Microarray analysis of gene expression during early development: a cautionary overview

Claude Robert¹,²

¹Centre de recherche en biologie de la reproduction, Faculté des Sciences de l’agriculture et de l’alimentation and ²Laboratory of Functional Genomics of Early Embryonic Development, Laval University, Pavillon Comtois, Local 4221 Université Laval, Québec, Québec, Canada G1K 7P4

Correspondence should be addressed to C Robert at Laboratory of Functional Genomics of Early Embryonic Development, Laval University; Email: claude.robert@fsaa.ulaval.ca

Abstract

The rise of the ‘omics’ technologies started nearly a decade ago and, among them, transcriptomics has been used successfully to contrast gene expression in mammalian oocytes and early embryos. The scarcity of biological material that early developmental stages provide is the prime reason why the field of transcriptomics is becoming more and more popular with reproductive biologists. The potential to amplify scarce mRNA samples and generate the necessary amounts of starting material enables the relative measurement of RNA abundance of thousands of candidates simultaneously. So far, microarrays have been the most commonly used high-throughput method in this field. Microarray platforms can be found in a wide variety of formats, from cDNA collections to long or short oligo probe sets. These platforms generate large amounts of data that require the integration of comparative RNA abundance values in the physiological context of early development for their full benefit to be appreciated. Unfortunately, significant discrepancies between datasets suggest that direct comparison between studies is difficult and often not possible. We have investigated the sample-handling steps leading to the generation of microarray data produced from prehatching embryo samples and have identified key steps that significantly impact the downstream results. This review provides a discussion on the best methods for the preparation of samples from early embryos for microarray analysis and focuses on the challenges that impede dataset comparisons from different platforms and the reasons why methodological benchmarking performed using somatic cells may not apply to the atypical nature of prehatching development.

Introduction

The term ‘omics’ is attributed to any high-throughput methodology that allows surveying and, at least partially, quantification of specific types of cellular molecular components for comparing the following biological treatments. These extremely efficient procedures are being developed at a fast pace. Every level of the classical gene expression cascade, from the DNA sequence to the metabolites produced following a cellular response, can be targeted. Any aspect pertaining to DNA is referred to as genomics, e.g. DNA polymorphisms (from single nucleotide substitution (SNP) Lander et al. 2001, Gibbs et al. 2009) to more complex sequence variations (insertion/deletion and copy number variants; Bae et al. 2010, Conrad et al. 2010), whereas the transcribed genome is studied by the field of transcriptomics. Once protein-coding transcripts have been translated, proteomics takes over and covers protein content as well as posttranslational protein modifications, which include the specific addition of a wide array of residues ranging from chemical groups (phosphorylation, methylation, acetylation, etc.) to sugars (glycosylation and ADP ribosylation) and lipids (merystoylation) or even derivatives of amino acids or addition of small peptides (ubiquitination and sumoylation). In addition, protein–protein interactions (dimerization and oligomerization) also fall under the scope of proteomics. The study of cellular metabolism has its own ‘omics,’ aptly named metabolomics, in which the actions of proteins and of some RNAs responsible for cellular activities that result in by-products are being studied. Finally, the emerging field of epigenomics takes into account the DNA and protein molecules in the context of cellular metabolism through the study of the influence of the surrounding macro- and microenvironments. The impressive level of complexity of the ‘omics’ fields of research is a clear reflection of the complex nature of cell biology and physiology.

Like most fields of research, the study of early embryo development has used the ‘omics’ to study the underlying physiology. The tremendous increment in throughput offered by ‘omics’ has changed the scientific methods by going from a very Platonic perspective, in which the study of a handful of candidates is justified by
a strong _a priori_ background and a clear hypothesis to be tested, to a Socratic scientific approach where observations dictate conclusions and generate hypotheses. Indeed, high-throughput methods rely on the simple assumption that they survey, for instance, the entire protein-coding transcripts and that some of those transcripts must therefore show differential abundance when comparing samples exhibiting different phenotypes. The integration and interpretation of data is therefore performed _a posteriori_.

In the early days of these high-throughput technologies (less than a decade ago), many projects received criticism and were considered fishing expeditions lacking a clear objective. Nowadays, ‘omics’ are buzzwords that suggest ‘state-of-the-art,’ ‘must-use,’ and ‘well-accepted’ technologies.

In the field of embryo development, the scarce source of the starting material, which are oocytes and early embryos, has emphasized transcriptomics since these methods can incorporate global amplification steps to generate sufficient cDNA or aRNA for effective probing of microarrays. The field of genomics also provides the benefit of global amplification when oocytes and embryos are the source of starting material. In contrast, it is still extremely challenging to work at the protein and metabolite levels because of the scarce nature of the starting material and the absence of effective methods to amplify protein and generate sufficient material for screening purposes.

As microarray technologies have been around for nearly a decade, it is widely believed that the characterization of embryonic development at the transcript level is already completed. However, this assumption is far from the truth, since a great proportion of the available data are more or less impossible to fully interpret, if not deeply biased. Indeed, despite the numerous reports found in the literature presenting global gene expression analyses of early development in various species, in-depth interpretations of these datasets that go beyond uninformative gene lists are rare, making it difficult to achieve the overall goal of these analyses, which is to gain valuable insights into the molecular program controlling early development. The focus of this review is to put into perspective the numerous pitfalls that litter the path toward the production of accurate high-throughput datasets of RNA abundance levels in an attempt to describe early embryo development.

**Discussion**

**What is a transcriptome?**

The transcriptome is defined as being composed of all transcribed elements, regardless of whether they are protein-coding sequences or not. When strictly adhering to this definition, we find that none of the current high-throughput platforms dedicated to the study of RNA populations, including deep sequencing, can claim to study the entire transcriptome. For instance, small RNA molecules such as microRNAs are handled (i.e. isolated and processed) differently from longer RNA types. Traditionally, the main focus of transcriptomics has been directed at protein-encoding RNAs (mRNAs), as their translation leads to the production of the main cellular effectors, i.e. proteins.

The decryption of genomes has shown that the number of protein-coding genes contained in a mammalian genome, which was originally thought to be in the vicinity of 100,000, was more likely to be between 20,488 (Pennisi 2007) and 23,244 (Griffiths-Jones 2007). At first, this substantially lower gene count raised many questions regarding the apparent lack of complexity, which did not correlate with the established estimation of the number of different proteins. This discrepancy between the number of genes and the number of resulting proteins led to the proposition that a gene, defined in its simplest form as being a single locus composed of a promoter region and a downstream-transcribed sequence coding for introns and exons, can produce different proteins or different versions of a protein. So far, little effort has been invested on the investigation of splice variants, although they may represent an extremely important aspect that could explain many physiologically relevant cellular states. In this case, the number and position of microarray probes could significantly influence results and downstream interpretations.

Aside from mRNAs, the transcriptome is composed of essential RNA molecules that have functional roles without being translated. As such, the RNA subtypes involved in the maturation process of mRNAs and rRNAs, respectively, localized in the nucleus (small nuclear RNAs) and in the nucleolus (small nucleolar RNAs) are important components of the transcriptome. Additional critical RNA components required for protein production include tRNAs and rRNAs. In fact, it is well appreciated that the proportion of mRNAs in a transcriptome account for only 1–2% of total RNA content in somatic cells (Griffiths-Jones 2007).

More recently, very short RNA molecules that are involved in the control of gene expression have been identified. These very short RNA molecules have the ability to manage other RNA molecules by influencing their translation (micro-RNAs; Huttenhofer et al. 2005, Filipowicz et al. 2008, Shyu et al. 2008), by specifically targeting the decay of longer RNA molecules (small interfering RNAs, piwi interacting RNAs, and repeat associated small interfering RNAs; Choudhuri 2009), or by influencing the epigenome program to induce specific methylation of genomic sequences (still unclear but this function has been proposed for piRNAs; Choudhuri 2009). These newly identified RNA molecules definitely add another level of complexity to the transcriptome. Moreover, long noncoding RNAs (IncRNAs) are slowly emerging as an abundant RNA class that may have
important functions (Ponting et al. 2009, Wilusz et al. 2009). The field of early development has been at the center of the characterization of three of these IncRNAs whose expression is suspected to be influenced by the embryonic culture environment. The first two form a pair of IncRNAs that are involved in X-inactivation. The mechanisms by which the X-inactivation-specific transcript (Xist) and its complementary coding sequence (Tsix) are involved in dosage compensation of the extra X chromosome in mammalian females are still mostly uncharacterized (Brown & Chow 2003). However, it has been proposed that the gender bias favoring males observed in some in vitro embryo production conditions may result, at least in part, from the incapacity of the female embryo to properly inactivate the extra X chromosome, thus leading to early embryonic mortality (Wrenzycki et al. 2002, Bermejo-Alvarez et al. 2010). The third IncRNA that plays a role during embryogenesis is the H19 gene, which pairs up with the neighboring Igf2 gene to form an imprinted locus, and is believed to be abnormally methylated under diverse conditions during the use of assisted reproductive technologies (Lewis & Murrell 2004, Thurston et al. 2008). The underexpression of H19 can indicate an overexpression of Igf2, which likely contributes to the origin of the Large Offspring syndrome observed in ruminants (Thurston et al. 2008). The compiling of deep-sequencing data clearly showed a increase in the repertoire of IncRNAs (Huttenhofer et al. 2005). This is indicative of the shallowness of our true understanding of the transcriptome. Early embryo development incorporates all of these transcriptomic regulatory mechanisms that involve every RNA class of the transcriptome. So far, virtually all studies with a wide enough scope to be qualified as ‘transcriptomics’ applied to the investigation of gene expression in the early embryo have solely targeted protein-coding RNAs (without any specific focus on splice variants). Although these studies are of importance, we can conclude that only the tip of the iceberg (1–5%) has been surveyed, while the largest portion of the transcriptome still remains unaddressed.

**Unique nature of oocytes and early embryos**

It should be noted that in the context of this review, early embryo development is defined as spanning final maturation of the oocyte up to hatching of the blastocyst. As the term ‘preimplantation’ only applies to rodents and primates and that a complete embryo–uterine interaction may occur much later in development (days 30–35 in cows; Melton et al. 1951), prehatching development will rather be used to describe this developmental window. The unique nature of prehatching development significantly impairs our ability to gain all that we could from the large volume of accumulating transcriptomic data. This developmental interval includes a period of transcriptional silence that ends with the activation of the embryonic genome (EGA). The duration of transcriptional silence is species specific; the mouse genome is activated right after the fertilization and is first detected in the male pronucleus (maternal–zygotic transition (MZT); Schultz 1993) while other mammals activate their genome after two (pigs) or three cell cycles (humans, cows, and rabbits; Braude et al. 1988, Memili et al. 1998, Brunet-Simon et al. 2001, Kanka 2003). In the absence of transcription, protein synthesis is sustained by the translation of stored mRNAs accumulated during oogenesis. The complete mechanism by which these RNAs remain dormant and avoid degradation in storing particles (messenger ribonuclease-protein complexes) remains to be elucidated. However, it is well accepted that storage involves shortening of the messenger’s poly(A) tail (Bachvarova et al. 1985, Wickens 1990, Bachvarova 1992, Paynton & Bachvarova 1994). The extent of deadenylation is not known but it can be assumed to be variable. It has been reported that the average poly(A) tail length for a nonstored mRNA contains more than 200 residues, whereas stored RNAs are found with a short poly(A) tail of <50 residues (Paynton & Bachvarova 1994, Kim & Richter 2006). Conversely, the recruitment process involves the elongation of the poly(A) tail (Paynton & Bachvarova 1994). So far, maternal RNA management is believed to fall under the ‘closed-loop model’ in which the maskin and CPEB proteins are initially involved in the stabilization of the mRNA (Groisman et al. 2002). An unknown recruitment signal triggers phosphorylation events that release the mRNA, allowing the poly(A) polymerase to elongate the 3’ tail. The binding of poly(A) binding proteins (PABPs) induces the recruitment of initiation factors by joining the 3’ end to the 5’ cap structure, which in turn recruits the translation machinery (Groisman et al. 2002; Fig. 1).

Many aspects of this general model still remain to be elucidated. However, poly(A) tail length does vary greatly during oocyte maturation (Brevini et al. 2002, Lequarré et al. 2004). Furthermore, the study of total RNA content, both in mice and cows, indicates that a large proportion of the stored RNAs are lost during oocyte maturation (Lequarré et al. 2004). The purpose of this first wave of maternal RNA depletion is unknown, as de novo protein synthesis analysis during oocyte maturation suggests that mRNAs are most probably not solely destroyed following translation, but some must be degraded without first being translated (Fig. 1).

The oocyte’s legacy to embryonic development is believed to last at least until genome activation and even beyond. In the mouse, it has been proposed that the maternal RNA must first be depleted, largely through an unspecific degradation event, for MZT to occur (Schultz 2002, Schier 2007). It is not known whether this general degradation event takes place in all species or whether it involves the modification of poly(A) tail length before their decay. Following the EGA, blastomeres rapidly...
acquire a ‘somatic’ RNA management strategy in which the genome sustains basic cellular functions and also reacts and copes with the changing surrounding microenvironment. However, during the pre-EGA window, the content in the two large rRNAs is atypical of the general somatic profile (Gilbert et al. 2009b). Together, these findings clearly underscore the transient state that characterizes prehatching development, during which cell size and number fluctuate in addition to transcription potential, total RNA content, proportion of each RNA type, and adenylation state of the mRNAs. In contrast to the general process of gene expression in somatic cells, the fate of an mRNA found in a blastomere is multiple; it can be stored, used in the translation process, or simply sent for decay. These factors are important considerations for the selection of sample-handling procedures and of the type of transcriptomic platform, and ultimately for the interpretation of data to define physiological functions.

Why is it not possible to compare datasets generated by different laboratories?

The main culprit: the sample-handling steps to generate microarray hybridization samples

Although it is clear the embryo production system can significantly affect embryonic development and the resulting abundance of transcripts found in these samples, the present review will not address the discrepancies found between systems used across laboratories. It is unarguably an important aspect that could explain some of the contradictory reports. The objective of this review is rather to draw attention to the fact that even in the presence of samples produced using identical systems, the sample-handling procedures can profoundly affect the resulting datasets, to the extent of rendering the generated datasets incompatible.

It is a fact that microarray data generation requires considerable sample-handling steps, and this is especially true for mammalian oocytes and prehatching embryos as they represent a rare source of biological material, thus requiring extensive care in handling and the introduction of a global amplification step. For the establishment of our own microarray platform, we conducted many tests in order to define and control the sources of methodological variations (Gilbert et al. 2010). Every step is important, even the composition of the buffer in which the oocytes/embryos are washed and stored, as the presence of poly-vinyl alcohol negatively impacts the yield during RNA extraction (author’s personal observations, unpublished). As expected, some steps have a more significant impact on endpoint results. The following overview presents the main considerations that arise with sample processing when dealing with early embryos.

RNA extraction

Undertaking a transcriptomic study requires the extraction of the RNA content of the samples of interest. Many different commercially available options exist, which can be globally classified into three main categories: those targeting total RNA, those aim to select the mRNAs based on the presence of a poly(A) tail, and kits that are dedicated to the isolation of small RNA molecules. The scope of the present discussion is not to detail the efficiency of these different approaches but rather to draw attention to the fact that their selection will influence RNA abundance measurements and consequently the interpretation of the microarray data (Fig. 2). Subfractionation of the RNA population may impact...
downstream relative quantification and most certainly requires a different physiological interpretation than similar work based on total RNA samples.

Furthermore, the concern over the RT priming strategy will be compounded by the quality of the RNA, as sample degradation will lead to a partial or complete loss of poly(A) tails. Consequently, it is important to assess sample integrity, and currently the most efficient method is to perform electrophoretic profiles of all samples. This requires the use of instruments that can handle small sample size such as the 2100 BioAnalyzer (Agilent) or the Experion System (Bio-Rad). Careful interpretation of the profiles must be done during early development because profiles observed during the early developmental stages do not fit the ones observed in the somatic cells (Gilbert et al. 2009b). It is noteworthy to mention that the generally used u.v. absorbance measurement, which is instrumental in determining sample quantity, is not appropriate to determine its quality since it is blind to partial degradation. When using approaches subjecting whole-cell lysis to global amplification, sample integrity can only be presumed. The presence of cellular debris can reduce downstream reactions in a dose-dependent manner leading to more variable outputs and thus requiring additional considerations for controlling this aspect (author’s personal observations, unpublished).

Sample amplification

As mentioned, the scarcity of the starting material when working with mammalian oocytes and prehatching embryos dictates the requirement for global amplification (the only other alternative being to collect thousands of prehatching stage embryos). Many different types of global amplification procedures have been proposed and tested for their potential to do so without introducing a considerable bias. Many different global amplification methods are commercially available. They can be grouped in three main categories:

1) PCR-based method, which is characterized by cycling the reaction’s incubation temperature. The amplification is driven by a thermostable DNA polymerase. The reaction outputs DNA fragments from both strands of the template.

2) In vitro transcription-based method, which involves copying the templates containing a double-stranded RNA polymerase promoter sequence at one end. The reaction is isothermic and outputs single-stranded RNA molecules.

3) Isothermic amplification driven by a DNA polymerase exhibiting a strong strand displacement activity. Many versions of this type of amplification are commercially available, most of them designed for whole-genome amplification. The use of such a strategy to amplify RNA samples is becoming commercially available. The reaction is isothermic and can be randomly primed, therefore creating multiple initiation sites, even on molecules in the process of being elongated, thus producing a branched structure. Some other amplification methods target initiation of the amplification by introducing a specific priming site at one end of the RNA molecules. The cycling is thus conducted by the use of a DNA:RNA hybrid primer. Once the hybrid primer binds the target sequence, it recruits the DNA polymerase that initiates the amplification. The elongation is carried out by displacing the strand in front of the polymerase. The RNA portion of the primer is then removed by digestion with RNAsase H, liberating the priming site for another intact hybrid primer that will initiate another copy of the template that will displace the previously synthesized strand. If the reaction is randomly primed, both strands are amplified, whereas targeted priming outputs single-stranded DNA molecules.

All of these global amplification strategies reach their goal of amplifying the templates that have been
generated by the upstream steps. They have been tested for their potential to preserve the relative representative abundance of all mRNAs in the sample (Pabon et al. 2001, Puskas et al. 2002, Jenson et al. 2003, Stirewalt et al. 2004, Patel et al. 2005, van Haften et al. 2006, Peano et al. 2006, Wilhelm et al. 2006, Zhu et al. 2006). The general consensus is that, although it is expected that some representation will be lost due to different amplification efficiency between templates, the skew only affects a minority of templates. These conclusions have been made following the comparison of the abundance values of transcripts between amplified and unamplified samples. Considering the limited potential to obtain sufficient input material to conduct such tests from the prehatching samples, they all have been conducted with somatic tissues where the status of the poly(A) tail length is much less of a concern. However, when studying early development, the unique nature of the samples requires additional factors to be considered. These factors had not received enough attention before the field quickly and widely accepted these amplification strategies for routine applications aimed at studying the embryonic transcriptome. When a study requires the comparison of transcript abundance between different developmental stages, as is the case for developmental profiling, the naturally occurring differences in the RNA content must be taken into consideration. So far, this has been omitted in nearly all studies involving different developmental stages. We have recently noticed that the well accepted in vitro transcription-based global amplification method, routinely used and implemented in high-end transcriptomic platforms, leads to a uniform yield for all samples, irrespective of their intrinsic initial RNA content. For example, it is known that in cow, the eight-cell stage has the lowest total RNA content of prehatching development, whereas the germinal stage oocyte, which contains a full stock of maternal RNAs, has a significantly higher RNA content (Lequarre et al. 2004, Gilbert et al. 2009b). Therefore, samples containing equal numbers of oocytes and eight-cell embryos should, if the global amplification method were truly linear, yield cDNA/aRNA levels representative of the initial differences in RNA content. Instead, we noticed that the amplification yields equivalent outputs and therefore erases the natural differences in RNA content between these developmental stages. This loss of physiological mRNA level representation also occurs during microarray hybridization, as equal amounts of sample are deposited on the slides. Again, comparison based on identical amount of material for samples known to bear different RNA contents will lead to a loss of physiological representation of transcript abundance. In other words, how should we interpret the comparison between 1000 eight-cell embryos and 5 GV oocytes? We have recently demonstrated that failing to account for these natural differences in mRNA abundance can lead to a false-discovery rate of more than 60% (Gilbert et al. 2009a). In an unpublished study using microarrays to compare GV oocytes and eight-cell stage embryos in order to identify newly synthesized transcripts in the early stage of EGA in the cow, the validation by qRT-PCR of 40 candidate mRNAs from the generated gene list proved them to be all false positives (Dr M A Sirard, personal communication). This clearly indicates that although sound statistical methods were used to generate the gene list, if the relative mRNA level representation of a sample is lost during sample preparation, the data will most definitely not be physiologically relevant. In contrast, the comparison of samples of a similar nature, e.g. same embryo stage comparisons, is much less prone to the introduction of this type of experimental error. Overall, to avoid the introduction of profound bias, it is critical that sample amplification be controlled and that microarray data be prenormalized using spiked-in controls (Gilbert et al. 2009a, Vallee et al. 2009).

We have also conducted tests to contrast the end results of different amplification strategies in order to determine the extent of data consistency between microarray platforms. Benchmarking involved the use of a single RNA source and identical postamplification sample-handling and processing methods throughout the microarray platform. The datasets showed surprisingly low correlation values (in the range of 0.3; Gilbert et al. 2010; Fig. 3). To determine which one provided physiologically relevant outcomes, qRT-PCR was applied as the gold standard as it does not use preamplified samples. Interestingly, the lack of correlation did not arise from a loss in the relative relationship between candidates (Fig. 4) but rather from the fact that some amplification strategies favor 3’ ends while others are

Figure 3 Comparison of the impact of two-sample amplification approaches on microarray performance. Aliquots from a single RNA pool were subjected to global amplification using different commercial kits (RiboAmp, Molecular Devices, Sunnyvale, CA, USA versus RampLUP, Genisphere, Hatfield, PA, USA) following the manufacturers’ recommendations. A cDNA microarray was hybridized with labeled samples from both methods. The overall correlation was low (0.30), indicating the significant impact that sample processing can have on the resulting gene list. The presence of red and green spots indicates that the sample processing protocols generated templates that hybridize to different microarray probes.
directed toward the 5′ end, and the lack of apparent correlation between datasets arose from the position of the microarray probe relative to the 3′ end. For example, when a microarray probe is located in the coding sequence and a specific candidate bears a long 3′ UTR, some amplification methods will generate samples containing molecules that do not extend into the coding sequence (Fig. 5). Therefore, the discrepancy between datasets results in the detection or apparent absence of many transcripts (as shown in Fig. 3 by the presence of bright red and green spots). This implies that the absence of signal is not a clear indication that the transcript is absent from the sample as it is also dependent on the position of the detection probe. The difference in probe sets between microarray platforms will be discussed later.

The potential to copy sequences located at the 5′ end of a transcript depends on several factors. First, the priming strategy during the RT step greatly influences the potential to copy 5′ located sequences. By definition, the use of an oligo-dT to prime the RT reaction requires the whole transcript to be copied in order to reach the 5′ end. In addition, for oligo-dT priming, the RNA must bear a poly(A) tail of sufficient length (Fig. 2). As mentioned, this is an important consideration for the prehatching embryonic stage samples since stored mRNAs are largely deadenylated. Overall, the RT step remains fairly inefficient in providing long reads on a consistent basis. Although it is true that reverse transcriptase can read through long templates, as observed during the cloning of long cDNAs, this observation does not indicate that the enzyme generates long reads every time. Evidence of this lack of reliable efficacy of the reverse transcriptase can be observed with microarray platforms that have several probes targeting the same transcript located at various positions. Signals most often diminish when the probes are located upstream of the 3′ end.

**Sample labeling, hybridization, and microarray scanning**

Following amplification, samples must be labeled prior to being deposited onto the microarray, which will then be washed following incubation. Sample labeling can be performed using different means, such as the introduction of modified nucleotides bearing either a fluorescent dye or a chemical group (e.g. amino-allyl) that will serve as an anchor to bind the dye in a downstream reaction. In all cases where modified nucleotides are added into the polymerization mix, the yield of the reaction is negatively affected (Gilbert et al. 2010). Other strategies involve coupling of the dyes after the amplification step to prevent the kinetics of the amplification reaction from being influenced by labeling. This can be performed by introducing a specific nucleotide sequence on all templates, which will then generate a chimeric molecule.
bearing the complementary nucleic sequence and a dye-docking part (Yu et al. 2002). Another means of postamplification sample labeling is the use of the universal linkage system that targets guanine residues through the reactivity of a cisplatin group coupled with the appropriate fluorescent dye (van Gijlswijk et al. 2001). In our experience, the sample labeling strategy is not a source of variation that profoundly influences the compatibility of microarray datasets (author’s personal observations, unpublished).

Similarly, the microarray hybridization procedure itself, which either uses a simple static chamber or a state-of-the-art automated hybridization and washing station, is not an important source of variation that prevents data compatibility. In our experience, the use of an automated station increases reproducibility and therefore increases the potential for discovery by creating lower technical variance, which increases the power to find statistically significant differences. The newly developed microarray scanners that feature extended dynamic ranges (XDR) have the power to better detect extremely high and extremely low intensity signals than conventional scanners. Consequently, comparison of data produced by a scanner featuring XDR with data outputted from a conventional scanner could lead to statistically valid gene lists composed of artificial results originating from the differences in dynamic ranges between the scanners. The use of universal control standard spikes for all microarray platforms, as proposed by the Reference Consortium (Baker et al. 2005), would be of great value for defining more precise normalization algorithms that would offer full compatibility between datasets outputted from different scanners.

Data processing and analysis

Once fluorescence values have been generated from the microarray, data must be preprocessed prior to statistical analysis. The standard preprocessing steps include background correction and intra- and interarray normalizations. Anyone who has worked with microarray analysis has been boggled by the overwhelming choices that are available for any of these steps. The following is not intended to be a complete review of the possibilities but rather a brief overview of the process and some specific considerations for early embryo studies. As the present topic aims to highlight the compatibility issues with microarray data produced with early embryos, it must be stated that the ultimate comparison is better performed when data are processed using identical data processing steps. To do so, raw data from other groups can be retrieved from microarray repository such as ArrayExpress (Parkinson et al. 2009) and the NCBI Gene Expression Omnibus (Barrett et al. 2009).

In the case of background correction, some have proposed to refrain from any background correction (Yang et al. 2001, Tran et al. 2002) while others report complex strategies (Ritchie et al. 2007, Silver et al. 2009). In any case, it can be argued that background correction does not greatly influence endpoint data when the microarray hybridization resulted in low nonspecific binding. However, in the case of a messy set of hybridizations, background correction can help salvage valuable information. Intra-array normalization can also detect some of the nonspecific variations that can occur due to hybridization artifacts. The underlying logic is that when probes are randomly distributed on the surface of the microarray, hybridization signals should not be clustered in a specific area of the slide. Interarray normalization is used to correct for differences in average signal between microarrays. It is expected that replicates should be similar, therefore offering a standardization reference.

The current state of knowledge makes it extremely difficult to determine whether the premise of the standard and well-accepted normalization algorithms apply to the specific context of early embryo development. For instance, many normalizations are based on the assumption that most genes are not differentially expressed and that an equal number of candidates are over- and underexpressed (Yang et al. 2002). Using the same aforementioned example comparing the GV oocytes and eight-cell embryos, one can wonder whether this assumption holds, as it is expected that nearly all candidates will be more abundant in the oocyte, which contains the entire pool of maternal RNAs, whereas at the eight-cell stage they are mostly depleted. This question is currently unanswered. Again, the use of spiked controls may be a part of the normalization strategy that should be used when performing across-stage comparisons.

The data normalization issue when comparing naturally different samples applies to every type of dataset, including protein quantification. The investigation of mechanisms controlling prehatching development cannot be undertaken without the means to properly normalize data. One of the main benefits of working with early development samples is the potential to collect samples containing a perfectly defined number of oocytes or embryos. Such a situation is not possible when working with somatic cells, which thus rely on a specific amount of input RNA. Since RNA content varies widely between the stages (Gilbert et al. 2009b), the physiological interpretation of across-stage results generated from samples containing equal RNA content is dependent on the knowledge of how many embryos of each developmental stage were required to produce the RNA samples. Such information is never available, making the use of samples containing a defined number of oocytes or embryos more relevant. Data normalization in such a fluctuating context is thus more challenging.

It has been clearly shown by qRT-PCR that the expression of all housekeeping candidates generally used data from the somatic cells fluctuates across early development (Robert et al. 2002, Goossens et al. 2005,
Bettegowda et al. 2006, Vigneault et al. 2007), and while the proposed use of spiked-in controls may offer the opportunity to account for technical inaccuracies accumulated along sample-handling procedures, the main limitation of exogenous controls is their incapacity to account for the condition of the initial sample. For instance, if one replicate or one treatment is partially degraded, spikes will be immune to this issue. Spiked-in controls can behave perfectly and offer a way to control downstream steps but will still lead to inaccurate data when a sample is already partially degraded prior to the addition of the exogenous reference. The use of spikes for microarray data normalization has been proposed (Baker et al. 2005, Shi et al. 2006) and could represent one method of normalization for across-stage comparisons. Completely degraded samples raise a different concern as this condition can be assessed by microelectrophoresis and will directly affect the downstream global amplification. In fact, sample integrity can only be accounted for using endogenous controls that, as mentioned, fluctuate from stage to stage in embryos. To resort to the use of such endogenous transcripts, a method of data normalization across development must be determined. The use of the geometric mean from a panel of several so-called housekeeping candidates is so far the best strategy for data normalization (Goossens et al. 2005) but a few criteria must be met. First, the candidates must be chosen from independent cellular pathways to avoid the impact of co-regulation, which would skew normalization. Second, the underlying concept of a geometric mean is that although some candidates may fluctuate, it is expected that they will do so in opposite directions and that, overall, the use of several unrelated data points will prove to be more robust than one point. Again, in the context of early embryonic development, it is still possible that all seemingly unrelated candidates fluctuate in the same direction, therefore introducing a bias in the normalized dataset (Fig. 6). In fact, stable candidates across development are extremely rare (Vigneault et al. 2007). As mentioned above, before EGA, in the absence of transcription, the majority of transcripts are expected to drop as early development progresses. An opposite situation is found following EGA. As a consequence, it is highly improbable that the survey of RNA abundance across development targeting the entire RNA pool through random priming will highlight the presence of a candidate whose expression is stable enough through-out prehatching development to serve as an ideal endogenous standard for normalizing transcript abundance. In contrast, it could be possible to find a stable candidate when priming the RT reaction using an oligo-dT as the equilibrium between stored mRNAs that do not bear a poly(A) stretch long enough for the poly(T) primer to bind, and conversely, the elongation of the poly(A) tail during recruitment may lead to a constant replenishment of targetable templates.

Currently, the issue of data normalization is not settled either for qRT-PCR or microarray, especially when considering across-stage comparisons. In fact, the literature contains several algorithms that perform qRT-PCR data normalization using the geometric mean that generates three significantly different normalization datasets when tested side by side with a unique dataset (author’s personal observations, unpublished). Taken together, these observations imply that across-stage data normalization is problematic both for qRT-PCR and microarray datasets. Efforts are needed to determine benchmarked means for microarray data normalization across development, as it is an important source of variation that leads to discrepancies in gene lists or validation efficiency between studies.

**Microarray platforms**

To present some of the limitations pertaining to microarrays, it is interesting to look back at their technological development, which started a little more than a decade ago. Arrays are in fact reverse northern blots in which probes are immobilized on a solid surface and the sample is labeled. Northern blots have the benefit of assessing the integrity of the sample,
confirming the length of the transcript, and detecting the presence of splice variants. In contrast, microarrays offer the throughput to interrogate the same sample about a large number of candidates simultaneously. Microarrays can only inform on the nature of the transcript by detecting one or a few pieces of the transcript. Handmade macroarrays containing several hundred candidates yielded valuable information on RNA abundance during early development (Robert et al. 1995, Bernard et al. 2005). The scaling down of this technology using automates introduced the use of spotted glass slides containing several thousands of cDNAs. The limitation of these homemade platforms came from the need to manually assemble the cDNA collection to be printed. Even though these platforms contain significant numbers of candidates, they are tagged as being ‘boutique’ arrays containing a limited and incomplete set of candidates. Furthermore, since they are PCR products, amplicon size is generally a few hundred bp and, consequently, cDNA microarrays have been criticized for their potential lack of specificity, as gene family members could bind to the same probe (Kothapalli et al. 2002).

The recent genome decryption efforts have paved the way for the use of this genomic information to design oligos for every known protein-encoding gene. This opportunity is predominantly supported by corporate or large consortium initiatives, while boutique cDNA arrays are mainly developed using academic support. Oligonucleotide microarrays can be found in different formats that can be classified based on the number of bases of their probes, either short (the affymetrix platform offers 25-mer oligonucleotides; Albert et al. 2003) or long (70-mers; Barczak et al. 2003, Lee et al. 2007). Since it is all about characterizing relative abundance differences between treatments, every format of the microarray platform can perform adequately to provide a list of candidates following statistical analysis. However, the compatibility of the results between platforms differs widely and as mentioned, sample processing in combination with microarray format can have a considerable effect on the output gene list.

**Microarrays versus RNAseq**

Ten years ago, profiling the abundance of a dozen transcripts was a considerable task. The introduction of two main technologies changed this perspective. As mentioned, microarrays have grown in popularity ever since but have been heavily criticized and challenged right from the start by other high-throughput methods such as the serial analysis of gene expression (SAGE; Velculescu et al. 1995) and the massive parallel signature sequencing (MPSS; Brenner et al. 2000). Both of these methods are based on the presence of a poly(A) stretch on mRNAs. The SAGE method was designed to generate a single short tag (14–27 bp; Velculescu et al. 1995, Gowda et al. 2004) from every poly(A)-bearing RNA molecule, concatemerize them to produce a library of tags and systematically tally each transcript in a sample by sequencing the library. MPSS is a different approach where cDNA molecules are immobilized individually on a bead and duplicated to populate its surface. The entire batch of beads is then subjected to several rounds of an ingenious DNA sequencing strategy involving type IIS restriction enzymes, which cut the DNA at a defined distance from their binding site, and a mixture of adaptors bearing different fluorescent dyes targeting the diverse potential combination of resulting cohesive ends.

Both of these methods are often referred to as true transcriptomic methods since they do not require knowledge of the transcriptome to be performed. This aspect has become the main criticism toward microarray technologies where the method confines the user to investigate only known candidates whose probes are featured on the array and thus limits the opportunity for generating novel unexpected elements. However, although these methods are extremely powerful, the requirement for large amounts of starting material and the challenging identification of the short tags (for SAGE) or short reads (for MPSS), especially in the absence of a reference genome, prevented the wide use of these methods in the study of prehatching development.

These limitations are now being overcome with the development and availability of very high-throughput sequencing procedures, in addition to the fact that genomes of model species have been sequenced completely and several livestock as well as wild species have been sequenced with at least sufficient coverage to allow chromosome positioning and identification of genes from small DNA sequences. So far, the challenges arising from deep sequencing (or RNAseq in the case of a transcriptome (Wang et al. 2009)) are linked to the efficient handling of the data flow, in addition for requiring the same considerations for the amount of starting material and the targeted RNA population being studied as microarrays. Apart from the potential bias introduced during sample processing, postsequencing normalization can be challenging if the goal were to generate an overall perspective of relative abundance values across transcripts as well as across developmental stages. Sample fragmentation introduces a bias related to the size of the transcripts as longer ones will generate more fragments and will thus be counted more often. Strategies have been proposed to account for this size-dependent count, such as considering a defined interval from the poly(A) tail (Robinson & Oshlack 2010). So far, the issue of the fluctuating RNA composition and poly(A) length occurring during early development has not been addressed for RNAseq.

Although the throughput of these sequencing methods represents huge amounts of sequences, on the order of giga-bases, most of them still generate relatively short reads (35 bases for ABI Solid and 51–76 bases for
Illumina Solexa) that need to be sorted back onto a frame provided by a reference sequence such as Refseq or a complete genome. Deep sequencing undoubtedly provides a high-performance tool to study RNA abundance and its discovery potential is undeniable.

Through our projects, in which transcripts were obtained from a normalized library (thus some artifacts may have been introduced by the loss of closely related transcripts during the normalization process), we found that a pool of embryos encompassing all stages of the prehatching window generates transcripts coding for about 14 500 genes (author’s personal observations, unpublished). However, the study of RNA abundance and of the impact of splice variants during development may not be undertaken with the ease that is currently expected. In the case of splice variants, the extent of fragmentation required to reduce sequence length and ensure complete coverage somehow limits the power of deep sequencing to discover novel transcript variants, as these fragments must lie directly at the junction of a variant for it to be identified. Even with the longer reads provided by the 454 titanium technology from Roche, a little over 1.2 million reads of an average length of 323 bases were able to identify only 1100 transcript variants of the exon skipping type (author’s personal observations, unpublished). Considering that nearly all of these splicing events were found <5 times in the entire dataset and that the number of splice variants per gene is estimated to be 2.5 in mammals (Brett et al. 2002), these results suggest that coverage was insufficient to properly identify most splice variants. For instance, detecting 50% of splice junctions requires 17–32 million 32-nucleotide-long reads dataset, and to reach the level of detecting 92% of all splice variants in a sample, coverage must reach >500 reads per 100 nucleotides (Pan et al. 2008).

Third-generation deep-sequencing procedures will soon be available, and some are predicting the demise of microarrays. Until then, the cost of deep-sequencing runs is a limiting factor when considering numerous sample types in addition to technical replication. So far, the main bottleneck of high-throughput sequencing remains data processing, normalization, and analysis. The development of benchmarked data processing tools to properly handle the tremendous datasets is underway. As both technologies are complementary, microarrays are also still evolving and as such, whole-genome tiling arrays may provide a valuable alternative to study transcript variants. They are not available yet for livestock species, are also expensive to run, and have been criticized for their lack of sensitivity due to inherent high background signals (Wang et al. 2009). The development of the exon arrays or restricted tiling arrays solely containing probes targeting all transcripts known to be expressed in the tissue of interest determined by deep sequencing may provide an acceptable alternative.

**Interpretation of differential RNA abundance**

Although it is not clear whether the transcriptome faithfully reflects the proteome (Chen et al. 2002, Griffin et al. 2002, Imrler et al. 2008, Rogers et al. 2008), the study of transcribed elements is justified because of the greater efficacy in defining the relative abundance of all or at least a wide array of candidates. Additionally, the study of protein-encoding RNAs is supported by the assumption that in response to a treatment, signal transduction will induce shifts in transcript production or, more drastically, result in transcription inhibition or activation of previously silent sequences. This situation generally applies well to somatic cells but the unique nature of prehatching development requires additional considerations. As mentioned, the atypical RNA management needs to be accounted for in the interpretation of RNA abundance fluctuations. For example, in the somatic cells, a higher abundance value for a specific transcript would generally be interpreted as an increase in transcription, whereas during early development, the transcriptional potential (pre-EGA versus post EGA) will significantly influence its interpretation. A higher pre-EGA abundance value could indicate an unexpected early transcription event, a lower RNA depletion activity (following translation or not), or an increase in the adenylation status of the transcript if sample processing involved the use of an oligo-dT primer during the RT step. This clearly indicates that artificially higher RNA levels can be created methodologically. The study of RNA abundance values across development is increasingly prone to such erroneous conclusions, especially when a global amplification step is required, since natural RNA abundance differences between developmental stages are alleviated both during sample amplification (Fig. 6A) and during microarray hybridization. All high-throughput methods, including microarrays and RNAseq, require preamplification of the sample to generate sufficient starting material. A second source of technical bias is due to the lack of a consistent and reliable approach for data standardization. Data normalization is required to provide a comparable basis between samples and treatments but can also be a source of variation in the absence of an appropriate standard. As a consequence, overnormalization can distort RNA abundance values (Fig. 6B).

The impact of the presence of transcript variants can be an additional source of false abundance values. Surprisingly, it was recently estimated that 61% of differential expression detected by microarrays may in fact be associated with the probe detecting a shift in transcript processing, e.g. the presence of variants (Kwan et al. 2008). Such a significant false-discovery rate due to transcript maturation processes is surprising. However, some structural modifications in transcripts may be of importance during early development. It was recently shown that the length of the 3′ UTR fluctuates
according to the differential potential or stemness of the cells (Ji & Tian 2009). The true prevalence and physiological significance of these transcript modification events during early development and their impact on RNA abundance measurements are areas requiring further investigation.

It has been suggested that following gene list generation, the second degree of analysis, namely the pathway/network analysis that targets co-regulated elements, may be more robust or rather less prone to erroneous conclusions since the results are based on proportional enrichment probabilities. In other words, in the case of occasional false discoveries polluting the gene list, they will most likely not all be associated with a single pathway/network, thus enhancing the value of gene enrichment. However, this situation does predominate, and the outcomes can still be misleading when the gene list contains a very important proportion of false discoveries, as may be the case with across-stage comparisons. Some lists can harbor more than 60% false positives (Gilbert et al. 2009a). Another aspect to consider is the fact that pathway analysis relies on known knowledge and is consequently conservative in revealing novel functions. It is expected that early development involves many atypical cell functions, as exemplified by the oocyte being the only cell to re-initialize a programmed genome or the entire pre-EGA developmental window to coordinate and manage the use of large amounts of stored mRNAs. Nonetheless, integration of gene lists in a physiological context is required for the work to be of any significance. The use of high-throughput technologies leads to the generation of an overwhelmingly complex dataset and software to aid and facilitate the interpretation of such information is constantly evolving, thus becoming more powerful and useful.

Conclusion

The description of early development samples through the comparison of RNA abundance values obtained using microarrays has become a common practice. The unique characteristics of prehatching development, especially in terms of RNA management, combined with the numerous options for sample and data processing, profoundly impact the resulting gene list. These factors are too often neglected, as the introduction of methodological artifacts is common for RNA abundance reports during prehatching development. For instance, through a retrospective data compatibility assessment performed using the published RNA abundance profiles produced by qRT-PCR of 15 candidate genes (Vigneault et al. 2004, 2009, McGraw et al. 2007) and a published microarray dataset aimed to describe the bovine prehatching development (Kues et al. 2008), we observed important discrepancies in the relative transcript abundance patterns among immature oocyte, eight-cell embryo, and blastocyst for 80% of the candidates (12/15). Without the use of a way to account for the natural RNA abundance differences between the developmental stages during sample amplification, hybridization, and data normalization, it is challenging to determine whether methods could have introduced a bias that deeply influenced the downstream conclusions.

In order to minimize the methodological differences and provide a strong basis for comparison, complete evaluation of the impacts of each source of methodological variations needs to be performed. Until then, compatibility between platforms is questionable. On a more positive note, the true gain of knowledge from RNA abundance studies resides in downstream interpretation. This objective may be somewhat buffered from the discrepancies that arise between platforms as the main consideration lies within relative abundance values or more precisely the extent and the nature of the change, rather than absolute transcript copy numbers.

The true description of early embryo development, so far mostly limited to stage-specific datasets, will most certainly need to consider a multifaceted strategy that will integrate results from all of the different ‘omics’ and the characterization of the program controlling prehatching development. This is the mandate of the newly created field of research called system biology, which is still in its infancy. With increasingly compatible datasets, it will be possible to perform meta-analyses as a first step in this objective of global modeling/perspectives. Until we can pursue this type of comprehensive analysis, our present efforts are equivalent to trying to appreciate a beautiful painting by looking at it through a pinhole. The current issue is not necessarily to concern ourselves with how many transcripts are missed but rather with how to integrate the overwhelming amount of information being produced into a clear and comprehensible picture that truly reflects the developmental and physiological functions that arise in the early embryo during prehatching development.

Declaration of interest

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This review was supported by the Natural Sciences and Engineering Research Council of Canada (grant numbers 300712-04 and 340825-06); the Fond québécois de la recherche sur la nature et les technologies (grant number 121560).
Acknowledgements
The author acknowledges the work of Dr Andrew J Watson (University of Western Ontario, Canada) for his help in editing the final manuscript in addition to Dr Julie Nieminen (Université Laval, Canada) and Dr Marc-André Laniel for linguistic revision.

References


Coudhuri S 2009 Lesser known relatives of miRNA. Biochemical and Biophysical Research Communications 388 177–180. (doi:10.1016/j.bbrc.2009.08.039)


www.reproduction-online.org

Pittfalls of gene expression in early embryos 799


Wickens M 1990 In the beginning is the end: regulation of poly(A) addition and removal during early development. Trends in Biochemical Sciences 15 320–324. (doi:10.1016/0968-0004(90)90022-4)


Received 19 April 2010
First decision 28 May 2010
Revised manuscript received 19 July 2010
Accepted 10 September 2010