids. The final step in spermatogenesis consists of the transformation of spermatids into spermatozoa and is called spermiogenesis. References: 1, Divina *et al.* (2005); 2, Shima *et al.* (2004); 3, Almstrup *et al.* (2004); 4, Ellis *et al.* (2004); 5, Clemente *et al.* (2006); 6, Schultz *et al.* (2003); 7, Yao *et al.* (2004); 8, Fox *et al.* (2003); 9, Huang *et al.* (2005); 10, Zhu *et al.* (2006); 11, Cagney *et al.* (2005); 12, Ro *et al.* (2007); 13, Yan *et al.* (2009); 14, Yan *et al.* (2007); 15, Tamminga *et al.* (2008); 16, Govin *et al.* (2006); 17, Wu *et al.* (2004); 18, Namekawa *et al.* (2006); 19, Chalmel *et al.* (2007b); 20, Schlecht *et al.* (2004); 21, Rolland *et al.* (2007).

non-human spermatozoa, including those of the mouse (Baker *et al.* 2008b), rat (Baker *et al.* 2008a) and bull (Peddinti *et al.* 2008), and also of the fruitfly (Dorus *et al.* 2006). Interestingly, the latter study was appealing in that marked correlations were found between the *Drosophila* sperm proteome and accessory structures of the mouse sperm flagellum, suggesting

the existence of molecular pathways that are critical for sperm function and conserved from invertebrates to mammals.

Attempts to characterize the complete sperm proteome were successful in producing large amounts of data that contributed to get a better understanding of sperm function (figure 4). However, the evaluation

Figure 4. The main components and maturation of spermatozoa. Once released from the seminiferous epithelium, spermatozoa are transferred to the epididymis. During their transit in the epididymis, spermatozoa become motile. However, they are not yet competent for fertilization. Within the female reproductive tract, they finally acquire the ability to fertilize eggs, during a time-dependent process called capacitation, making it possible for them to undergo the acrosomal reaction and to bind to the oocyte membrane. References: 1, Ostermeier *et al.* (2002); 2, Ostermeier *et al.* (2005); 3, Wang *et al.* (2004); 4, Nguyen *et al.* (2008); 5, Zhao *et al.* (2006); 6, Dorus *et al.* (2006); 7, Baker *et al.* (2008b); 8, Cao *et al.* (2006); 9, Krisfalusi *et al.* (2006); 10, Stein *et al.* (2006); 11, Baker *et al.* (2008a); 12, Peddinti *et al.* (2008); 13, Martinez-Heredia *et al.* (2006); 14, Johnston *et al.* (2005); 15, Baker *et al.* (2007); 16, Kim *et al.* (2006); 17, Baker *et al.* (2005); 18, Sleight *et al.* (2005); 19, Nixon *et al.* (2009); 20, Lefievre *et al.* (2007); 21, Ficarro *et al.* (2003); 22, Lalancette *et al.* (2006); 23, Platt *et al.* (2009); 24, Zhao *et al.* (2007); 25, Martinez-Heredia *et al.* (2008); 26, de Mateo *et al.* (2007); 27, Bohring *et al.* (2001); 28, Bohring & Krause (2003); 29, Fijak *et al.* (2005); 30, Paradowska *et al.* (2006); 31, Domagala *et al.* (2007).

of protein's functional aspects also relies on the analysis of their subcellular localization as defined by Anderson & Anderson (1998), especially in the case of the heavily compartmented sperm cells. Thus, several studies took advantage of sample pre-fractionation to investigate the protein content of diverse spermatozoa subcellular compartments such as the flagellum (Cao *et al.* 2006) and/or fibrous sheath (Kim *et al.* 2006; Krisfalusi *et al.* 2006), the acrosomal content or the sperm head membrane (Stein *et al.* 2006). As an example, sucrose gradient fractionation of SDS-resistant tail structures and their analysis by 2DE combined with MALDI-TOF/TOF-MS/MS led to the identification of 60 flagellar proteins (Cao *et al.* 2006). One-dimensional electrophoresis (1DE) and tandem mass spectrometry (MS/MS) carried out on mouse (Krisfalusi *et al.* 2006) and human (Kim *et al.* 2006) sperm fibrous sheath also led to the identification of several metabolic enzymes such as GAPDHS, ALDOA, LDHC and PK, suggesting this structure to be an anchoring platform for glycolytic machinery providing an autonomous production of ATP, independently from mitochondrial oxidation. In order to focus on proteins more likely to be involved in fecundity, both the acrosomal content (soluble

proteins released after the acrosomal reaction) and the membrane constituents (surface-biotinylated proteins of intact sperm and proteins from acrosomal vesicles released after acrosomal reaction) were investigated in mouse by 1DE combined with LC-MS/MS identification (Stein *et al.* 2006). Several hundreds of proteins were identified, 114 of which were predicted to be transmembrane or signal peptide-containing proteins. One third of male mice with the corresponding gene deletions were found to be sterile or subfertile, confirming the pertinence of this strategy.

Proteomics has also been used to investigate specific modifications during sperm maturation and capacitation, as well as to identify proteins responsible for some of these modifications (figure 4). 2D-DIGE was thus applied to rat spermatozoa from the cauda and caput epididymis, to highlight changes in the protein profiles during epididymal sperm transit (Baker *et al.* 2005). Sixty protein spots exhibited significant changes and eight proteins were identified by MALDI-TOF MS, including one protein undergoing serine-phosphorylation as demonstrated by two-dimensional Western blotting. Non-capacitated and *in vitro* capacitated mouse spermatozoa were also compared to investigate membrane protein

redistribution during capacitation (Sleight *et al.* 2005; Nixon *et al.* 2009). In total, 25 proteins were shown to be dissociated from lipid rafts and therefore, potentially involved in the signalling pathways associated with the initiation of capacitation. Following similar objectives, Nixon *et al.* (2009) used a combination of 2DE-MS and reversed phase nano-LC/MS/MS to identify 100 proteins from the mouse sperm detergent resistant membranes, among which 21 factors (such as Izumo or Ace) had been previously described as important for sperm–egg interaction. This study also revealed the presence of numerous molecular chaperone proteins, suggesting that specific mechanisms operate during capacitation to reshape sperm surface and enhance ovocyte interaction.

Post-translational modifications of sperm surface proteins are major events of the capacitation process and have also been specifically addressed using proteomic studies (figure 4). For example, Lefievre and coworkers investigated S-nitrosylation of proteins in human sperm, an important modification for the enhancement of capacitation and egg-binding ability (Lefievre *et al.* 2007). By incubating human spermatozoa with nitric oxide and further analysing it with a 1DE/MS/MS approach, they identified 240 putative targets for S-nitrosylation, including proteins from the AKAP family or heat shock proteins. Several studies were also conducted to identify proteins that undergo phosphorylation during the capacitation process. Combining anti-phospho-tyrosine two-dimensional Western blots, IMAC, post-IMAC dephosphorylation and LC-MS/MS analysis, Ficarro *et al.* (2003) reported the identification of more than 60 tyrosine-phosphorylation sites in 15 proteins from human capacitated sperm and 16 additional proteins undergoing tyrosine-phosphorylation during capacitation. Proteins potentially responsible for these tyrosine-phosphorylation events were investigated in the capacitated bull sperm (Lalancette *et al.* 2006). A cytosolic fraction enriched in tyrosine kinase activity was generated by poly-Glu-Tyr affinity chromatography, with MALDI-TOF MS, QTOF MS and LC-MS/MS analyses subsequently used to identify 126 proteins. These proteins included four different tyrosine kinases, but also putative partners, activators, inhibitors or substrates to these enzymes. Recently, isotopic labelling and IMAC-based phosphopeptides enrichment from trypsin digested sperm samples allowed the identification of 55 phosphorylation sites on capacitated and non-capacitated mouse sperm proteins (Platt *et al.* 2009). Interestingly, the use of quantization methods based on the comparison of MS spectra signal intensities allowed sites presenting a capacitation-associated phosphorylation state to be highlighted. In the fascinating field of sperm proteome modifications, the sperm glycome is becoming a major issue for the understanding of male fertility. In addition to the suspected role of sugar moieties in the mediation of sperm–egg interactions (for review, see Schroter *et al.* 1999), a recent analysis by mass spectrometry of human sperm glycopeptides showed that sperm membrane carries several class of N-glycans. These glycans inhibit both the innate and adaptative immune response in the female genital

tract by preventing sperm cells from NK cell-mediated lysis (Pang *et al.* 2007). These results were recently strengthened by a wide-scale analysis of the human seminal plasma glycome that revealed the abundance of such immunomodulating glycans in the male genital tract (Pang *et al.* 2009).

Finally, proteomic strategies were also used to directly investigate human reproductive physiopathology, with the aim of providing new tools for the detection of sperm alterations at the protein level (figure 4). Several studies performed 2DE differential maps of samples from both normal and asthenozoospermic patients (Zhao *et al.* 2007; Martinez-Heredia *et al.* 2008) and identified many proteins whose decreased expression could be correlated with reduced sperm motility. Thus, de Mateo *et al.* (2007) identified 15 sperm proteins whose altered expression profiles were strongly correlated with either DNA damage or protamine content of the ejaculated spermatozoa, two parameters that are widely used to measure sperm cell integrity (Andrabi 2007). The latter study clearly confirmed the potentiality of a coordinated evaluation of several biological parameters in the clinical assessment of sperm quality. Several groups also used antisperm antibodies (ASA) from seminal plasma to identify proteins potentially involved in immunological diseases causing infertility (Bohring *et al.* 2001; Bohring & Krause 2003; Fijak *et al.* 2005; Paradowska *et al.* 2006). Recently, Domagala *et al.* (2007) performed two-dimensional Western blots on normal sperm proteins with sera or semen antibodies from infertile women and men. Mass spectrometry analysis of the immunodetected protein spots lead to the identification of 35 proteins, 10 of which being sperm-specific variants. Interestingly, several factors previously described as putative targets for male contraception were identified, confirming the relevance of these results.

Overall, these studies demonstrated the great power of proteomic-based approaches to address the dynamic of sperm maturation and capacitation within the male reproductive tract. However, it should be kept in mind that the considerable changes occurring in spermatozoa from spermiation to fertilization are almost the exclusive reflect of an intense environment-induced protein reshaping. Male gamete biology should thus be regarded as a multidimensional network in which protein modifications of various kinds are timely and spatially regulated. We can then assume that integrated and transverse analyses of the extended datasets already available on the transcriptome, proteome, glycome and secretome of the male reproductive peripheral organs (Turner *et al.* 2006; Dacheux *et al.* 2009) and of the female genital tract (Shaw *et al.* 2007) will be decisive to help decipher the inter-cellular molecular networks that drive sperm cells biology.

(h) The complexity of data analysis: towards holistic biology of spermatogenesis

Recent advances in genomics and the outcome of new technologies for large-scale gene expression analyses have greatly improved our knowledge of male

reproduction, making it possible to investigate the molecular mechanisms underlying this process at a genome wide level. The studies reviewed herein have led to the identification of factors likely to be of importance for particular steps in testicular development or required for male fertility. However, most of these studies have failed to exploit the large amounts of data generated by such high-throughput approaches to provide useful and meaningful results to enhance our overall understanding of the cellular events involved. Several strategies have been used, focusing on a restricted group of relevant genes or proteins likely to be selected for further investigations. Such strategies include tissue-profiling experiments or cross-species comparisons for the detection of testis-specific genes or evolutionarily conserved genes, the specific expression and conserved expression profiles of which may be correlated with essential functions (figure 5). These filtering processes have indeed demonstrated their efficiency for identifying the key factors from huge lists of anonymous candidates, but they cannot bridge the gap between the identification of thousands of coexpressed genes or proteins and an understanding of the connections between them, providing the ultimate explanation of the correct progression of a complete biological process.

The unravelling of the complete set of transcriptional regulation mechanisms, leading to the coordinated expression of a full set of functional products and the identification of the protein network interactions occurring during normal testis and germ cell development are crucial if one wants to understand pathological disorders of the human testis and their origins. Considerable effort has been made in the last few years to integrate data from large-scale experiments and to develop tools enabling the researchers not only to describe a group of genes or proteins with similar expression profiles, but also to propose and develop new hypotheses from their analyses. One common way of analysing such experiments is to use the gene description (annotation) of the gene ontology (GO) consortium for functional data mining (Ashburner *et al.* 2000). The GO consortium provides the scientific community with predictions concerning gene function, the process in which the gene product is involved and subcellular components with which it is associated, using a controlled and structured vocabulary (ontologies). This makes it possible to study GO terms enrichment in a set of genes/proteins and thus to demonstrate objectively that specific functions are significantly associated with a given process. Although this approach remains quite descriptive, it both facilitates the identification of unexpected important pathways and predicts functions for uncharacterized genes. It can significantly provide additional insight into observed transcriptional profiles and/or facilitate the identification of genes or proteins those belonging to the same complex from sets of coexpressed genes or proteins. Such a multifaceted analysis strategy was recently proposed for deeper analyses of data from a gene expression profiling study of the male germline (Chalmel *et al.* 2007a). These authors used *in silico* promoter analysis to evaluate the occurrence of

known transcription factors binding sites (TFBSs) within the regulatory sequences of genes co-expressed during mouse spermatogenesis. Combining TFBSs predictions, DNA conservation and high quality expression data, Chalmel and coworkers found that the cAMP response element of the spermiogenesis-related factor *Crem* was indeed more significantly enriched in the genomic regions of *loci* specifically expressed in spermatids than it would be expected at random (figure 5b). It is obvious that the systematic extension of such analyses should make it easier to identify and to discover relevant regulatory elements involved in the establishment of the germline expression programme. The authors also investigated the use of large protein network data now available for validation of the biological significance of clusters of co-expressed genes in terms of protein complexes. They focused on a small group of genes selected by filtering on the basis of their testis-specific and conserved expression in mammals and for which at least one interacting factor was known. This analysis yields a global amount of 87 interacting factors from the initial set of 15 genes, corresponding to genes expressed in the testis as well as other genes not detected on microarrays (figure 5c). The large number of interactions detected for genes shown to be important for male reproduction clearly demonstrates the relevance of this approach for the confirmation and extension of expression data. Finally transcript and protein expression profiles were compared during male germ cell development, leading to the detection of several genes displaying apparently delayed transcription and translation during spermiogenesis. An analysis similar to the one made for genomic sequences could be carried out to identify known or new motifs within untranslated regions in such transcripts, making it possible to predict the involvement of specific RNA-binding proteins in these translation regulation mechanisms during spermatogenesis. Such a project has recently been undertaken in our laboratory. In this context, a recent study by Liu *et al.* (2007) reporting the identification of a number of genes encoding mRNAs specifically subject to alternative 3'-processing during meiosis and postmeiotic development is of great interest.

The study by Chalmel and collaborators undoubtedly paves the way towards a systems-based analysis of male sexual reproduction by highlighting the possibility of bridging the gap between DNA sequence, transcriptional activity and translation regulation right up to the reconstitution of functional protein complexes, at each step of spermatogenesis. Modelling the entire process of spermatogenesis will, however, remain a huge challenge as it should include regulation by various hormones as well as continuous, complex communication between all cell types present in the testis. The magnitude of this challenge has recently increased with new data from the ENCODE project concerning exhaustive analyses of the transcription features of about 1 per cent of the human genome (Birney *et al.* 2007). This study has revealed the pattern of gene expression to be incredibly more complex than expected initially, with dispersed

Decoding how genes and proteins interact and participate in signal networks remains one of the most fundamental questions for biologists. Moreover, as

Figure 5. Data mining: going beyond expression profiles. Figure adapted from Stein *et al.* (2006). (a) Filtering strategy. Cross-species comparison and tissue profiling identified 80 testis-specific orthologues with conserved expression patterns in mouse, rat and human spermatogenesis. Genes differentially expressed during spermatogenesis were identified for each species. Differentially expressed genes corresponding to orthologues represented on microarrays for all three species were identified, using the HomoloGene database. Finally, expression data for 17 mouse somatic tissues downloaded from the GEO web server were used to select testis-specific genes. SE, SPG, SC, ST, TU and TT correspond to Sertoli cell, spermatogonia, pachytene spermatocyte, early spermatid, seminiferous tubule and total testis samples, respectively. SO, MI, MEI and PM correspond to the somatic, mitotic, meiotic and postmeiotic expression clusters, respectively. (b) Promoter analysis. Automated multi-step promoter analysis showed mouse testis-specific postmeiotic gene promoter regions to be specifically enriched in the CRE motif. We show here the results for *Capza3*, a gene also identified as present within the interaction node of conserved and testis-specific genes. (c) Interaction network analysis. Data from IntAct, MINT and BioGRID were used to monitor interactions between conserved and testis-specific gene products. Some proteins important for male reproduction were found within these networks and were found to associate with a large number of interacting factors. As not all the interacting factors were detected on microarray analyses, interaction network analysis might help to extend expression data.

we have pointed out before, mRNA and protein expression in spermatogenesis seems especially regulated and complex. Measuring simple levels of expression in a testicular context, therefore, can be misleading without deeper knowledge of isoform regulation and the differential fates of products. It is now possible, using novel all-exon and tiling arrays, to measure independently all known exons in the complete genomes of rodents and *Homo sapiens*. These experiments should make it possible to identify specific isoforms and novel non-coding RNAs not detected by previous approaches (Cheng *et al.* 2005). Armed with this information, the targets of spermatogenesis-relevant epigenetic regulation should gradually become known to us. No doubt that striking biological discoveries will stem from the new breed of comparative omics studies.

While most omics approaches have reached some sort of a maturity, proteomics remains somewhat the growing child. Partly this has been due to a greater variety of technologies and strategies being explored simultaneously. Secondly, the problem of resolving the wide dynamic range in protein expression remains since there is no technology available to amplify low-copy number proteins in a biological sample. Nevertheless, instrument sensitivity is rapidly improving and, although still incomplete, extensive proteome coverage can now be achieved in large proteome experiments (de Godoy *et al.* 2006, 2008). Likewise, absolute quantification of proteins in complex mixtures is becoming accessible thanks to innovative strategies such as the protein standard absolute quantification (PSAQ) technique (Brun *et al.* 2007). Tools for charting the plethora of post-translational modifications are also expanding rapidly such as for phosphorylation (Hilger *et al.* 2009) although this remains largely a specialist field at present. Despite recent advances in our understanding of protein carbohydrate modifications, the throughput of protein glycosylation analyses remains too low for exhaustive glycoproteomic analysis. One of the most astonishing breakthroughs in proteomics has been achieved very recently with the SILAC-mouse approach, a versatile tool by which organ and cell proteomes of different mouse strains, including knock-out mouse models, can be quantitatively compared under complex *in vivo* conditions (Kruger *et al.* 2008). Burgeoning advances in mass spectrometry imaging will inevitably attract biologists and clinicians as the ready pickings from this technology become more widely known. However, newcomers to proteomics should proceed with caution as this field remains highly technical. Fortunately, expert platforms are available and can collaborate with providers of high-quality biological material, enabling specialists to concentrate on the biological questions they wish to address.

The large series of testis-relevant 'omics' datasets that have been collected may seem a largely unmapped landscape for most researchers in the field of male reproduction. Certainly it is common for data producers to concentrate on a particular biological aspect or subset of elements in their post-production analyses before moving on. One hopes that other explorers might take different paths. Certainly, when more

minds can be encouraged to look into this world of data, new insights into the molecular events which underlie spermatogenesis and, by extension, human reproductive disorders, will follow. What tools and resources can be provided to encourage this? A rational compilation of large- and small-scale omics data into a reproduction-oriented repository system, such as the GermOnline database (Primig *et al.* 2003) would be a useful bridge for many. The GermOnline knowledgebase (<http://www.germonline.org>) assembles studies relevant to the cell cycle, gametogenesis and fertility. It incorporates a cross-species systems browser to provide DNA sequence annotation, evolutionary relationships, gene expression and functional annotation. The database, built around a sophisticated genome browser (Ensembl), covers eight model organisms and *H. sapiens*, for which complete genome annotation data are available. Current efforts by Primig, Pineau and coworkers to integrate proteome datasets into GermOnline will further enhance this tool as an environment for decision support and hypothesis generation. In the same context, now seems also an appropriate time to construct and coordinate, with the Human Proteome Organization [www.hupo.org] a 'testis/epididymis proteome initiative'. Such an initiative would, ideally, facilitate mining and comparison between omics datasets at various levels for researchers in the field of male reproduction.

Taking a hypothesis to the bioinformatic playing field usually entails off-web analysis. For this, many software environments currently exist, each differently enabled to perform specific steps of data management and analysis. As an example, the AMEN software package developed by Chalmel & Primig (2008) offers intuitive access to key components of the powerful R scripting language. This stand-alone, open source, unified suite of tools enables biological and medical researchers with basic bioinformatics training to manage and explore genome annotation, chromosomal mapping, protein-protein interaction, expression profiling and proteomics data. The current version provides modules for (i) uploading and pre-processing data from microarray expression profiling experiments, (ii) detecting groups of significantly co-expressed genes, and (iii) searching for enrichment of functional annotations within those groups. Moreover, the software user interface is designed to simultaneously visualize several types of data such as protein-protein interaction networks in conjunction with expression profiles and cellular co-localization patterns. Development is currently underway to integrate next-generation datasets such as the testis microRNAome and metabolome.

We thank James Moore, Frederic Chalmel and Aurelie Lardenois for stimulating discussions and corrections.

REFERENCES

- Aebersold, R. & Mann, M. 2003 Mass spectrometry-based proteomics. *Nature* **422**, 198–207. (doi:10.1038/nature01511)
- Almstrup, K., Nielsen, J. E., Hansen, M. A., Tanaka, M., Skakkebaek, N. E. & Leffers, H. 2004 Analysis of cell-type-specific gene expression during mouse

- spermatogenesis. *Biol. Reprod.* **70**, 1751–1761. (doi:10.1095/biolreprod.103.026575)
- Anderson, N. L. & Anderson, N. G. 1998 Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853–1861. (doi:10.1002/elps.1150191103)
- Andrabi, S. M. 2007 Mammalian sperm chromatin structure and assessment of DNA fragmentation. *J. Assist. Reprod. Genet.* **24**, 561–569. (doi:10.1007/s10815-007-9177-y)
- Ashburner, M. *et al.* 2000 Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29.
- Baker, M. A., Witherdin, R., Hetherington, L., Cunningham-Smith, K. & Aitken, R. J. 2005 Identification of post-translational modifications that occur during sperm maturation using difference in two-dimensional gel electrophoresis. *Proteomics* **5**, 1003–1012. (doi:10.1002/pmic.200401100)
- Baker, M. A., Reeves, G., Hetherington, L., Müller, J., Baur, I. & Aitken, R. J. 2007 Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. *Proteomics Clin. Appl.* **1**, 524–532. (doi:10.1002/prca.200601013)
- Baker, M. A., Hetherington, L., Reeves, G., Muller, J. & Aitken, R. J. 2008a The rat sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. *Proteomics* **8**, 2312–2321. (doi:10.1002/pmic.200700876)
- Baker, M. A., Hetherington, L., Reeves, G. M. & Aitken, R. J. 2008b The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. *Proteomics* **8**, 1720–1730. (doi:10.1002/pmic.200701020)
- Beverdam, A. & Koopman, P. 2006 Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum. Mol. Genet.* **15**, 417–431. (doi:10.1093/hmg/ddi463)
- Birney, E. *et al.* 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816. (doi:10.1038/nature05874)
- Bohring, C. & Krause, W. 2003 Characterization of spermatozoa surface antigens by antisperm antibodies and its influence on acrosomal exocytosis. *Am. J. Reprod. Immunol.* **50**, 411–419. (doi:10.1034/j.1600-0897.2003.00103.x)
- Bohring, C., Krause, E., Habermann, B. & Krause, W. 2001 Isolation and identification of sperm membrane antigens recognized by antisperm antibodies, and their possible role in immunological infertility disease. *Mol. Hum. Reprod.* **7**, 113–118. (doi:10.1093/molehr/7.2.113)
- Bouma, G. J., Affourtit, J. P., Bult, C. J. & Eicher, E. M. 2007 Transcriptional profile of mouse pre-granulosa and Sertoli cells isolated from early-differentiated fetal gonads. *Gene Expr. Patterns* **7**, 113–123. (doi:10.1016/j.modgep.2006.05.011)
- Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F. & Garin, J. 2007 Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell Proteomics* **6**, 2139–2149.
- Cagney, G., Park, S., Chung, C., Tong, B., O'dushlaine, C., Shields, D. C. & Emili, A. 2005 Human tissue profiling with multidimensional protein identification technology. *J. Proteome Res.* **4**, 1757–1767. (doi:10.1021/pr0500354)
- Calvel, P., Kervarec, C., Lavigne, R., Vallet-Erdtmann, V., Guerrois, M., Rolland, A., Chalmel, F., Jegou, B. & Pineau, C. 2009 Clph, a novel casein kinase 2-phosphorylated disordered protein, is specifically associated with post-meiotic germ cells in rat spermatogenesis. *J. Proteome Res.* **9**, 9.
- Cao, W., Gerton, G. L. & Moss, S. B. 2006 Proteomic profiling of accessory structures from the mouse sperm flagellum. *Mol. Cell Proteomics* **5**, 801–810. (doi:10.1074/mcp.M500322-MCP200)
- Chalmel, F. & Primig, M. 2008 The annotation, mapping, expression and network (AMEN) suite of tools for molecular systems biology. *BMC Bioinform.* **9**, 86. (doi:10.1186/1471-2105-9-86)
- Chalmel, F., Lardenois, A. & Primig, M. 2007a Toward understanding the core meiotic transcriptome in mammals and its implications for somatic cancer. *Ann. N.Y. Acad. Sci.* **1120**, 1–15. (doi:10.1196/annals.1411.010)
- Chalmel, F. *et al.* 2007b The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. USA* **104**, 8346–8351. (doi:10.1073/pnas.0701883104)
- Charbonnier, S., Gallego, O. & Gavin, A. C. 2008 The social network of a cell: recent advances in interactome mapping. *Biotechnol. Annu. Rev.* **14**, 1–28. (doi:10.1016/S1387-2656(08)00001-X)
- Chen, G. *et al.* 2002 Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell Proteomics* **1**, 304–313.
- Chen, C. *et al.* 2005 ERM is required for transcriptional control of the spermatogonial stem cell niche. *Nature* **436**, 1030–1034. (doi:10.1038/nature03894)
- Cheng, J. *et al.* 2005 Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154. (doi:10.1126/science.1108625)
- Chu, D. S., Liu, H., Nix, P., Wu, T. F., Ralston, E. J., Yates, J. R. & Meyer, B. J. 2006 Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. *Nature* **443**, 101–105. (doi:10.1038/nature05050)
- Clemente, E. J., Furlong, R. A., Loveland, K. L. & Affara, N. A. 2006 Gene expression study in the juvenile mouse testis: identification of stage-specific molecular pathways during spermatogenesis. *Mamm. Genome* **17**, 956–975. (doi:10.1007/s00335-006-0029-3)
- Com, E., Evrard, B., Roepstorff, P., Aubry, F. & Pineau, C. 2003 New insights into the rat spermatogonial proteome: identification of 156 additional proteins. *Mol. Cell Proteomics* **2**, 248–261. (doi:10.1074/mcp.M300010-MCP200)
- Com, E., Rolland, A. D., Guerrois, M., Aubry, F., Jegou, B., Vallet-Erdtmann, V. & Pineau, C. 2006 Identification, molecular cloning, and cellular distribution of the rat homolog of minichromosome maintenance protein 7 (MCM7) in the rat testis. *Mol. Reprod. Dev.* **73**, 866–877. (doi:10.1002/mrd.20453)
- Conrads, K. A. *et al.* 2005 A combined proteome and microarray investigation of inorganic phosphate-induced pre-osteoblast cells. *Mol. Cell Proteomics* **4**, 1284–1296.
- Costoya, J. A., Hobbs, R. M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K. E., Wolgemuth, D. J. & Pandolfi, P. P. 2004 Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat. Genet.* **36**, 653–659. (doi:10.1038/ng1367)
- Dacheux, J. L., Belleanne, C., Jones, R., Labas, V., Belghazi, M., Guyonnet, B., Druart, X., Gatti, J. L. & Dacheux, F. 2009 Mammalian epididymal proteome. *Mol. Cell Endocrinol.* **306**, 45–50. (doi:10.1016/j.mce.2009.03.007)
- Dadoune, J. P. 2009 Spermatozoal RNAs: what about their functions? *Microsc. Res. Tech.* **12**, 12.
- de Godoy, L. M., Olsen, J. V., de Souza, G. A., Li, G., Mortensen, P. & Mann, M. 2006 Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**, R50. (doi:10.1186/gb-2006-7-6-r50)

- de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C. & Mann, M. 2008 Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254. (doi:10.1038/nature07341)
- de Mateo, S., Martinez-Heredia, J., Estanyol, J. M., Dominguez-Fandos, D., Vidal-Taboada, J. M., Ballesca, J. L. & Oliva, R. 2007 Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. *Proteomics* **7**, 4264–4277. (doi:10.1002/pmic.200700521)
- de Rooij, D. G. 2001 Proliferation and differentiation of spermatogonial stem cells. *Reproduction* **121**, 347–354. (doi:10.1530/rep.0.1210347)
- de Rooij, D. G. & de Boer, P. 2003 Specific arrests of spermatogenesis in genetically modified and mutant mice. *Cytogenet. Genome Res.* **103**, 267–276.
- Divina, P., Vlcek, C., Strnad, P., Paces, V. & Forejt, J. 2005 Global transcriptome analysis of the C57BL/6J mouse testis by SAGE: evidence for nonrandom gene order. *BMC Genomics* **6**, 29. (doi:10.1186/1471-2164-6-29)
- Domagala, A., Pulido, S., Kurpisz, M. & Herr, J. C. 2007 Application of proteomic methods for identification of sperm immunogenic antigens. *Mol. Hum. Reprod.* **13**, 437–444. (doi:10.1093/molehr/gam026)
- Dorus, S., Busby, S. A., Gerike, U., Shabanowitz, J., Hunt, D. F. & Karr, T. L. 2006 Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat. Genet.* **38**, 1440–1445. (doi:10.1038/ng1915)
- Ellis, P. J. et al. 2004 Modulation of the mouse testis transcriptome during postnatal development and in selected models of male infertility. *Mol. Hum. Reprod.* **10**, 271–281. (doi:10.1093/molehr/gah043)
- Essader, A. S., Cargile, B. J., Bundy, J. L. & Stephenson Jr, J. L. 2005 A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **5**, 24–34. (doi:10.1002/pmic.200400888)
- Ficarro, S. et al. 2003 Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J. Biol. Chem.* **278**, 11579–11589. (doi:10.1074/jbc.M20325200)
- Fijak, M., Iosub, R., Schneider, E., Linder, M., Respondek, K., Klug, J. & Meinhardt, A. 2005 Identification of immunodominant autoantigens in rat autoimmune orchitis. *J. Pathol.* **207**, 127–138. (doi:10.1002/path.1828)
- Fox, M. S., Ares, V. X., Turek, P. J., Haqq, C. & Reijo Pera, R. A. 2003 Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol. Reprod. Dev.* **66**, 403–421. (doi:10.1002/mrd.10364)
- Garrido, N., Martínez Conejero, J. A., Jauregui, J., Horcajadas, J. A., Simón, C., Remohí, J. & Meseguer, M. 2009 Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. *Fertil. Steril.* **91**, 1307–1310. (doi:10.1016/j.fertnstert.2008.01.078)
- Ge, R. S., Dong, Q., Sottas, C. M., Chen, H., Zirkin, B. R. & Hardy, M. P. 2005 Gene expression in rat Leydig cells during development from the progenitor to adult stage: a cluster analysis. *Biol. Reprod.* **72**, 1405–1415. (doi:10.1095/biolreprod.104.037499)
- Govin, J., Caron, C., Escoffier, E., Ferro, M., Kuhn, L., Rousseaux, S., Eddy, E. M., Garin, J. & Khochbin, S. 2006 Post-meiotic shifts in HSPA2/HSP70.2 chaperone activity during mouse spermatogenesis. *J. Biol. Chem.* **281**, 37 888–37 892. (doi:10.1074/jbc.M608147200)
- Guillaume, E., Dupaix, A., Moertz, E., Courtens, J., Jégou, B. & Pineau, C. 2000 Proteome analysis of spermatogonia: identification of a first set of 53 proteins. *Proteome*, <http://link.springerlink.com/content/56v8hr6fjehtq31q?pf=f89a98711db2463a92dad16921c38d088pi=0>.
- Guillaume, E., Evrard, B., Com, E., Moertz, E., Jégou, B. & Pineau, C. 2001a Proteome analysis of rat spermatogonia: reinvestigation of stathmin spatio-temporal expression within the testis. *Mol. Reprod. Dev.* **60**, 439–445. (doi:10.1002/mrd.1108)
- Guillaume, E., Pineau, C., Evrard, B., Dupaix, A., Moertz, E., Sanchez, J. C., Hochstrasser, D. F. & Jégou, B. 2001b Cellular distribution of translationally controlled tumor protein in rat and human testes. *Proteomics* **1**, 880–889. (doi:10.1002/1615-9861(200107)1:7<880::AID-PROT880>3.0.CO;2-2)
- Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. 1999 Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720–1730.
- Hamra, F. K., Schultz, N., Chapman, K. M., Grellhesl, D. M., Cronkhite, J. T., Hammer, R. E. & Garbers, D. L. 2004 Defining the spermatogonial stem cell. *Dev. Biol.* **269**, 393–410.
- Han, B. K., Kim, J. N., Shin, J. H., Kim, J. K., Jo, D. H., Kim, H. & Han, J. Y. 2005 Proteome analysis of chicken embryonic gonads: identification of major proteins from cultured gonadal primordial germ cells. *Mol. Reprod. Dev.* **72**, 521–529. (doi:10.1002/mrd.20374)
- Hayashi, K. et al. 2008 MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS ONE* **3**, e1738. (doi:10.1371/journal.pone.0001738)
- He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. 2008 The antisense transcriptomes of human cells. *Science* **322**, 1855–1857. (doi:10.1126/science.1163853)
- He, Z., Kokkinaki, M., Pant, D., Gallicano, I. & Dym, M. 2009 Small RNA molecules in the regulation of spermatogenesis. *Reproduction* **24**, 24.
- Hilger, M., Bonaldi, T., Gnad, F. & Mann, M. 2009 Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Mol. Cell Proteomics* **8**, 1908–1920.
- Hofmann, M. C., Braydich-Stolle, L. & Dym, M. 2005 Isolation of male germ-line stem cells; influence of GDNF. *Dev. Biol.* **279**, 114–124. (doi:10.1016/j.ydbio.2004.12.006)
- Huang, S. Y., Lin, J. H., Chen, Y. H., Chuang, C. K., Lin, E. C., Huang, M. C., Sunny Sun, H. F. & Lee, W. C. 2005 A reference map and identification of porcine testis proteins using 2-DE and MS. *Proteomics* **5**, 4205–4212. (doi:10.1002/pmic.200401284)
- Hubbard, T. J. et al. 2009 ENSEMBL 2009. *Nucleic Acids Res.* **37**, D690–D697. (doi:10.1093/nar/gkn828)
- Humphery-Smith, I. 2004 A human proteome project with a beginning and an end. *Proteomics* **4**, 2519–2521. (doi:10.1002/pmic.200400866)
- Iguchi, N., Tobias, J. W. & Hecht, N. B. 2006 Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc. Natl Acad. Sci. USA* **103**, 7712–7717. (doi:10.1073/pnas.0510999103)
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. & Sakaki, Y. 2001 A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574. (doi:10.1073/pnas.061034498)
- Johnson, J. M. et al. 2003 Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144. (doi:10.1126/science.1090100)

- Johnston, D. S., Wooters, J., Kopf, G. S., Qiu, Y. & Roberts, K. P. 2005 Analysis of the human sperm proteome. *Ann. N. Y. Acad. Sci.* **1061**, 190–202. (doi:10.1196/annals.1336.021)
- Kettman, J. R., Coleclough, C., Frey, J. R. & Lefkovits, I. 2002 Clonal proteomics: one gene—family of proteins. *Proteomics* **2**, 624–631. (doi:10.1002/1615-9861(200206)2:6<624::AID-PROT624>3.0.CO;2-I)
- Kim, Y. H., Haidl, G., Schaefer, M., Egner, U. & Herr, J. C. 2006 Compartmentalization of a unique ADP/ATP carrier protein SFEC (Sperm Flagellar Energy Carrier, AAC4) with glycolytic enzymes in the fibrous sheath of the human sperm flagellar principal piece. *Dev. Biol.* **10**, 10.
- Kokkinaki, M., Lee, T. L., He, Z., Jiang, J., Golestaneh, N., Hofmann, M. C., Chan, W. Y. & Dym, M. 2009 The molecular signature of spermatogonial stem/progenitor cells in the 6-day-old mouse testis. *Biol. Reprod.* **80**, 707–717. (doi:10.1095/biolreprod.108.073809)
- Krisfalusi, M., Miki, K., Magyar, P. L. & O'Brien, D. A. 2006 Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa. *Biol. Reprod.* **75**, 270–278. (doi:10.1095/biolreprod.105.049684)
- Kruger, M. *et al.* 2008 SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353–364. (doi:10.1016/j.cell.2008.05.033)
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K. & Saitou, M. 2008a Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617–1635. (doi:10.1101/gad.1649908)
- Kurimoto, K., Yamaji, M., Seki, Y. & Saitou, M. 2008b Specification of the germ cell lineage in mice: a process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. *Cell Cycle* **7**, 3514–3518.
- Lalancette, C., Faure, R. L. & Leclerc, P. 2006 Identification of the proteins present in the bull sperm cytosolic fraction enriched in tyrosine kinase activity: a proteomic approach. *Proteomics* **6**, 4523–4540. (doi:10.1002/pmic.200500578)
- Lander, E. S. *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)
- Lee, T. L., Pang, A. L., Rennert, O. M. & Chan, W. Y. 2009 Genomic landscape of developing male germ cells. *Birth Defects Res. C Embryo Today* **87**, 43–63. (doi:10.1002/bdrc.20147)
- Lefevre, C. & Mann, J. R. 2008 RNA expression microarray analysis in mouse prospermatogonia: identification of candidate epigenetic modifiers. *Dev. Dyn.* **237**, 1082–1089. (doi:10.1002/dvdy.21482)
- Lefevre, L., Chen, Y., Conner, S. J., Scott, J. L., Publicover, S. J., Ford, W. C. & Barratt, C. L. 2007 Human spermatozoa contain multiple targets for protein S-nitrosylation: an alternative mechanism of the modulation of sperm function by nitric oxide? *Proteomics* **7**, 3066–3084. (doi:10.1002/pmic.200700254)
- Liu, D., Brockman, J. M., Dass, B., Hutchins, L. N., Singh, P., Mccarrey, J. R., Macdonald, C. C. & Graber, J. H. 2007 Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. *Nucleic Acids Res.* **35**, 234–246. (doi:10.1093/nar/gkl919)
- Maatouk, D. M., Loveland, K. L., Mcmanus, M. T., Moore, K. & Harfe, B. D. 2008 Dicer1 is required for differentiation of the mouse male germline. *Biol. Reprod.* **79**, 696–703. (doi:10.1095/biolreprod.108.067827)
- Mann, M. & Jensen, O. N. 2003 Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261. (doi:10.1038/nbt0303-255)
- Martinez-Heredia, J., Estanyol, J. M., Ballesca, J. L. & Oliva, R. 2006 Proteomic identification of human sperm proteins. *Proteomics* **6**, 4356–4369. (doi:10.1002/pmic.200600094)
- Martinez-Heredia, J., de Mateo, S., Vidal-Taboada, J. M., Ballesca, J. L. & Oliva, R. 2008 Identification of proteomic differences in asthenozoospermic sperm samples. *Hum. Reprod.* **23**, 783–791. (doi:10.1093/humrep/den024)
- Matzuk, M. M. & Lamb, D. J. 2002 Genetic dissection of mammalian fertility pathways. *Nat. Cell Biol.* **4**, s41–s49.
- Mise, N. *et al.* 2008 Differences and similarities in the developmental status of embryo-derived stem cells and primordial germ cells revealed by global expression profiling. *Genes Cells* **13**, 863–877. (doi:10.1111/j.1365-2443.2008.01211.x)
- Mishima, T. *et al.* 2008 MicroRNA (miRNA) cloning analysis reveals sex differences in miRNA expression profiles between adult mouse testis and ovary. *Reproduction* **136**, 811–822. (doi:10.1530/REP-08-0349)
- Mockler, T. C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S. E. & Ecker, J. R. 2005 Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15. (doi:10.1016/j.ygeno.2004.10.005)
- Modrek, B., Resch, A., Grasso, C. & Lee, C. 2001 Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* **29**, 2850–2859. (doi:10.1093/nar/29.13.2850)
- Mueller, M., Martens, L. & Apweiler, R. 2007 Annotating the human proteome: beyond establishing a parts list. *Biochim. Biophys. Acta* **1774**, 175–191.
- Namekawa, S. H., Park, P. J., Zhang, L. F., Shima, J. E., Mccarrey, J. R., Griswold, M. D. & Lee, J. T. 2006 Post-meiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667. (doi:10.1016/j.cub.2006.01.066)
- Nef, S. *et al.* 2005 Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev. Biol.* **287**, 361–377. (doi:10.1016/j.ydbio.2005.09.008)
- Nguyen, M. T., Delaney, D. P. & Kolon, T. F. 2008 Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. *Fertil. Steril.* **5**, 5.
- Nixon, B., Bielaniowicz, A., McLaughlin, E. A., Tanphaichitr, N., Ensslin, M. A. & Aitken, R. J. 2009 Composition and significance of detergent resistant membranes in mouse spermatozoa. *J. Cell Physiol.* **218**, 122–134. (doi:10.1002/jcp.21575)
- O'Shaughnessy, P. J., Fleming, L., Baker, P. J., Jackson, G. & Johnston, H. 2003 Identification of developmentally regulated genes in the somatic cells of the mouse testis using serial analysis of gene expression. *Biol. Reprod.* **69**, 797–808. (doi:10.1095/biolreprod.103.016899)
- Oatley, J. M. & Brinster, R. L. 2008 Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.* **24**, 263–286. (doi:10.1146/annurev.cellbio.24.110707.175355)
- Oatley, J. M., Avarbock, M. R., Telaranta, A. I., Fearon, D. T. & Brinster, R. L. 2006 Identifying genes important for spermatogonial stem cell self-renewal and survival. *Proc. Natl Acad. Sci. USA* **103**, 9524–9529. (doi:10.1073/pnas.0603332103)
- Ostermeier, G. C., Dix, D. J., Miller, D., Khatri, P. & Krawetz, S. A. 2002 Spermatozoal RNA profiles of normal fertile men. *Lancet* **360**, 772–777. (doi:10.1016/S0140-6736(02)09899-9)
- Ostermeier, G. C., Goodrich, R. J., Moldenhauer, J. S., Diamond, M. P. & Krawetz, S. A. 2005 A suite of novel human spermatozoal RNAs. *J. Androl.* **26**, 70–74.
- Pang, P. C., Tissot, B., Drobnis, E. Z., Sutovsky, P., Morris, H. R., Clark, G. F. & Dell, A. 2007 Expression of

- bisecting type and Lewisx/Lewis^y terminated N-glycans on human sperm. *J. Biol. Chem.* **282**, 36 593–36 602. (doi:10.1074/jbc.M705134200)
- Pang, P. C., Tissot, B., Drobnis, E. Z., Morris, H. R., Dell, A. & Clark, G. F. 2009 Analysis of the human seminal plasma glycome reveals the presence of immunomodulatory carbohydrate functional groups. *J. Proteome Res.* **2**, 2.
- Papaioannou, M. D. et al. 2009 Sertoli cell Dicer is essential for spermatogenesis in mice. *Dev. Biol.* **326**, 250–259. (doi:10.1016/j.ydbio.2008.11.011)
- Paradowska, A., Bohring, C., Krause, E. & Krause, W. 2006 Identification of evolutionary conserved mouse sperm surface antigens by human antisperm antibodies (ASA) from infertile patients. *Am. J. Reprod. Immunol.* **55**, 321–330. (doi:10.1111/j.1600-0897.2006.00367.x)
- Peddinti, D., Nanduri, B., Kaya, A., Feugang, J. M., Burgess, S. C. & Memili, E. 2008 Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Syst. Biol.* **2**, 19. (doi:10.1186/1752-0509-2-19)
- Platt, M. D., Salicioni, A. M., Hunt, D. F. & Visconti, P. E. 2009 Use of differential isotopic labeling and mass spectrometry to analyze capacitation-associated changes in the phosphorylation status of mouse sperm proteins. *J. Proteome Res.* **2**, 2.
- Platts, A. E. et al. 2007 Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum. Mol. Genet.* **16**, 763–773. (doi:10.1093/hmg/ddm012)
- Primig, M. et al. 2003 GermOnline, a new cross-species community annotation database on germ-line development and gametogenesis. *Nat. Genet.* **35**, 291–292. (doi:10.1038/ng1203-291)
- Reinke, V. et al. 2000 A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616. (doi:10.1016/S1097-2765(00)00059-9)
- Ro, S., Park, C., Sanders, K. M., Mccarrey, J. R. & Yan, W. 2007 Cloning and expression profiling of testis-expressed microRNAs. *Dev. Biol.* **311**, 592–602. (doi:10.1016/j.ydbio.2007.09.009)
- Rolland, A. D., Evrard, B., Guitton, N., Lavigne, R., Calvel, P., Couvet, M., Jegou, B. & Pineau, C. 2007 Two-dimensional fluorescence difference gel electrophoresis analysis of spermatogenesis in the rat. *J. Proteome Res.* **6**, 683–697. (doi:10.1021/pr060436z)
- Rossi, P., Lolicato, F., Grimaldi, P., Dolci, S., Di Sauro, A., Filippini, D. & Geremia, R. 2008 Transcriptome analysis of differentiating spermatogonia stimulated with kit ligand. *Gene Expr. Patterns* **8**, 58–70. (doi:10.1016/j.modgep.2007.10.007)
- Schlecht, U., Demougin, P., Koch, R., Hermida, L., Wiederkehr, C., Descombes, P., Pineau, C., Jegou, B. & Primig, M. 2004 Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. *Mol. Biol. Cell* **15**, 1031–1043. (doi:10.1091/mbc.E03-10-0762)
- Schmidt, J. A., Avarbock, M. R., Tobias, J. W. & Brinster, R. L. 2009 Identification of glial cell line-derived neurotrophic factor-regulated genes important for spermatogonial stem cell self-renewal in the rat. *Biol. Reprod.* **81**, 56–66. (doi:10.1095/biolreprod.108.075358)
- Schroter, S., Osterhoff, C., Mcardle, W. & Ivell, R. 1999 The glycocalyx of the sperm surface. *Hum. Reprod. Update* **5**, 302–313. (doi:10.1093/humupd/5.4.302)
- Schultz, N., Hamra, F. K. & Garbers, D. L. 2003 A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc. Natl Acad. Sci. USA* **100**, 12 201–12 206. (doi:10.1073/pnas.1635054100)
- Shaw, J. L., Smith, C. R. & Diamandis, E. P. 2007 Proteomic analysis of human cervico-vaginal fluid. *J. Proteome Res.* **6**, 2859–2865. (doi:10.1021/pr0701658)
- Shima, J. E., Mclean, D. J., Mccarrey, J. R. & Griswold, M. D. 2004 The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol. Reprod.* **71**, 319–330. (doi:10.1095/biolreprod.103.026880)
- Sleight, S. B., Miranda, P. V., Plaskett, N. W., Maier, B., Lysiak, J., Scrable, H., Herr, J. C. & Visconti, P. E. 2005 Isolation and proteomic analysis of mouse sperm detergent-resistant membrane fractions: evidence for dissociation of lipid rafts during capacitation. *Biol. Reprod.* **73**, 721–729. (doi:10.1095/biolreprod.105.041533)
- Small, C. L., Shima, J. E., Uzumcu, M., Skinner, M. K. & Griswold, M. D. 2005 Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biol. Reprod.* **72**, 492–501. (doi:10.1095/biolreprod.104.033696)
- Stein, K. K., Go, J. C., Lane, W. S., Primakoff, P. & Myles, D. G. 2006 Proteomic analysis of sperm regions that mediate sperm–egg interactions. *Proteomics* **6**, 3533–3543. (doi:10.1002/pmic.200500845)
- Takemori, N. & Yamamoto, M. T. 2009 Proteome mapping of the *Drosophila melanogaster* male reproductive system. *Proteomics* **2**, 2.
- Tamminga, J., Kathiria, P., Koturbash, I. & Kovalchuk, O. 2008 DNA damage-induced upregulation of miR-709 in the germline downregulates BORIS to counteract aberrant DNA hypomethylation. *Cell Cycle* **7**, 3731–3736.
- Turner, T. T., Johnston, D. S. & Jelinsky, S. A. 2006 Epididymal genomics and the search for a male contraceptive. *Mol. Cell Endocrinol.* **250**, 178–183. (doi:10.1016/j.mce.2005.12.042)
- Venter, J. C. et al. 2001 The sequence of the human genome. *Science* **291**, 1304–1351. (doi:10.1126/science.1058040)
- Wang, H., Zhou, Z., Xu, M., Li, J., Xiao, J., Xu, Z. Y. & Sha, J. 2004 A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. *J. Mol. Med.* **82**, 317–324. (doi:10.1007/s00109-004-0526-3)
- Wang, Z., Gerstein, M. & Snyder, M. 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63. (doi:10.1038/nrg2484)
- Waterston, R. H. et al. 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562. (doi:10.1038/nature01262)
- Wilhelm, D., Huang, E., Svingen, T., Stanfield, S., Dinnis, D. & Koopman, P. 2006 Comparative proteomic analysis to study molecular events during gonad development in mice. *Genesis* **44**, 168–176. (doi:10.1002/dvg.20200)
- Wilhelm, D., Palmer, S. & Koopman, P. 2007 Sex determination and gonadal development in mammals. *Physiol. Rev.* **87**, 1–28. (doi:10.1152/physrev.00009.2006)
- Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphrey-Smith, I., Hochstrasser, D. F. & Williams, K. L. 1996 Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**, 19–50.
- Wrobel, G. & Primig, M. 2005 Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction. *Reproduction* **129**, 1–7. (doi:10.1530/rep.1.00408)
- Wu, S. M. et al. 2004 Analysis of mouse germ-cell transcriptome at different stages of spermatogenesis by SAGE: biological significance. *Genomics* **84**, 971–981. (doi:10.1016/j.ygeno.2004.08.018)

- Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y. & Saitou, M. 2006 Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biol. Reprod.* **75**, 705–716. (doi:10.1095/biolreprod.106.053686)
- Yan, N., Lu, Y., Sun, H., Tao, D., Zhang, S., Liu, W. & Ma, Y. 2007 A microarray for microRNA profiling in mouse testis tissues. *Reproduction* **134**, 73–79. (doi:10.1530/REP-07-0056)
- Yan, N. *et al.* 2009 Microarray profiling of microRNAs expressed in testis tissues of developing primates. *J. Assist. Reprod. Genet.* **26**, 26.
- Yao, J., Chiba, T., Sakai, J., Hirose, K., Yamamoto, M., Hada, A., Kuramoto, K., Higuchi, K. & Mori, M. 2004 Mouse testis transcriptome revealed using serial analysis of gene expression. *Mamm. Genome* **15**, 433–451. (doi:10.1007/s00335-004-2347-7)
- Zhao, Y., Li, Q., Yao, C., Wang, Z., Zhou, Y., Wang, Y., Liu, L., Wang, L. & Qiao, Z. 2006 Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. *Hum. Reprod.* **21**, 1583–1590. (doi:10.1093/humrep/del027)
- Zhao, C., Huo, R., Wang, F. Q., Lin, M., Zhou, Z. M. & Sha, J. H. 2007 Identification of several proteins involved in regulation of sperm motility by proteomic analysis. *Fertil. Steril.* **87**, 436–438. (doi:10.1016/j.fertnstert.2006.06.057)
- Zhao, C. *et al.* 2009 Role of translation by mitochondrial-type ribosomes during sperm capacitation: an analysis based on a proteomic approach. *Proteomics* **9**, 1385–1399. (doi:10.1002/pmic.200800353)
- Zhou, Q. *et al.* 2008 Expression of stimulated by retinoic acid gene 8 (Stra8) and maturation of murine gonocytes and spermatogonia induced by retinoic acid *in vitro*. *Biol. Reprod.* **78**, 537–545. (doi:10.1095/biolreprod.107.064337)
- Zhu, Y. F. *et al.* 2006 Proteomic analysis of effect of hyperthermia on spermatogenesis in adult male mice. *J. Proteome Res.* **5**, 2217–2225. (doi:10.1021/pr0600733)