THE ROLE OF TRANSCRIPTOMICS: PHYSIOLOGICAL EQUIVALENCE BASED ON GENE EXPRESSION PROFILES

Shiori Miura¹, Takehiro Himaki¹, Junko Takahashi², and Hitoshi Iwahashi¹

¹ The United Graduate School of Agricultural Science, Gifu University. 1-1 Yanagido, Gifu-shi, Gifu 501-1193, Japan
² Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

ABSTRACT

DNA microarray and RNA sequencing platforms enable us to simultaneously measure the expression levels of a large number of genes. These platforms serve as emerging techniques for the comprehensive understanding of gene expression and profiling and are generally used to detect genes that are differentially expressed in response to several conditions including the environment, stress, and genetic background. Therefore, these can function as tools to understand biological responses. Large-scale gene expression studies are also useful in biological experiments to confirm the assumptions of models and data uniformity. Accordingly, these approaches are imperative for ensuring the accuracy of investigations of the in vivo vital reaction. In this review, we describe the role of transcriptomics in the evaluation of physiological equality through gene expression profiles.

Keywords: gene expression, microarray, RNA-seq, transcriptome


1. RNA and transcriptomics

The "Central dogma of molecular biology", as defined by Francis Crick, explains the flow of genetic information within a biological system (Crick, 1970). This concept states that genetic information encoded in DNA is transcribed to RNA, and RNA is translated to protein. The biological activities of cells, tissues, and organisms are based on the central dogma of molecular biology. RNA is an important factor required to mediate gene and protein expression. There are two types of RNA, non-coding RNA (ncRNA) and messenger RNA (mRNA or protein-coding RNA). As indicated in Table 1, ncRNAs play several key roles in gene regulation including transcriptional and post-transcriptional regulation, regulation of alternative splicing, control of transcription factor binding, chromatin modification, and protein-coding RNA stabilization (Louro et al., 2009; Pertea, 2012). ncRNAs include ribosomal RNA (rRNA) and transfer RNA (tRNA) and are classified into two broad groups by size. Long ncRNAs (lncRNA) are greater than 200 nucleotides and small ncRNAs (sncRNA) are 200 nucleotides or less. lncRNAs include large intergenic ncRNAs, long intronic ncRNAs, antisense RNAs, and pseudogene RNAs. lncRNAs play critical and specialized roles in numerous biological processes including the regulation of gene expression, and pre- and post-transcriptional modulation of epigenetic regulation (Rinn and Chang, 2012; Guo et al., 2015). sncRNAs also have several functions: microRNAs (miRNAs) and small interfering RNAs (siRNAs) modulate post-transcriptional gene expression by binding to specific mRNAs; small nucleolar RNAs chemically modify tRNAs and other RNAs; and piwi-interacting RNAs inhibit transposon function through the PIWI protein, and maintain genomic constancy in germline cells. Other ncRNAs include small nuclear RNA, transcription initiation RNA, X-inactivation RNA, and promoter-associated RNA. Dysfunction of ncRNA is associated with complex diseases such as cancer, and neurological, developmental, and cardiovascular diseases (Taft et al., 2010; Esteller, 2011). More than 90% of the genome is transcribed into RNA, and it is estimated that mRNA constitutes...
approximately 62% of the transcripts (Pertea, 2012). mRNA plays a key role in transcription and reflects the information of almost all expressed genes. The complete set of mRNA, or primary transcripts, under a specific condition or in a specific cell is defined as the transcriptome. Transcriptomics is the study of the transcriptome and enables researchers to elucidate gene expression dynamics under different circumstances. This leads to a deeper understanding of the function of each gene in the genome, and the regulation of gene expression, and aids in the elucidation of molecular mechanisms of abnormal states such as diseases.

In this review, we describe the general application of transcriptomics and recommend approaches for analyzing whole genome expression profiling data. Fig. 1 shows a schematic diagram of recommended application of transcriptomics.

### Table 1: List of major non-coding RNAs and their main functions

<table>
<thead>
<tr>
<th>Function category</th>
<th>Type</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein synthesis</td>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td></td>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>transcriptional modification, replication of DNA</td>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td></td>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>regulation of expression</td>
<td>piRNA</td>
<td>PIWI-interacting RNA</td>
</tr>
<tr>
<td></td>
<td>miRNA</td>
<td>micro RNA</td>
</tr>
<tr>
<td></td>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td></td>
<td>xiRNA</td>
<td>X-inactivation RNA</td>
</tr>
<tr>
<td></td>
<td>tiRNA</td>
<td>transcription initiation RNA</td>
</tr>
<tr>
<td></td>
<td>lincRNA</td>
<td>large intergenic RNA</td>
</tr>
<tr>
<td></td>
<td>pRNA</td>
<td>promoter-associated RNA</td>
</tr>
</tbody>
</table>

2. Methods of transcriptomics

The transcriptome of a living organism is complex. Over the past decade, several approaches have been developed to elucidate its intricacy and recent dramatic advances in analytical technologies have allowed researchers to further appreciate the transcriptome. In the 1990s, expressed sequence tag (EST) sequencing was employed to rapidly identify expressed genes and gene fragments (Schuler et al., 1996). Although EST sequencing is a high-throughput technique, it is expensive. Tag-based methods including serial analysis of gene expression, cap analysis of gene expression, and massively parallel signature sequencing were developed, but were unable to discriminate between genetic isoforms and were very expensive to apply on a large scale. Microarray was developed for genome-wide analysis, and has become the most widely used approach for transcriptomics. Recently, RNA sequencing (RNA-seq) using next generation sequencing technology has allowed the transcriptome to be characterized, and the number of studies using RNA-seq have gradually increased (Ghosh and Qin, 2010; Yu and Lin, 2016). Microarray and RNA-seq have become the main tools used in transcriptome research. These tools allow researchers to simultaneously analyze the expression of a large number of genes and to focus on physiological equivalence. Several studies have compared the accuracy of microarray and RNA-seq measurements (Marioni et al., 2008; Fu et al., 2009; Su et al., 2011; Zhang et al., 2015). It was found that the biological interpretation of detected genes was mostly consistent between the data of the two technologies, and that RNA-seq provides better detection sensitivity than microarray. To validate the accuracy of these techniques, it is necessary to quantitatively evaluate the expression levels of genes detected as having altered expression by DNA Microarray or RNA-seq. Therefore, Real-Time Reverse-Transcriptase PCR is performed after global analysis in most studies.

2.1 DNA microarray

DNA microarray analysis was established in 1995 as a new technology to analyze gene expression (Schena et al., 1995). Since then, microarrays are being widely used across biological
disciplines. The number of published papers using this technology continues to increase, as does the number of commercial suppliers of microarrays, associated reagents, and analysis hardware and software (Kawasaki, 2006; Yauk and Berndt, 2007).

The microarray technique is based on the basic principle of "DNA Hybridization" and uses DNA chips consisting of...
many oligonucleotides (probes) spotted onto a glass slide. Probes are deposited onto glass slides via the direct (in situ) synthesis of oligonucleotide probes onto the chip surface using photolithographic methods (Gao et al., 2004) or by deposition methods, which include contact-spotting using pins and deposition by ink jet (Hughes et al., 2001; Sethi et al., 2008). Presently, these arrays comprise mostly 40- to 70-mer oligonucleotides spotted on a glass slide. This technique makes it possible to semi-quantitatively measure the expression levels of large numbers (1,000-40,000) of genes simultaneously. In conjunction with computational analysis tools, microarray analysis enables the identification of genes that vary in expression in different biological contexts (Schena et al., 1995; Quackenbush, 2001).

The microarray method consists of several processes. Total RNA is extracted from the sample and reverse-transcribed into complementary DNA (cDNA). cDNA is labeled with fluorescent dyes, hybridized to the DNA chip, and scanned to produce microarray image data. The intensity of fluorescence, which reflects the degree of hybridization and transcript copy number, is digitized by appropriate software. However, microarray has several limitations including requiring information about genome sequence to produce the oligonucleotides, and inaccuracy of data owing to high background from nonspecific cross-hybridization.

2.2 RNA-seq

RNA-seq is a new approach involving next-generation sequencing and allows a quantitative analysis of all expressed genome regions (Wang et al., 2009). The RNA-seq method consists of several steps: extraction of total RNA, reverse transcription of RNA to cDNA, construction of a cDNA fragment library, sequencing using a high throughput sequencing platform, generation of single-end or paired-end reads 30-400 base pairs in length, and sequence alignment (Wang et al., 2009; Griffith et al., 2015). Recently, a method allowing direct single molecule RNA sequencing, without prior conversion of RNA to cDNA, was reported (Ozsolak et al., 2009).

RNA-seq captures RNA directly to build a sequence, allowing for the detection of new transcription products, fused sequences, and single nucleotide polymorphisms of unknown genes without gene-specific biases. Additional advantages of RNA-seq include low background noise, large and dynamic signal range, and detection with no requirement for prior sequence information. More recently, RNA-seq has emerged as the preferred approach for genome-wide expression analysis (Wang et al., 2009; Rowley et al., 2011; Su et al., 2014).

3. Transcriptomics for physiological difference

DNA microarray and RNA-seq technology provide a wide range of novel application opportunities relating to gene expression profiles, which can be applied to various studies. The availability of transcriptomic technology has provided new opportunities for researchers to characterize global gene expression profiles. Recent advances in transcriptomics have allowed us to identify specific genes, gene families, and pathways associated with biological responses. The mechanisms regulating biological reactions, as well as the identification of genes implicated in these responses, are of great interest to the research community. These techniques could serve to assign functions to previously unannotated genes and to allocate gene groups to functional pathways (DeRisi et al., 1996; Han et al., 2015). Additionally, these techniques contribute to our understanding of biological mechanisms and responses to environmental stimuli (Miller and Tang, 2009). The identification of differentially expressed genes is helpful to show the biological distinction and physiological difference between two different sets of conditions.

3.1 Evaluation for genetic background

Organisms sustain biological activity based on information contained within the genome. The characteristics of the individual are distinctively determined by genetic information, the expression of which defines the phenotype (sex, aging, tissues, individuals, and species). Detection of genes differentially expressed between these phenotypes helps to characterize the sample and allows us to evaluate individual subjects.

Differences in gene expression resulting from changes in intrinsic conditions such as sex (Balakrishnan et al., 2012; Blekhman et al., 2010; Conforto and Waxman, 2012; Jansen et al., 2014; Caetano-Anolles et al., 2015; Sakashita et al., 2015) and aging (Brink et al., 2009; Wilson et al., 2010; Takahashi et al., 2011; Naumova et al., 2012; Steegenga et al., 2014; Roux et al., 2015; Wei et al., 2015) can be identified using microarray and RNA-seq techniques. Liu et al. (2013) reported that aging results in sexually dimorphic changes in the skeletal muscle transcriptome, and they detected differential expression of genes related to oxidative phosphorylation, immune function, and muscle protein catabolism. Data show that gene expression dynamics related to aging vary according to sex, and suggest that older women tend to be more predisposed to loss of muscle function with aging. Many studies show that normal tissues have their own gene expression profiles and have identified organ specific gene sets that are highly expressed in a tissue selective manner in mouse (Su et al., 2002), rat (Walker et al., 2004), dog (Briggs et al., 2011), pig (Hornshøj et al., 2007), and human (Hisao et al., 2001; Shmueli et al., 2003; Son et al., 2005; Kilpinen et al., 2008). Transcriptomics can also be used to compare gene expression data across species (Chan et al., 2009; Merkin et al., 2012; Sudmant et al., 2015). These studies provide new insights into the molecular basis of tissue and organismal diversity. Transcriptomics can be used to identify genes that contribute to this diversity, and can be
utilized to build a biological gene database.

3.2 Evaluation for environmental effects

Recent progress in transcriptomics enables us to identify genes and pathways associated with responses to exogenous abiotic stresses (Ma et al., 2012; Jogaiah et al., 2013; Deshmukh et al., 2014; Evans, 2015). Many studies have used DNA microarrays to infer how organisms respond to different environments (Gracey and Cossins, 2003; Cossins et al., 2006; Gracey, 2007; Evans and Hofmann, 2012), such as temperature (Murata et al., 2006; Yang et al., 2010; Long et al., 2012; Aguado-Urda et al., 2013; Logan and Buckley, 2015), osmolality (Posas et al., 2000; Evans and Somero, 2008; Melamed et al., 2008; Halbeisen and Gerber, 2009), oxygen (Ton et al., 2003; Garniczarska, 2006; Swiderek et al., 2008; Otsuka et al., 2010; Gracey et al., 2011; Shinde et al., 2015) and pH (Leaphart et al., 2006; Serrano et al., 2006; Worden et al., 2009; Evans et al., 2013). Detection of genes with expression changes in response to environmental change helps to predict the fragility, resistance, and adaptability of an organism, tissue, or cell in the environment. Additionally, genes with constant expression in the presence of environmental change can be identified and are potentially important for overall survival. Transcriptomics using RNA-seq in fishes have reported many new genes that participate in metabolic functions, protein folding and degradation, developmental processes, oxygen transport, and protein synthesis (Liu et al., 2013). These studies have also identified heat shock protein genes that are differentially expressed following alterations in temperature (Liu et al., 2013; Smith et al., 2013). Coble et al. (2014) identified differential expression of genes related to decrease internal temperatures, reduced hyperthermia-induced apoptosis, and promotion of tissue repair occurring in the liver of heat-exposed broiler chickens. In addition, they also found that the expression of genes involved in the regulation of perturbed cellular calcium changes following heat exposure.

Moreover, an adaptive response to various exogenous environmental stresses, including osmotic pressure and starvation, was validated at a genome level using RNA-seq (Xia et al., 2013; Johnson et al., 2015). RNA-seq is being applied to study the stress response similar to how microarray is being used assess the physiological state.

Transcriptomics has been especially useful in the field of experimental embryology, where it has been used to evaluate the in vitro and in vivo environments. Transcriptomic data of embryonic cells produced in vitro and developed in vivo have been compared in mouse (Ren et al., 2015), pig (Östrup et al., 2013; Whitworth et al., 2015), bovine (Driver et al., 2012; Degrelle, 2015), and sheep (Wei et al., 2016). These studies have identified crucial discordances between the in vitro and in vivo expression of several genes and gene pathways. Individual genes, and pathways, function in complex biological processes. Minor changes in the expression of several genes may perturb a pathway and possibly have drastic biological effects (Han et al., 2015). Transcriptomics helps researchers to identify differences in embryo gene expression in vivo and in vitro. These analyses may assist in improving culture conditions so that in vitro analyses can more accurately represent in vivo physiological conditions.

3.3 Evaluation for toxicological damage

Evaluation of chemical stress or toxicogenomics is critically important for transcriptome analysis. Afshari et al. (1999) and Nuwaysir et al. (1999) demonstrated the efficacy of microarray as a tool for assessing chemical and environmental toxicity in a bioassay. Our group has applied microarrays to examine the molecular response of a yeast model using various toxic materials. As summarized in Table 2, we have detected differentially expressed genes and determined the mechanism against the toxic matter for each of terpinene (Parveen et al., 2004), dimethyl sulfoxide (Murata et al., 2003), mycotoxin citrinin (Iwahashi et al., 2007), thiorium (Mizukami-Murata et al., 2006), cadmium (Momose and Iwashashi, 2001), and thiuram (Kitagawa et al., 2002).

The application of microarray to toxicogenomics is not limited to yeast cells. Various different models including fungus (Zhao et al., 2015), plants (Oono et al., 2014; He et al., 2015; Xu et al., 2015; Oono et al., 2016), mollusk (Meng et al., 2013), fish (Bougas et al., 2013), mouse (Hu et al., 2014), and human cells (Cartularo et al., 2016) are being used to assess cadmium stress. Huang et al. (2012) reported that perfluorooctane sulfonate affects the expression of genes related to neurobehavioral defects, mitochondrial dysfunction, and the metabolism of proteins and fats. Identification of differentially expressed genes helps to reveal statistical significance (e.g. fold change and significance test) and/or biological significance (Tseng et al., 2012) and to clarify the mechanisms regulating adaptive responses. RNA-seq can also be applied to study toxicogenomics and has been used to evaluate changes in miRNA expression in response to multiple environmental factors including arsenic (Zhang et al., 2016), cigarette smoke (Bean et al., 2011; Hackett et al., 2012), the carcinogen benzo[a]pyrene (Van Delft et al., 2012) and gamma-irradiation (Moskalev et al., 2014).

Humans are exposed to many substances, including drugs, additives, and toxic chemicals, that can have a direct or indirect influence on our body functions. Risk assessment for these materials using transcriptomics is a means to evaluate the degree of toxicity, or risk, that such substances pose to an organism.

3.4 Evaluation for diseases

Comparison of genome-wide expression patterns among patient samples presents measurable information and helps to identify genes
Using microarrays, biomarkers have been identified using microarray. RNA-seq is emerging as a very powerful tool to identify biomarkers of cancer (Wood et al., 2007; Berger et al., 2010; Pfueger et al., 2011; Cancer Genome Atlas Research Network, 2013; Fumagalli et al., 2014; Kosti et al., 2016), Alzheimer’s disease (Satoh et al., 2014), tuberculosis (Zhang et al., 2014), and cirrhosis (Tan et al., 2014). RNA-seq may be a more suitable platform for the search of precise biomarkers than traditional omics approaches, including microarray or proteomics because of its ability to detect novel genes/exons, RNA editing, fusion transcripts, and allele-specific expression. Still, RNA-seq does have limitations and issues resulting from several biases including experimental/technical procedures, downstream computational analyses, and informatics infrastructures (Costa et al., 2013).

Genome-wide analysis has boosted the biomarker diagnostics industry and contributes to disease subtype classification, disease diagnosis and prognosis, selection of therapeutic treatments, and disease prevention (He et al., 2006; Sun et al., 2013; Su et al., 2014; Aibar et al., 2015).

4. Transcriptomics for physiological equality

Most transcriptome analyses have been aimed at detecting genes with altered expression levels. Many researchers analyze expression patterns to identify a characteristic expression pattern following exposure to change, as detailed in the previous section. Although transcriptome analysis has been a powerful tool for biological and biomedical studies, it remains to be determined whether these technologies can be applied with high accuracy and precision.

---

Table 2: Evaluation of toxicity by yeast DNA microarray

<table>
<thead>
<tr>
<th>Toxic material</th>
<th>Function of differentially gene</th>
<th>Main mechanisms of action of toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antifungal drug</td>
<td>Up-regulate gene</td>
<td>Synthesis, transcription</td>
</tr>
<tr>
<td>Lipid and fatty acid metabolism, cell wall structure and organization, detoxification, cellular transport</td>
<td></td>
<td>Cell wall structures</td>
</tr>
<tr>
<td>Dimethyl sulfoxide</td>
<td>Phospholipid biosynthesis, the methionine synthesis</td>
<td>Energy (mitochondrial respiration-related genes), ion homeostasis, protein synthesis</td>
</tr>
<tr>
<td>Phospholipid biosynthesis, cell rescue, defense and virulence, energy</td>
<td></td>
<td>Synthesis and maintenance of the cell wall</td>
</tr>
<tr>
<td>Mycotoxin</td>
<td>Metabolism, cell rescue, defense and virulence, energy</td>
<td>Glutathione synthesis, cell rescue, defense and virulence</td>
</tr>
<tr>
<td>Cytokinin</td>
<td></td>
<td>Oxidative stress</td>
</tr>
<tr>
<td>Thorium</td>
<td>Oxidative stress, glycosogen and trehalose metabolism, cell wall damage</td>
<td></td>
</tr>
<tr>
<td>Thiuram</td>
<td>Cellular transport, carbohydrate metabolism, protein destination, transport facilitation, detoxification, stress response</td>
<td>Transcription, cell growth, protein synthesis, carbohydrate metabolism, cellular communication, DNA, lipid etc. metabolism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Membrane factors, transport activity, protein denaturing and degradation, oxidative stress, DNA damage</td>
</tr>
</tbody>
</table>

That would be reasonable targets for therapeutic intervention (Afshari et al., 1999; Chin and Kong, 2002; Dudda-Subramanya et al., 2003; Suci and Omidi, 2011). The invention of technologies for transcriptomics, using genome-wide analysis and computational approaches, has made it possible to identify the prognostic significance of individual gene expression changes from thousands of markers. For example, in cancer studies, this technology is employed to obtain comprehensive gene expression profiles in both normal tissues (Saito-Hisaminato et al., 2002) and cancer tissues including those from hepatocellular carcinomas, pancreatic cancers, and esophageal squamous cell carcinomas (Okabe et al., 2001; Han et al., 2002; Macgregor, 2003; Nakamura et al., 2004; Yamabuki et al., 2006; D’Angelo et al., 2014; Zhu and Tsao, 2014; Nishimura et al., 2015). Using microarrays, biomarkers for Parkinson’s disease (Alonso-Navarro et al., 2014; Sun et al., 2014) and myocardial infarction (Devaux et al., 2010; Głogowska-Ligus and Dąbek, 2012) have been identified. Additionally, DNA microarrays are used to study complex diseases, in which hundreds of genes are often implicated, such as allergies, diabetes, and obesity (Rome et al., 2009; Wang et al., 2009; Rodriguez-Acebes et al., 2010; Liu et al., 2013; Lu and Liao, 2015). Some reports indicate that these diseases rely on multiple gene interactions, rather than changes in a single causal gene, and that many different mechanisms and pathways are linked together (Benson and Breitling, 2006). While more illness biomarkers have been identified using microarray, RNA-seq is emerging as a very powerful tool to identify biomarkers of cancer (Wood et al., 2007; Berger et al., 2010; Pfueger et al., 2011; Cancer Genome Atlas Research Network, 2013; Fumagalli et al., 2014; Kosti et al., 2016), Alzheimer’s disease (Satoh et al., 2014), tuberculosis (Zhang et al., 2014), and cirrhosis (Tan et al., 2014). RNA-seq may be a more suitable platform for the search of precise biomarkers than traditional omics approaches, including microarray or proteomics because of its ability to detect novel genes/exons, RNA editing, fusion transcripts, and allele-specific expression. Still, RNA-seq does have limitations and issues resulting from several biases including experimental/technical procedures, downstream computational analyses, and informatics infrastructures (Costa et al., 2013).

Genome-wide analysis has boosted the biomarker diagnostics industry and contributes to disease subtype classification, disease diagnosis and prognosis, selection of therapeutic treatments, and disease prevention (He et al., 2006; Sun et al., 2013; Su et al., 2014; Aibar et al., 2015).
Proof of invariability or and constancy of gene expression profiles provides internal evidence of biological stability. Transcriptome analysis enables us to check the technical/biological uniformity using genome-wide screening for gene expression.

4.1 Capacity

Gene expression profiles reflect biological capacity and vary according to the situation at the time. For example, expression profiles change during the process of differentiation and generation (Mansergh et al., 2009; Goggolidou et al., 2013; Iruretagoyena et al., 2014; Shiraki et al., 2014; Alonso-Martin et al., 2016). DNA microarray is tool used to evaluate cell properties by comparing the expression profiles of all genes. In some stem cell research reports, the degree of differentiation and development is evaluated by examining the similarity of gene expression patterns. Global gene-expression patterns were compared between human induced pluripotent stem (iPS) cells and human embryonic stem (ES) cells using oligonucleotide DNA microarrays (Takahashi et al., 2007). A high correlation of global gene-expression patterns was found between iPS cells and ES cells, suggesting that established iPS cells are similar to ES cells. Hrvatin et al. (2014) reported that differentiated human stem cells are analogous to fetal β cells rather than adult β cells. Mishra et al. (2008) demonstrated that human bone marrow-derived mesenchymal stem cells exposed to tumor-conditioned medium over a prolonged period assumed a carcinoma associated fibroblast-like myofibroblastic phenotype. Handel et al. (2016) generated transcriptome data to compare iPS cell derived neurons to human fetal and adult brain and indicated that iPS cell-derived cortical neurons closely resembled primary fetal brain cells. Tanaka et al. (2013) confirmed that the human iPS cell derived myogenic differentiation cells were similar to those of perfectly differentiated human myoblast cells and quite divergent from those of undifferentiated iPS cells. Therefore, transcriptionomics provides evidence for the establishment of cell-specific identities.

Moreover, Datson et al. (2007) reported the comparison of gene expression profiles between tissues in the same individual. This study reported that a high correlation coefficient was obtained when comparing gene expression in marmoset neuronal tissues (hippocampus and cortex) indicating a high degree of similarity in expression profiles. Additionally, comparison of hippocampal gene expression with that of all peripheral tissues resulted in a severe drop in the correlation coefficient. Thus, transcriptome analysis is useful to demonstrate biological similarity between cells or tissues, and for the determination of genetic characteristics.

4.2 Reproducibility

Biological experiments need to be reproduced multiple times under the same experimental conditions. To demonstrate evidence of reproducibility is important for data to be comparable (Chen et al., 2007; Darbani and Stewart, 2014). The Micro Array Quality Control (MAQC) project was established to construct quality control and standardization tools using four titration samples which are measured on seven microarray platforms and three alternative gene expression technologies. The MAQC project had already proven the reproducibility of microarray data by the quantitative signal values and the qualitative detection calls (MAQC Consortium et al., 2006). To validate and extend these observations, numerous researchers have independently validated microarray data (Yang et al., 2002; Burgoon et al., 2005; Guo et al., 2006; Hoen et al., 2008).

In addition to the technical reproducibility of microarray results, biological and physiological reproducibility are also important. Iwahashi et al. (2009) and Takahashi et al. (2012) demonstrated the importance of reproducibility of expression profiles among individuals under the same experimental conditions. This reproducibility proved the stability of an experimental protocol that affected the biology and physiology. Therefore, the reproducibility of gene expression patterns observed under the same experimental conditions suggests that the experimental and analysis methods used are stable and robust. However, MAQC does not provide conclusions related to inter-platform compatibility. There are differences in the fluorescent intensities measured by different platforms, and even within each platform site-by-site variability exists (Chen et al., 2007).

Reproducibility of RNA-seq has been demonstrated by multi-group reports (Tang et al., 2015), across laboratories (Hoen et al., 2013), and among technical replicates (Marioni et al., 2008; Mortazavi et al., 2008; Anders and Huber, 2010; Bullard et al., 2010; Ozsolak et al., 2010; Roberts et al., 2011). Danielsson et al. (2015) compared RNA-seq data sets of human brain, heart, and kidney samples from different laboratories and studies and concluded that RNA-seq expression measurements show global consistency after log transformation and elimination of batch effects. To establish the reproducibility and comparability of RNA-seq, the RNA Sequencing Quality Control (SEQC) project was constituted and coordinated by the Food and Drug Administration. The role of the SEQC is to assess the performance of RNA-seq across laboratories and to dissect different sequencing platforms and data analysis pipelines (SEQC/MAQC-III Consortium, 2014).

4.3 Stability

It is essential that there is a high correlation and reproducibility within and between replicated experiments (within established standards) for data to be considered reliable and robust (Yauk and Berndt, 2007). In 2001, the Functional Genomics Data Society (http://fged.org/) described the Minimum Information About a Microarray Experiment (http://www.mged.org/Workgroups/MIAME/miame).
html) to establish a standard for recording and reporting microarray-based gene expression data (Brazma et al., 2001). They provide six factors for standardization: experimental design, array design, sample, hybridization, measurements, and normalization. However, this report does not indicate the most suitable standardized analytical methods, which may result in the variance of data.

To overcome the dispersion of microarray data, standardization is carried out in various experimental models such as yeasts (Mizukami et al., 2004; Taymaz-Nikerel et al., 2016), mice (Williams et al., 2004), rats (Guo et al., 2006), and non-human primates (Ebeling et al., 2011). Microarray is employed to prove the stability of both experimental conditions and experimental subjects. Iwahashi et al. (2009) used genomics to report on the physiology of medaka, which are used as a model animal for toxicity testing. Medaka mRNA expression was measured in individuals maintained within, as well as beyond, the Organization for Economic Cooperation and Development (OECD) guidelines for the fish acute toxicity test. They found that the toxic environment specified within the OECD guidelines did not affect the expression profiles of medaka and indicated that extraordinary conditions, beyond the guidelines, decreased reproducibility of data.

Takahashi et al. (2011) conducted microarray analysis to evaluate variations in whole blood gene expression patterns in different individual miniature pigs at different ages. The number of expressed genes and variation in gene expression intensity within miniature pigs of the same age were observed to converge with aging, and gene expression became uniform after 20 weeks of age. This report reveals the age at which genetic uniformity of the large animal model was reached. It is a basic concept of biological experiments that all conditions, except the variable being measured, must be the same. Demonstration of biological standardization and uniformity of genetic background using transcriptomics provides great value for laboratory animals in biological experiments. Such analyses allow for efficient and accurate experimental results and contribute to the standardization of breeding and rearing methods.

4.4 Margin of safety

Few studies have applied transcriptomics to the issue of food safety. We have introduced this new application of transcriptomics (Miura et al., 2017). We have used transcriptomics to demonstrate the safety of a diet consisting of a by-product of Japanese liquor production (shochu distilled water: SDW) for use as pig feed. We evaluated the expression profiles of pigs fed with SDW, hyperlipid diet, and feed containing toxicant. We observed a high correlation between the gene expression profiles of the control and SDW feeding groups. Furthermore, the expression profiles of these two groups were different from those of the hyperlipidemia and toxicant model groups. These data indicated that feeding with SDW did not have a physiological effect on the pigs and assessed such feeding as safe. Therefore, microarray can be used to test foods and demonstrate proof of similarity with the normal state, making it a valuable approach for evaluating safety.

In the medical field, preclinical and clinical drug safety studies are a key prerequisite of the drug approval process. Non-human primates and pigs are important models for such studies. Ebeling et al. (2011) reported the genetic similarity of the non-human primate Macaca fascicularis to human. However, from the viewpoint of animal welfare, usage of these species in animal experiments has declined, and the opportunity of animal experimentation is limited. Active utilization of transcriptomics, which can obtain a large volume of information simultaneously, improves our understanding of the in vivo pharmacokinetics of model organisms and provides a significant contribution to the global “3R” animal welfare initiative: reduce, refine, and replace animal experiments.

5. Conclusion

Transcriptomics has allowed us to simultaneously identify gene expression dynamics and differential gene expression. Transcriptomics is useful to identify illness biomarkers as well as biological responses to various stimulations and stresses, and plays a key role in advancing genomic and molecular biology research.

However, using this approach, we may overlook potentially important functions of genes that are not induced by the particular condition being examined. Not only should we concentrate on detecting specific differentially expressed genes, but we should also examine the entire expression profile. In addition, standardization of experimental conditions is essential and an absolute requirement for the legitimacy of the experiment. Uniformity of experimental conditions is brought about by adjustment of the technical, platform, and biological biases.

In conclusion, transcriptome analysis can be used to validate the standardization of an experiment by eliminating biological biases. As well as using transcriptomics to identify change or variability, genomic researchers can take advantage of these approaches as evidence of invariance, constancy, and reproducibility of their system of interest.

Acknowledgments

We acknowledge the help of Shinji Tokuyama, Associate Professor of Shizuoka University, for critical comments and advice, as well as Ms. Yolani Syaputri and Editage (www.editage.jp) for English language assistance.

References

’t Hoen PAC, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen GJB and den Dunnen JT (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab...


Gracey AY, Lee TH, Higashi RM and Fan T (2011) Hypoxia-induced mobilization of stored triglycerides in the euryoxic...


Leaphart AB, Thompson DK, Huang K, Alm E, Wan XF, Arkin...


Nakamura T, Furukawa Y, Nakagawa H, Tsunoda T, Ohigashi H, Murata K, Ishikawa O, Ohgaki K, Kashimura N, Miyamoto M,
Sakashita A, Kawabata Y, Jincho Y, Tajima S, Kumamoto S,


