Title:

Linking transcriptomics and proteomics in spermatogenesis

Frédéric Chalmel[1, §] and Antoine D. Rolland[1, §]

[1] Inserm U1085-Irset; Université de Rennes 1; F-35042 Rennes, France.

[§] To whom correspondence should be addressed

Email addresses:

FC: frederic.chalmel@inserm.fr

ADR: antoine.rolland@univ-rennes1.fr

Key words: integrative genomics; proteogenomics; reprogenomics.

Abstract

Spermatogenesis is a complex and tightly regulated process leading to the continuous production of male gametes, the spermatozoa. This developmental process requires the sequential and coordinated expression of thousands of genes, including many that are testis-specific. The molecular networks underlying normal and pathological spermatogenesis have been widely investigated in recent decades, and many high-throughput expression studies have studied genes and proteins involved in male fertility. In this review, we focus on studies that have attempted to correlate transcription and translation during spermatogenesis by comparing the testicular transcriptome and proteome. We also discuss the recent development and use of new transcriptomic approaches that provide a better proxy for the proteome, from both qualitative and quantitative perspectives. Finally, we provide illustrations of how testis-derived transcriptomic and proteomic data can be integrated to address new questions and of how the "proteomics informed by transcriptomics" technique, by combining RNA-seq and MS-based proteomics, can contribute significantly to the discovery of new protein-coding genes or new protein isoforms expressed during spermatogenesis.

Introduction: Unraveling testis specificities with omics technologies

From a genomist's point of view, spermatogenesis, especially in mammals, is arguably one of the most exciting objects of study available. Not only does this developmental process offer incredible molecular dynamics, but it also embodies several expression specificities and striking genomic features.

To make motile sperm capable of fertilization, germ cells must undergo unique processes, such as meiosis, and develop specific organelles and cell structures, including the acrosome, the flagellum, and a highly condensed nucleus. This extreme differentiation process involves the functions of specific molecular factors, many of them expressed only during spermatogenesis. High-throughput tissue-profiling experiments thus regularly identify the testis as the organ that expresses the greatest number of tissue-specific genes and proteins (Chalmel *et al*., 2007, 2012; Kouadjo *et al*., 2007; Fagerberg *et al*., 2014; Uhlen *et al*., 2015). Similarly, the finding that the testis contains the highest number of alternative splicings (Xu *et al*., 2002; Yeo *et al*., 2004; Kan *et al*., 2005; de la Grange *et al*., 2010) indicates that what is true for genes and proteins also applies to isoforms.

Evolutionarily speaking, genes involved in male germ cell development are also quite remarkable. For instance, testis-expressed genes show the highest divergence rate between species for both sequence and expression (Khaitovich *et al*., 2005; Voolstra *et al*., 2007). Additionally, testicular transcripts have, yet again, the highest number of diverged alternative splicings (Kan *et al*., 2005). This fast evolution of male fertility-related factors is thought to result from sexual selection, a specific pressure selection that enables mutations providing a reproductive advantage to be transmitted more easily to progeny and thus fixed within a species relatively quickly.

Another striking genomic feature observed during germ cell development concerns sexual chromosomes and especially X-linked genes (for review, see Hu and Namekawa, 2015). Throughout the evolution of eutherian species, the Y chromosome has become progressively shorter, thus preventing the X and Y chromosomes from aligning/pairing with any precision during meiotic prophase I. Accordingly, to prevent misalignments and recombinations, sexual chromosomes condense into a specific nuclear structure named the sex- or XY-body (Solari, 1974; McKee and Handel, 1993). The strong condensation of X and Y chromosomes within this structure leads to their complete transcriptional silencing (Turner, 2007). This so-called meiotic sex chromosome inactivation (MSCI) has been demonstrated clearly at the genome-wide level: several high-throughput expression studies have failed to find the expression of a single X-linked gene during the meiotic phase of spermatogenesis (Namekawa *et al*., 2006; Chalmel *et al*., 2007). Additionally, to compensate for the cessation of transcription of crucial genes for any cell, a large number of X-linked genes have, over the course of evolution, been transposed onto autosomes and acquired specific meiotic and post-meiotic expression (Potrzebowski *et al*., 2008, 2010). Finally, because the heterozygous nature of sexual chromosomes in males allows advantageous reproductive traits to be fixed quickly, X-linked genes are overrepresented among those preferentially expressed in testicular somatic cells, spermatogonia, and post-meiotic spermatids (Khil *et al*., 2004; Chalmel *et al*., 2007; Mueller *et al*., 2008).

Taken together, the specificities of the male germ cell expression program provide a rich environment for studying regulatory mechanisms of gene expression at various levels as well as for the discovery of new genes and protein isoforms.

As many past studies investigating spermatogenesis with high-throughput approaches have been reviewed elsewhere (Rolland *et al*., 2008; Calvel *et al*., 2010; Chocu *et al*., 2012), here we focus on studies that have attempted to link the transcriptome and proteome in spermatogenesis or have combined transcriptomic and proteomic data to gain insight into testicular functions and germ cell biology (Table 1).

Integrative omics strategies to study spermatogenesis

The integration of various types of omics data, *e.g.*, epigenomic, transcripomic, proteomic, interactomic, or regulomic, represents a powerful tool for going far beyond basic descriptive analyses. Combining information from different samples and technologies makes it possible to improve data consistency by refining candidate selection, as well as to address more specific questions and to build new hypotheses (Figure 1A).

For example, microarray and proteomic data from mouse testes lacking DICER in the Sertoli cells ($DCR^{fx/fx}$*;MisCre*) were compared to investigate the miRNA-mediated post-transcriptional regulation in these cells (Papaioannou *et al*., 2011). This approach allowed the authors to identify miRNA-targets within Sertoli cells, *i.e.*, proteins whose abundance increases in KO mice, even though the expression of their corresponding mRNAs does not change. Subsequently they performed 3'UTR luciferase assays to validate SOD-1 as a likely direct target of miR-125a-3p, miR-872 and miR-24.

Many studies have also combined proteomic and transcriptomic data to improve the characterization of the expression landscape during spermatogenesis. Chalmel *et al.,* using biopsies from infertile patients with spermatogenesis arrested at various stages of germ cell development, first identified genes preferentially expressed in each type of testicular cell (Chalmel *et al*., 2012).

Gene expression data from various tissues and antibody-based profiling data from the Human Protein Atlas (http://www.proteinatlas.org; Uhlen *et al*., 2010) were then used to identify, respectively, the genes and gene products showing testis-specific expression, most of them being found to be expressed in meiotic and post-meiotic germ cells. Finally, taking advantage of available information on protein-gene interactions (*i.e.*, regulomic data), these authors filtered for a core network of transcription factors and DNA-binding proteins that are likely to drive the germ cell-specific expression program. Conversely, Djureinovic and colleagues sought to determine the human testis-specific proteome, beginning with the RNA-seq analysis of 27 tissues (Djureinovic *et al*., 2014). They then interrogated the Human Protein Atlas about the testis-enriched transcripts they identified to confirm the testis-specificity of corresponding protein products and to identify the testicular cell type(s) in which they were expressed.

The combination of transcriptomic and proteomic approaches has also been very successful in helping to interpret the content of biological fluids or transcriptionally-inactive cells, such as spermatozoa. For instance, Rolland *et al.* compiled several human seminal plasma proteomic studies and compared the resulting proteome to gene expression data for the organs contributing to this biological fluid, *i.e.*, the testis, epididymis, seminal vesicle, and prostate (Rolland *et al*., 2013). This allowed the identification of protein biomarkers for each of the male genital tract organs; importantly these biomarkers, including germ cell markers, can thus be monitored in semen.

A recent investigation of the intricate question of Sertoli-germ cell communication used another such integrative approach (Chalmel *et al*., 2014a); it analyzed the testicular fluid proteome of rats and rams and then combined it with transcriptomic data from isolated testicular cells (Chalmel *et al*., 2007) and with protein-protein interaction data.

The authors were therefore able to identify testicular fluid proteins likely to be secreted by Sertoli cells and to interact with germ cell membrane proteins and, conversely, proteins secreted by germ cells that might interact with Sertoli cell membrane proteins. Among these, the interactions of APOH and CDC42 and of APP and NGFR were further investigated and validated *in situ.* Finally, Wang *et al.* compiled several proteomic studies to determine the proteome of human spermatozoa and used gene expression tissue-profiling data to identify the sperm proteins specifically expressed in the testis (Wang *et al*., 2013). With this candidate list they next queried the drug target information available in Drugbank (Wishart *et al*., 2008) to identify potential male contraceptive molecules. Disulfiram and propofol, two molecules thought to target cilia proteins, were then shown to inhibit sperm motility.

Correlating transcription and translation rates during spermatogenesis

Transcriptomic studies often assume that the abundance of mRNAs and that of their corresponding proteins are well correlated. This hypothesis is considered to justify extrapolation from changes in gene expression to changes in protein expression and, ultimately, to their potential functional impact. The rationale of this hypothesis may appear quite reasonable: typical microarray or RNA-seq protocols involve an oligo-dT selection of polyadenylated mRNAs, which are thought to be actively translated, unlike those that are not polyadenylated.

The fate of an mRNA, however, is tightly regulated by a complex interplay of modification, processing, storage, decay, and translation, all involving protein-RNA interactions through messenger ribonucleoprotein (mRNP) complexes. Some of these assembled complexes are conducted directly to translation while other are diverted towards storage and translational repression (for review see Müller-McNicoll and Neugebauer, 2013).

While post-transcriptional and translational gene regulation is a common mechanism in all cell types, it is especially striking during spermatogenesis. As spermatids start to elongate, nuclear histones are sequentially replaced by transition proteins and protamines (for review see Rathke *et al*., 2014). This substitution allows the progressive condensation of chromatin and thus leads to the complete cessation of transcription from mid-spermiogenesis onwards (Kierszenbaum and Tres, 1975). As a consequence, many genes that are required for the development and/or functioning of spermatozoa are transcribed much earlier during germ cell differentiation, then translationally repressed, and finally translated several days after the mRNA production, thanks to a complex interplay of RNA-binding proteins and non-coding RNA (for review, see Kleene, 2013). In this context, it is interesting to note the existence of the chromatoid body, a germ cell-specific RNA processing center suggested to be involved in the sequestration and translation repression of several mRNAs during spermiogenesis (Kotaja and Sassone-Corsi, 2007) and whose RNA and protein content was recently analyzed (Meikar *et al*., 2014). However, the direct contribution of this organelle to translational regulation remains to be clearly demonstrated (for review, see Kleene and Cullinane, 2011). More importantly, the use of cross-linking immunoprecipitation (CLIP) together with microarray analysis (CLIP-chip) or high-throughput sequencing (HITS-CLIP or CLIP-seq) has allowed some potential direct targets of RNA-binding proteins to be identified in male germ cells (Reynolds *et al*., 2005; Grellscheid *et al*., 2011; Vourekas *et al*., 2012; Zhang *et al*., 2015). The combination of such approaches with proteomic analyses of mutant mice would in turn help identify which mRNAs are actually translationally regulated by these specific factors during spermatogenesis.

Because of this prominent uncoupling between transcription and translation, the testis is often seen as an organ in which transcriptome and proteome are not necessarily linked.

This low correlation between mRNA and protein concentrations within the testis was clearly evidenced in a tissue-profiling experiment that used multidimensional protein identification technology (MudPIT) for human tissue (Cagney *et al*., 2005). In this study, the abundance of 683 proteins and their corresponding transcripts were measured and compared in eight organs. Interestingly, the gene profiles of all organs clustered together, as did their protein profiles. This finding suggests that transcriptomes or proteomes from different organs are more similar than the transcriptome and proteome of the same organ. Importantly, the correlation between transcriptome and proteome data was weakest for the testis, and highest for the liver (correlation coefficients of 0.138 and 0.432, respectively). To investigate the relation between proteins and mRNA levels during spermatogenesis in more detail, Gan and collaborators used isolated type A spermatogonia, pachytene spermatocytes, round spermatids, and elongated spermatids for an iTRAQ-based proteomic analysis of male germ cell differentiation (Gan *et al*., 2013a) and compared their results with those of a previously published microarray dataset (Namekawa *et al*., 2006). While they found a consistent match for a subset of transcriptomic and proteomic profiles, they also observed that several regulation mechanisms - including transcript degradation, translation repression, translation de-repression, and protein degradation - affected most genes and may account for the low correlation between mRNAs and proteins, at both the mitosis/meiosis transition (Pearson correlation of 0.55) and the meiosis/post-meiosis transition (Pearson correlation of 0.41).

Investigating the translatome of testicular cells

Another exciting possibility for bridging the gap between gene expression and protein abundance lies in methods that allow investigation of the translatome, *i.e.*, the measurement of transcripts that are actively processed by the translational machinery (Figure 1B).

These analyses usually involve the purification of ribosomes or polysomes and the subsequent measurement of associated transcripts. Iguchi and collaborators first applied one such approach to address the question of translational regulation during the meiotic and post-meiotic phases of male germ cell development (Iguchi *et al*., 2006). The authors monitored the mRNAs associated with free RNPs and polysomes in the testes from mice at various postnatal stages and identified translationally up- and down-regulated transcripts, *i.e.*, mRNAs significantly redistributed between free RNPs and the polysomal fraction during testis development. Not surprisingly, translation increased for most of these mRNAs, in elongating spermatids; this increase reflects a common mechanism compensating for the cessation of transcription from mid-spermiogenesis onwards. Nonetheless, they also identified a small cluster of meiotically-induced mRNAs that were actively translated only in post-meiotic stages.

More recently, the development of genetically modified organisms that express an affinity-tagged ribosomal protein has provided a straightforward means of isolating ribosomes along with their bound mRNAs. Interestingly, the expression of these tagged proteins can be driven by a tissue/cell-specific promoter, such as the Cre-lox system in mice, which enables the capture of tagged ribosomes from an entire organ or tissue without the need to isolate the cells of interest (for review, see King and Gerber, 2014). These methodologies, initially called translating ribosome affinity purification (TRAP) in the mouse (Doyle *et al*., 2008; Heiman *et al*., 2008), have been used several times to capture the translatome of various testicular cell types. For instance, Sanz and collaborators took advantage of Cyp17iCre and Amh-Cre mice to investigate the translatome of adult Leydig cells and Sertoli cells, respectively (Sanz *et al*., 2013). They also used this approach to investigate the regulation of Leydig cells by LH and that of Sertoli cells by FSH and testosterone in gonadotropin-depleted mice.

They notably found that the early response to LH (within one hour) was characterized by the induction of several transcription factors and genes involved in cell cycle while the secondary response to LH (after four hours) involved the up-regulation of genes involved in steroid metabolism and FGF signaling and the down-regulation of several transcription repressors. The regulatory role of testosterone was also successfully examined by De Gendt and colleagues who combined a TRAP approach that used Amh-Cre mice, a mouse model lacking a functional androgen receptor (AR) in Sertoli cells, and RNA-seq analysis (De Gendt *et al*., 2014). After determining the Sertoli cell translatomes of prepubertal and adult mice, which they found to be very similar, the authors compared these wild-type Sertoli cell translatomes to that of Sertoli cells lacking the AR and identified androgen-regulated genes at the onset of meiosis, which included many plasma membrane and cytoskeleton factors involved in cell junction and adhesion. Finally, another study took advantage of this method to investigate the translatome of neonatal testicular germ cells at the onset of meiosis (Evans *et al*., 2014). Using a synchronized spermatogenesis model, the authors identified the changes in ribosome-bound mRNAs taking place in both differentiating spermatogonial cells (with Ngn3-Cre and Stra8-Cre mice) and maturing Sertoli cells (with Amh-Cre mice) after retinoic acid restoration.

Note that these ribosomal profiling analyses not only help to evaluate translation efficiency and estimate corresponding protein abundance more accurately than classical transcriptomic approaches, but, when coupled with RNA-seq, they can also provide information about ribosome occupancy, translation initiation, elongation, and termination at near-nucleotide resolution (for review, see Ingolia, 2014).

From gene expression measurement to new testicular protein isoform prediction

As mentioned above, the measurement of steady-state gene expression does not necessarily estimate the actual proteome well. This is true from both the quantitative and qualitative points of view and especially for microarray experiments. Specifically, because microarrays measure gene expression through the sequence-specific hybridization of RNAs to DNA probes, they cannot gather information about transcript structure outside the sequence targeted by the probes. Furthermore, because most probes recognize several transcript isoforms, they report average gene expression and fail to identify the specific isoforms actually expressed in a given sample.

In this regard, the recent advance of RNA-sequencing (RNA-seq) technologies, together with the development of associated analysis pipelines, has revolutionized the field of transcriptomics. RNA-seq is an efficient and cost-effective way to obtain large amount of transcriptome data and identify both new genes and new isoforms, by the sequencing of novel exons and/or novel exonic junctions. RNA-seq thus makes it simultaneously possible to determine the structure of thousands of transcripts and to measure their abundance. It thus provides a more accurate prediction of all corresponding protein isoforms (Figure 1B). Several RNA-seq analyses have already been conducted to investigate spermatogenesis in rodents, with either isolated cells (Gan *et al*., 2013b; Soumillon *et al*., 2013; Chalmel *et al*., 2014b) or testes at various stages of the first wave of spermatogenesis (Laiho *et al*., 2013; Schmid *et al*., 2013; Margolin *et al*., 2014). All these studies have led to the reconstruction of a plethora of transcripts, including known isoforms but also thousands of new isoforms of known genes and hundreds of uncharacterized transcripts that correspond to either new coding or non-coding genes. The amount of information generated in such RNA-seq studies is so huge that a single study cannot undertake and report on all the exploration possibilities.

For example, Chalmel and colleagues restricted their analysis to novel unannotated loci (Chalmel *et al*., 2014b) and used four different bioinformatics tools to distinguish between transcripts with high and low coding potential. Margolin and collaborators initiated a broad analysis of splicing events by identifying transcripts that contained novel splice junctions in which the open reading frame (ORF) was maintained (Margolin *et al*., 2014). Finally, Schmid and colleagues focused their study on the splicing dynamics of mRNAs in male germ cells, with a special emphasis on the newly identified splicing events that might affect protein isoform production during mouse meiosis, and identified significantly enriched motifs for PTB, TRA2B and STAR proteins in and around meiotically-regulated cassette exons (Schmid *et al*., 2013). Therefore, although all these RNA-seq studies have highlighted many potential new protein-coding transcripts expressed in male germ cells, a thorough analysis aimed specifically at identifying the variants that actually code for specific proteoforms with distinct biological functions is still needed. Most important, the functional relevance and validity of these findings still require experimental validation at the protein level.

Proteogenomic approaches applied to spermatogenesis

Compared with microarrays, which are intrinsically limited to studying the expression of genes for which probes are spotted on their surface, MS-based proteomics has long been considered to be more powerful in the sense that theoretically it can detect and quantify any protein entity within a given sample. Protein identification, however, typically involves the comparison of experimental masses obtained by mass spectrometry to that of *in silico*-digested protein databases. Therefore proteomic studies are also limited to the sequence content of the database that is used for identification purposes.

Several methods, called proteogenomics, have emerged to overcome this limitation and help identify novel peptides not present in reference protein sequence databases (for review, see Hernandez *et al*., 2014). These methods rely on the construction of customized protein sequence databases that include, for example, products resulting from the 6-frame translation of the reference genome or the 3-frame translation of transcripts, or both, regardless of whether these correspond to known mRNAs, non-coding RNAs, or pseudogenes. Large consortia seeking to decipher the complete human proteome through the analysis of several human tissues, including the testis (Kim *et al*., 2014; Wilhelm *et al*., 2014), have recently used such approaches.

Among the most promising proteogenomic methods is RNA-seq-based proteogenomics, also called "proteomics informed by transcriptomics" (PIT). The PIT strategy derives the customized protein sequence database queried for protein identification directly from RNA-seq data of the same or a similar sample (Evans et al., 2012). It therefore limits protein products in the custom database to those resulting from the 3- or 6-frame translation of the assembled transcripts that are indeed expressed in the organ, tissue, or cell type of interest. Compared with other proteogenomic approaches, PIT offers the advantage of a smaller database, which in turn reduces the number of false positives and increases sensitivity (Figure 1A).

Recently, this strategy was applied to the identification of new proteins expressed during late stages of rat spermatogenesis (Chocu *et al*., 2014). In this study, protein extracts from isolated rat pachytene spermatocytes and round spermatids were first trypsin-digested and analyzed by nano LC-MS/MS. Next, MS/MS spectra were queried against a customized database derived from a previous RNA-seq analysis of rat testicular cells (Chalmel *et al*., 2014b), which had identified almost 12,000 new transcript isoforms. It also reported the existence of more than 1400 completely new unannotated loci, most of them preferentially expressed in spermatocytes and/or spermatids.

Because of this high gene discovery potential, Chocu and colleagues voluntarily restricted their PIT approach to meiotic and post-meiotic germ cells. This experiment led to the identification of 44 novel coding genes expressed during rat spermatogenesis, including 14 that were initially thought to correspond to non-coding RNAs. This approach has also been used to study the testicular proteome of the red abalone, *Haliotis rufescens* (Palmer *et al*., 2013) and allowed the identification of almost 1000 proteins. This number of proteins is especially remarkable when we consider that the number of UniProt entries still does not exceed 140 for this non-model species (Release 2014_11).

There is no doubt that the increasing performance of mass spectrometers and the decreasing cost of RNA-seq will allow the rapid democratization of PIT studies and of proteogenomics in general. These approaches will be critical to the full characterization of both the transcriptome and the proteome of model organisms in various biological contexts, which in turn will help to annotate the corresponding genomes. PIT strategies are also a unique opportunity for non-model species, for which reference genome sequences are not available: their transcriptomes and proteomes can be thoroughly examined without requiring the use of nucleic or protein sequence databases from phylogenetically distant species. Finally, regardless of the model of interest, the combination of RNA-seq and mass spectrometry into a PIT study offers a straightforward method of investigating the correlation of transcriptomes and proteomes, because protein profiles can be directly compared to transcript profiles on which protein identifications are also performed (for review, see Wang *et al*., 2014).

Conclusion

The recent progress in next-generation sequencing technologies together with the improved performance of mass spectrometers has made possible a fruitful revisit of the testis genomic landscape. While we are now getting close to the complete identification of the molecular factors involved in spermatogenesis, an understanding of all the regulatory mechanisms that drive gene and protein expression during germ cell development and the identification of the key factors for male fertility both require additional work. This will notably imply the combination of all types of available data, *i.e.*, from epigenomic, regulomic, transcriptomic, proteomic, and interactomic studies, in order to link the flow of information from DNA to functional proteins and non-coding RNAs. The success of this integrative work will also depend on the development of new types of web servers, such as the ReproGenomics Viewer (http://rgv.genouest.org/; Darde *et al.*, 2015), which allows the visualization, mining, and comparison of various types of omics data (*e.g.*, ChIP-seq, RNA-seq, MS-based proteomics) in a multi- and cross-species manner. Finally, a current challenge in biology resides in the development of methods to investigate single cells at the genomic, transcriptomic, proteomic, and metabolomic level (for review, see Tsioris *et al.*, 2014). The use of these so-called single-cell approaches will mandate the more detailed study of the kinetics of germ cell differentiation and, most importantly, enable us to gain insight into the biology of discrete cell populations within the testis, such as the spermatogonial stem cell.

Funding

This work was supported by the Institut national de la santé et de la recherche médicale (Inserm), the Université de Rennes 1, the Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail [ANSES n°EST-13-081 to F.C.], the Fondation pour la recherche médicale [FRM n°DBI20131228558 to F.C.], and the European Union [FEDER to F.C].

Conflict of interest statement

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported

Acknowledgements

We thank all members of the Institute for research in Health, Environment and Work for stimulating discussions.

References

Cagney G, Park S, Chung C, Tong B, O'Dushlaine C, Shields DC and Emili A (2005) Human tissue profiling with multidimensional protein identification technology. *Journal of Proteome Research* **4** 1757–1767.

Calvel P, Rolland AD, Jégou B and Pineau C (2010) Testicular postgenomics: targeting the regulation of spermatogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365** 1481–1500.

Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J-J, Wolgemuth DJ, Jégou B *et al.* (2007) The conserved transcriptome in human and rodent male gametogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **104** 8346–8351.

Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C *et al.* (2012) Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Human Reproduction (Oxford, England)* **27** 3233–3248.

Chalmel F, Com E, Lavigne R, Hernio N, Teixeira-Gomes A-P, Dacheux J-L and Pineau C (2014a) An integrative omics strategy to assess the germ cell secretome and to decipher sertoli-germ cell crosstalk in the Mammalian testis. *PloS One* **9** e104418.

Chalmel F, Lardenois A, Evrard B, Rolland AD, Sallou O, Dumargne M-C, Coiffec I, Collin O, Primig M and Jégou B (2014b) High-resolution profiling of novel transcribed regions during rat spermatogenesis. *Biology of Reproduction* **91** 5.

Chocu S, Calvel P, Rolland AD and Pineau C (2012) Spermatogenesis in mammals: proteomic insights. *Systems Biology in Reproductive Medicine* **58** 179–190.

Chocu S, Evrard B, Lavigne R, Rolland AD, Aubry F, Jégou B, Chalmel F and Pineau C (2014) Forty-four novel protein-coding loci discovered using a proteomics informed by transcriptomics (PIT) approach in rat male germ cells. *Biology of Reproduction* **91** 123.

Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Le Bras Y, Jégou B, Collin O, Rolland AD and Chalmel F (2015) The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Research*.

Djureinovic D, Fagerberg L, Hallström B, Danielsson A, Lindskog C, Uhlén M and Pontén F (2014) The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Molecular Human Reproduction* **20** 476–488.

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML *et al.* (2008) Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. *Cell* **135** 749–762.

Evans VC, Barker G, Heesom KJ, Fan J, Bessant C and Matthews DA (2012) De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nature Methods* **9** 1207–1211.

Evans E, Hogarth C, Mitchell D and Griswold M (2014) Riding the spermatogenic wave: profiling gene expression within neonatal germ and sertoli cells during a synchronized initial wave of spermatogenesis in mice. *Biology of Reproduction* **90** 108.

Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics : MCP* **13** 397–406.

Gan H, Cai T, Lin X, Wu Y, Wang X, Yang F and Han C (2013a) Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Molecular & Cellular Proteomics : MCP* **12** 1144–1157.

Gan H, Wen L, Liao S, Lin X, Ma T, Liu J, Song C-X, Wang M, He C, Han C *et al.* (2013b) Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nature Communications* **4** 1995.

De Gendt K, Verhoeven G, Amieux PS and Wilkinson MF (2014) Research Resource: Genome-Wide Identification of AR-Regulated Genes Translated in Sertoli Cells In Vivo Using the RiboTag Approach. *Molecular Endocrinology (Baltimore, Md.)* **28** 575–591.

Grellscheid S, Dalgliesh C, Storbeck M, Best A, Liu Y, Jakubik M, Mende Y, Ehrmann I, Curk T, Rossbach K *et al.* (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2β in development. *PLoS Genetics* **7** e1002390.

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ *et al.* (2008) A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. *Cell* **135** 738–748.

Hernandez C, Waridel P and Quadroni M (2014) Database construction and peptide identification strategies for proteogenomic studies on sequenced genomes. *Current Topics in Medicinal Chemistry* **14** 425–434.

Hu Y-C and Namekawa SH (2015) Functional significance of the sex chromosomes during spermatogenesis. *Reproduction (Cambridge, England)* **149** R265–R277.

Iguchi N, Tobias JW and Hecht NB (2006) Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proceedings of the National Academy of Sciences of the United States of America* **103** 7712–7717.

Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews. Genetics* **15** 205–213.

Kan Z, Garrett-Engele PW, Johnson JM and Castle JC (2005) Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. *Nucleic Acids Research* **33** 5659–5666.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M and Pääbo S (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science (New York, N.Y.)* **309** 1850–1854.

Khil PP, Smirnova NA, Romanienko PJ and Camerini-Otero RD (2004) The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature Genetics* **36** 642–646.

Kierszenbaum AL and Tres LL (1975) Structural and transcriptional features of the mouse spermatid genome. *The Journal of Cell Biology* **65** 258–270.

Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S *et al.* (2014) A draft map of the human proteome. *Nature* **509** 575–581.

King HA and Gerber AP (2014) Translatome profiling: methods for genome-scale analysis of mRNA translation. *Briefings in Functional Genomics*.

Kleene KC (2013) Connecting cis-elements and trans-factors with mechanisms of developmental regulation of mRNA translation in meiotic and haploid mammalian spermatogenic cells. *Reproduction (Cambridge, England)* **146** R1–R19.

Kleene KC and Cullinane DL (2011) Maybe repressed mRNAs are not stored in the chromatoid body in mammalian spermatids. *Reproduction (Cambridge, England)* **142** 383–388.

Kotaja N and Sassone-Corsi P (2007) The chromatoid body: a germ-cell-specific RNA-processing centre. *Nature Reviews. Molecular Cell Biology* **8** 85–90.

Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M and St-Amand J (2007) Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8** 127.

De la Grange P, Gratadou L, Delord M, Dutertre M and Auboeuf D (2010) Splicing factor and exon profiling across human tissues. *Nucleic Acids Research* **38** 2825–2838.

Laiho A, Kotaja N, Gyenesei A and Sironen A (2013) Transcriptome Profiling of the Murine Testis during the First Wave of Spermatogenesis. *PLoS ONE* **8**.

Margolin G, Khil PP, Kim J, Bellani MA and Camerini-Otero RD (2014) Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics* **15** 39.

McKee BD and Handel MA (1993) Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102** 71–80.

Meikar O, Vagin V V, Chalmel F, Sõstar K, Lardenois A, Hammell M, Jin Y, Da Ros M, Wasik KA, Toppari J *et al.* (2014) An atlas of chromatoid body components. *RNA (New York, N.Y.)* **20** 483–495.

Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC and Turner JMA (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics* **40** 794–799.

Müller-McNicoll M and Neugebauer KM (2013) How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews. Genetics* **14** 275–287.

Namekawa SH, Park PJ, Zhang LF, Shima JE, McCarrey JR, Griswold MD and Lee JT (2006) Postmeiotic Sex Chromatin in the Male Germline of Mice. *Current Biology* **16** 660–667.

Palmer MR, Mcdowall MH, Stewart L, Ouaddi A, Maccoss MJ and Swanson WJ (2013) Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. *Molecular Reproduction and Development* **80** 460–465.

Papaioannou MD, Lagarrigue M, Vejnar CE, Rolland AD, Kühne F, Aubry F, Schaad O, Fort A, Descombes P, Neerman-Arbez M *et al.* (2011) Loss of Dicer in Sertoli cells has a major impact on the testicular proteome of mice. *Molecular & Cellular Proteomics : MCP* **10** M900587MCP200.

Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jégou B and Kaessmann H (2008) Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biology* **6** e80.

Potrzebowski L, Vinckenbosch N and Kaessmann H (2010) The emergence of new genes on the young therian X. *Trends in Genetics : TIG* **26** 1–4.

Rathke C, Baarends WM, Awe S and Renkawitz-Pohl R (2014) Chromatin dynamics during spermiogenesis. *Biochimica et Biophysica Acta* **1839** 155–168.

Reynolds N, Collier B, Maratou K, Bingham V, Speed RM, Taggart M, Semple CA, Gray NK and Cooke HJ (2005) Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. *Human Molecular Genetics* **14** 3899–3909.

Rolland AD, Jégou B and Pineau C (2008) Testicular development and spermatogenesis: harvesting the postgenomics bounty. *Advances in Experimental Medicine and Biology* **636** 16–41.

Rolland AD, Lavigne R, Dauly C, Calvel P, Kervarrec C, Freour T, Evrard B, Rioux-Leclercq N, Auger J and Pineau C (2013) Identification of genital tract markers in the human seminal plasma using an integrative genomics approach. *Human Reproduction (Oxford, England)* **28** 199–209.

Sanz E, Evanoff R, Quintana A, Evans E, Miller JA, Ko C, Amieux PS, Griswold MD and McKnight GS (2013) RiboTag Analysis of Actively Translated mRNAs in Sertoli and Leydig Cells In Vivo. *PLoS ONE* **8**.

Schmid R, Grellscheid SN, Ehrmann I, Dalgliesh C, Danilenko M, Paronetto MP, Pedrotti S, Grellscheid D, Dixon RJ, Sette C *et al.* (2013) The splicing landscape is globally reprogrammed during male meiosis. *Nucleic Acids Research* **41** 10170–10184.

Solari AJ (1974) The behavior of the XY pair in mammals. *International Review of Cytology* **38** 273–317.

Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A *et al.* (2013) Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Reports* **3** 2179–2190.

Tsioris K, Torres AJ, Douce TB and Love JC (2014) A new toolbox for assessing single cells. *Annual Review of Chemical and Biomolecular Engineering* **5** 455–477.

Turner JMA (2007) Meiotic sex chromosome inactivation. *Development (Cambridge, England)* **134** 1823–1831.

Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S *et al.* (2010) Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology* **28** 1248–1250.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A *et al.* (2015) Tissue-based map of the human proteome. *Science* **347** 1260419–1260419.

Voolstra C, Tautz D, Farbrother P, Eichinger L and Harr B (2007) Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Research* **17** 42–49.

Vourekas A, Zheng Q, Alexiou P, Maragkakis M, Kirino Y, Gregory BD and Mourelatos Z (2012) Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nature Structural & Molecular Biology* **19** 773–781.

Wang G, Guo Y, Zhou T, Shi X, Yu J, Yang Y, Wu Y, Wang J, Liu M, Chen X *et al.* (2013) In-depth proteomic analysis of the human sperm reveals complex protein compositions. *Journal of Proteomics* **79** 114–122.

Wang X, Liu Q and Zhang B (2014) Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. *Proteomics* **14** 2676–2687.

Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H *et al.* (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509** 582–587.

Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B and Hassanali M (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research* **36** D901–D906.

Xu Q, Modrek B and Lee C (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Research* **30** 3754–3766.

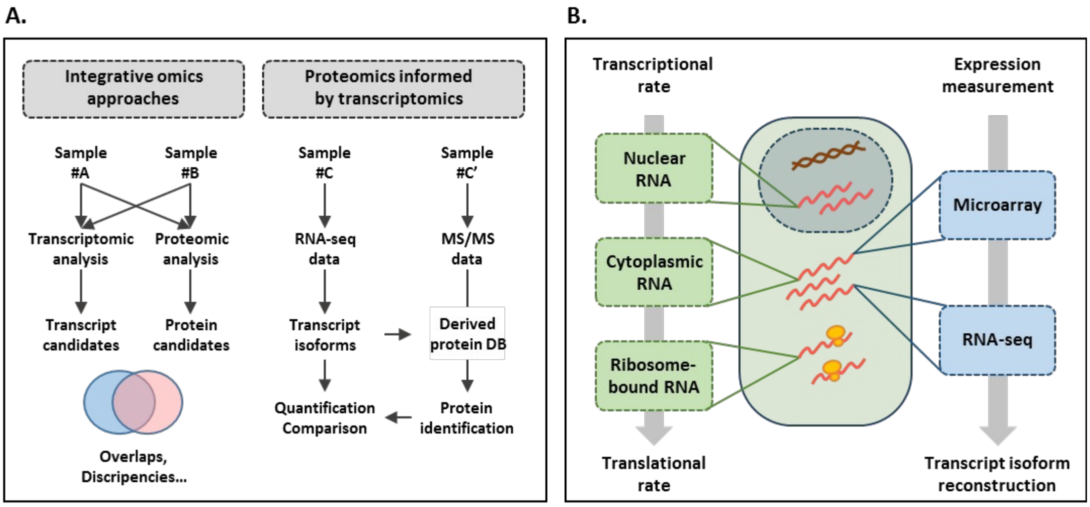
Yeo G, Holste D, Kreiman G and Burge CB (2004) Variation in alternative splicing across human tissues. *Genome Biology* **5** R74.

Zhang P, Kang J-Y, Gou L-T, Wang J, Xue Y, Skogerboe G, Dai P, Huang D-W, Chen R, Fu X-D *et al.* (2015) MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Research* **25** 193–207.

Legends to figures

Figure 1: When transcriptomics meets proteomics

A) Typical integrative omics approaches involve the combination of datasets originating from various technologies, most notably transcriptomics and proteomics. Such strategies are often used to identify more reliable candidates (*i.e.* factors evidenced at both the RNA and protein levels), but they can also be useful in order to compare and correlate transcription and translation rates. More recently, the combination of RNA-seq and mass spectrometry (MS)-based proteomic has led to the development of the so-called Proteomics Informed by Transcriptomics (PIT) approach. In this approach, the protein sequence database (DB) queried for protein identification purpose is directly derived from transcript sequences obtained following RNA-seq analysis of the same or equivalent sample as that used for MS/MS analysis. B) The characterization of the transcriptome has long been used as a proxy for the proteome. However, depending on whether nuclear, total cytoplasmic or ribosome-bound RNAs are analysed, the captured picture will reflect either more the transcriptional rate or the translational rate. Additionally, while both approaches perform equivalently from a quantitative point of view, RNA-seq overcomes microarray technology from a qualitative point of view as it allows full-length transcript reconstruction and can thus discriminate between distinct protein-coding isoforms.


A.
Integrative omics approaches
Sample #A
Sample #B
Transcriptomic analysis
Proteomic analysis
Transcript candidates
Protein candidates
Overlaps, Discripencies...
Proteomics informed by transcriptomics
Sample #C
Sample #C'
RNA-seq data
MS/MS data
Transcript isoforms
Derived protein DB
Quantification Comparison
Protein identification
B.
Transcriptional rate
Expression measurement
Nuclear RNA
Cytoplasmic RNA
Ribosome-bound RNA
Microarray
RNA-seq
Translational rate
Transcript isoform reconstruction