

## Review Article

# Transcriptome profiling and genomic atlas study for endocrine disrupting chemicals in whole body system of animals

Yejee Park, Min-Jae Jang, and Jun-Mo Kim\*

<sup>1</sup>Department of Animal Science and Technology, Chung-Ang University, Anseong, Gyeonggi-do 17546, Korea

\*Corresponding author: Jun-Mo Kim, Department of Animal Science and Technology, Chung-Ang University, Anseong, Gyeonggi-do, 17546, Korea. Tel: +82-31-670-3263, Fax: 82-31-675-3108, Email: junmokim@cau.ac.kr

## ABSTRACT

Understanding the biological phenomena encoded through the genome is important. Therefore, it is possible to identify the genes of each organism and infer the characteristics and similarities of each other through the genomic analysis. RNA-seq is a representative method among methods for comparing gene expression levels through transcriptome analysis. It is accompanied by next generation sequencing (NGS) and enables highly quantitative and wide-range precise measurement. A representative research method that includes whole genome contents is to construct a biological map. Using the atlas, it is possible to conduct research that includes all comprehensive genetic information to identify specific locations where genes are expressed through gene location mapping. In addition, by identifying the correlation between the gene and the biomarker in the network, select a significant biomarker and functional analysis could be performed. Endocrine disruptors cause diseases by disrupting endocrine function in the body. Bisphenol A, a representative endocrine disruptor, is most permeated in daily life, threatening human health, obesity and spermatogenesis. This review focuses on transcriptome profiling and genomic atlas construction that can provide comprehensive biological insights in animal genetics studies and information of endocrine disrupting chemicals.

**Keywords:** Atlas, Bisphenol A, Endocrine disrupting chemicals, Network analysis, Transcriptome profiling

## Transcriptome profiling

Information representing biological phenomena is encoded in the genome. Therefore, it is possible to identify the genes of each living organism through genome analysis and infer their characteristics and similarities to each other. Not all genes in the genome are always expressed, but living organisms respond to the external environment by expressing their genes according to specific circumstances. Transcriptome analysis is used for determining how an organism responds by comparing and confirming the expression level of a gene (Wolf, 2013).

### 1. Representative transcriptome profiling methods

The most popular genome-wide transcriptome profiling methods are to use microarray and RNA-sequencing (RNA-seq) (Table 1). Microarray is based on the hybridization of target strands with fixed complementary probe strands which are consisted of already known sequences. RNA derived from characteristic cells is treated with reverse transcriptase to make complementary DNA (cDNA), and then a fluorescent substance is added thereon to examine the level of gene expression (Ekins & Chu, 1999). Microarray has the advantage of confirming the result through a relatively simple reaction, and chips for specific organisms are commercially available. Since the microarray technique has been developed and commercialized a long time ago, its standards are well established and reliability is also high. On the other hand, the RNA-seq method is based

on next-generation sequencing (NGS) that reads all RNA sequences. It converts cDNA from each isolated transcript and then sequencing them with a massively parallel deep sequencing-based approach. The resulting short sequencing reads can be mapped to a reference genome to quantify the expression level of a gene relative to a condition of interest or absolute level (Wang, Gerstein, & Snyder, 2009). In the past, the microarray technique played an important role in whole transcriptome analysis. However, nowadays, RNA-seq has been preferred because it effectively overcomes the limitations of microarray. RNA-seq technology is not dependent on the prior knowledge of the reference transcriptome, unlike the microarray. Also, when compared to microarray, RNA-seq data includes a higher dynamic range of expression levels, lower background signals, and a relatively small amount of total RNA for quantification (Van Vliet, 2010).

**Table 1.** Comparison of expression microarray with RNA-sequencing technology.

	Microarray	RNA-seq
Principle	A hybridization-based technique Detect the presence of specific RNA sequences within a sample	A sequencing-based technique Detects the RNA sequences in a sample
Advantages	Well-defined protocols / analysis pipelines Standardized approaches for data submission Relatively low cost	Can identify new RNA sequences including SNP, noncoding RNA High dynamic range (no saturation) Direct sequence alignment, no hybridization
Disadvantages	Analysis only for pre-defined sequence Dynamic range limited by scanner High variance for low expressed genes	Requires relatively high-power computing facilities High set-up costs in limited places Complex analysis of splice variants
Biases	By sequence differences and errors during amplification	By fragmentation and sequencing technology used

## 2. Other transcriptome profiling methods

Serial analysis of gene expression (SAGE) is a sequence-based approach transcriptome technology that can be identified and quantitatively compared. It generates a short specific tag of a population of messenger RNA (mRNA) in a sample of interest based on generating a representative SAGE tags library. SAGE sequencing allows the high-throughput process of the frequency of libraries correlating with the relative amounts of their mRNAs (Velculescu, Zhang, Vogelstein, & Kinzler, 1995). In addition, massively parallel signature sequencing (MPSS) is the sequencing tool that is available to conduct in-depth expression profiling. It is an open-ended platform that counts the number of individual mRNA molecules produced by each gene and can be analyzed at the expression level. The datasets of MPSS are in a simplified digital format and no prior requirements are needed to identify and characterize the genes (Rani & Sharma, 2017; Reinartz et al., 2002).

## Atlas

Over the decades, scientists have strived to provide a comprehensive overview of research findings in the field of genomics. As a result, a growing number of overviews for different types of research have been published, and one of the most effective research methods that contain the entire content was constructing a biological atlas. The purpose of a biological atlas is to assist users by providing additional information and analyses of maps. In particular, the structure of the human body, including the location of bones and muscles, and nerves, is the most often shown in anatomy (Netter, 2014). There are also many atlases containing comprehensive genetic information, and it is possible to check the contents and use the database online.

## 1. Types of atlas

A brief introduction to representative atlas databases made for easy use by people: (1) The human protein atlas which is composed of 6 detailed atlas (tissue atlas, single cell type atlas, pathology atlas, brain atlas, blood atlas, and cell atlas) (Pontén, Jirström, Uhlen, & Ireland, 2008). The aim of this atlas is to map all human proteins by integrating system biology techniques such as antibody-based imaging and mass spectrometry-based proteomics. (2) The cancer genome atlas (TCGA) contains comprehensive information including molecular characteristics of the major cancer genome of over 20,000 primary cancer and matched normal samples spanning 33 cancer types (Weinstein et al., 2013). (3) The Allen brain atlas is a genome-wide atlas that integrates gene expression information and anatomical data generated in the mouse brain. In addition, the brain atlas maps and identifies genes expressed in all tissues of the brain in a three-dimensional space (Sunkin et al., 2012).

## 2. Methodologies of genomic atlas

There are several ways to display genomic information, and some of the main methods used in the transcriptome atlas are as follows (Table 2). Firstly, analyses such as multi-dimensional scaling (MDS), t-distributed stochastic neighbor embedding (t-SNE), and principal component analysis (PCA) are performed to check the similarity and distribution pattern of the data (Abdi & Williams, 2010; Hout, Papesh, & Goldinger, 2013; Van der Maaten & Hinton, 2008). Similarly, hierarchical clustering analysis is used to categorize samples with similar characteristics. There are various types of cluster agglomeration methods such as maximum or complete linkage clustering, minimum or single linkage clustering, mean or average linkage clustering, centroid linkage clustering, and Ward's minimum variance method, and the linkage method is selected differently among Euclidean distance, Manhattan distance, Pearson correlation distance, Spearman correlation distance, and Kendall correlation distance according to the characteristics of the data (Murtagh, Contreras, & Discovery, 2012). Through genetic location mapping, specific locations where genes are expressed can be identified by mapping genes or genetic variants to the expressed tissue or further to the expressed chromosome (Altshuler, Daly, & Lander, 2008; Mahfouz, Huisman, Lelieveldt, Reinders, & Function, 2017). In addition, this approach allows for annotating the reference genome to identify already known genomes and their biological function using a genome browser and other functional annotation databases. Furthermore, gene co-expression network (GCN) analysis is performed to find the correlation between genes and significant biomarkers among the genes inside the network. Representative examples of this analysis are using the partial correlation and information theory (PCIT) algorithm (Watson-Haigh, Kadarmideen, & Reverter, 2010) and the weighted gene co-expression network analysis (WGCNA) (Langfelder & Horvath, 2008). The selected biomarker can be subjected to subsequent analysis, such as showing changes in expression level or analyzing their function compared to other species (Fang et al., 2020).

**Table 2.** Summary of the major atlas construction approaches.

Approach	Analysis	Objective	Type	Method (visualization / program)	Reference
Element-based approach	Correlation	Simple, direct, and intuitive	Pearson's, Spearman's, Fisher's	Differentially expressed gene (DEG), Tissue specific gene (TSG)	(Costa-Silva, Domingues, & Lopes, 2017; Fang et al., 2020) (Murtagh et al., 2012)
	Clustering	Defined, and representative	Hierarchical, Non-hierarchical	K-means, Random forest,	
	Multivariate	Dynamic relation, and distribution	Distribution analysis, Regression	Multiple experiment viewer (MeV) Principal component analysis (PCA), Multi-dimensional scaling (MDS), t-distributed stochastic neighbor embedding (t-SNE)	(Anowar, Sadaoui, & Selim, 2021)
Pathway-based approach	Pathway analysis	Function identification	Kyoto encyclopedia of genes and genomes (KEGG), Gene ontology (GO)	Pathview, Gene set enrichment analysis (GSEA), Treemap	(Werner, 2008)
	Co-expression analysis	Assess the relationship between each gene	Gene co-expression network, Weighted correlation	Partial correlation and information theory (PCIT) algorithm, Weighted correlation network analysis (WGCNA)	(Langfelder & Horvath, 2008; Watson-Haigh et al., 2010)

# Endocrine disrupting chemicals (EDCs)

## 1. Types of EDC

Endocrine disrupting chemicals (EDCs), which are also known as environmental hormones, are chemicals that cause various diseases by disrupting the normal function of the endocrine system in the body. It is largely divided into naturally occurring EDC and synthetic EDC (Kabir, Rahman, Rahman, & pharmacology, 2015). Heavy metals such as lead, cadmium, and mercury are known as naturally occurring EDC. The first known synthetic EDC was diethylstilbestrol (DES), a powerful synthetic female hormone. It was administered to millions of pregnant women in the United States from 1948 to 1972, with the hope that it would be effective in preventing miscarriage, while the efficacy and safety of the drug had not been confirmed in clinical trials. Unfortunately, miscarriages increased, and the fetus would have a disorder of the reproductive system (Langston, 2008). Phytoestrogens, which exist in nature as endocrine disruptors, have been observed to have estrogen activity in more than 43 edible plants such as bean, apple, cherry, strawberry, wheat, corn, and cotton fruit. It is known that the concentration of these hormones is known to be only about a few thousandths of the natural estrogen secreted by the body, so in general, these foods do not act as endocrine disruptors (Reinli, Block, & cancer, 1996). The category of endocrine disruptors that has been the most problematic in recent years is environmental pollutants. Since they are both endocrine disruptors and environmental pollutants, they are sometimes referred to as environmental endocrine disruptors. Currently, more than 100 species are known. Typical examples are bisphenol analogs, dioxin, polychlorinated biphenyls (PCBs), dichloro-diphenyl-trichloroethane (DDT), organic chlorine pesticides, heavy metals, and plasticizers (Diamanti-Kandarakis et al., 2009).

## 2. Bisphenol A

Bisphenol A (BPA), an endocrine disrupting chemical that permeates daily life from receipts to plastic bags, is threatening human health (Rochester, 2013). In the case of infants and children, due to behavioral characteristics such as sucking fingers or sitting on the floor, the concentration of endocrine disruptors accumulated in the body is high, so caution is required. In particular, it was confirmed that the concentration of BPA in the urine of infants and children was twice that of adults. A typical problem with BPA exposure in children and adolescents is precocious puberty. In addition, BPA is a substance capable of binding to androgen receptors and inhibits spermatogenesis by antagonizing androgens at high concentrations (Radwan et al., 2018). BPA is an endocrine disruptor that has been proven to be the most correlated with obesity, and according to the US National Health and Nutrition Examination Survey, higher BPA concentration in both adults and children are associated with a higher risk of obesity and abdominal obesity (Carwile & Michels, 2011). An association between BPA and cardiovascular disease was reported in the 2008 US National Health and Nutrition Examination Survey data analysis (Lang et al., 2008). Similarly, a 10-year follow-up in the UK also showed that an increase in BPA concentration of 4.56 ng/ml increased the risk of coronary artery disease by 1.13 times, suggesting a correlation between BPA and cardiovascular disease (Melzer et al., 2012).

## Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1A6A1A03025159).

## References

- Abdi, H., & Williams, L. J. J. W. i. r. c. s. (2010). Principal component analysis. *2*(4), 433-459. Altshuler, D., Daly, M. J., & Lander, E. S. J. s. (2008). Genetic mapping in human disease. *322*(5903), 881-888.
- Anowar, F., Sadaoui, S., & Selim, B. J. C. S. R. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). *40*, 100378.
- Carwile, J. L., & Michels, K. B. J. E. r. (2011). Urinary bisphenol A and obesity: NHANES 2003–2006. *111*(6), 825-830.
- Costa-Silva, J., Domingues, D., & Lopes, F. M. J. P. o. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *12*(12), e0190152.
- Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L. C., Hauser, R., Prins, G. S., Soto, A. M., . . . Gore, A. C. J. E. r. (2009). Endocrine-disrupting chemicals: an Endocrine Society scientific statement. *30*(4), 293-342.
- Ekins, R., & Chu, F. W. J. T. i. b. (1999). Microarrays: their origins and applications. *17*(6), 217-218.
- Fang, L., Cai, W., Liu, S., Canela-Xandri, O., Gao, Y., Jiang, J., . . . Rosen, B. D. J. G. r. (2020). Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *30*(5), 790-801.
- Hout, M. C., Papesh, M. H., & Goldinger, S. D. J. W. I. R. C. S. (2013). Multidimensional scaling. *4*(1), 93-103.
- Kabir, E. R., Rahman, M. S., Rahman, I. J. E. t., & pharmacology. (2015). A review on endocrine disruptors and their possible impacts on human health. *40*(1), 241-258.
- Lang, I. A., Galloway, T. S., Scarlett, A., Henley, W. E., Depledge, M., Wallace, R. B., & Melzer, D. J. J. (2008). Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. *300*(11), 1303-1310.
- Langfelder, P., & Horvath, S. J. B. b. (2008). WGCNA: an R package for weighted correlation network analysis. *9*(1), 1-13.
- Langston, N. J. E. H. (2008). The retreat from precaution: Regulating diethylstilbestrol (DES), endocrine disruptors, and environmental health. *13*(1), 41-65.
- Mahfouz, A., Huisman, S. M., Lelieveldt, B. P., Reinders, M. J. J. B. S., & Function. (2017). Brain transcriptome atlases: a computational perspective. *222*(4), 1557-1580.
- Melzer, D., Osborne, N. J., Henley, W. E., Cipelli, R., Young, A., Money, C., . . . Wareham, N. J. J. C. (2012). Urinary bisphenol A concentration and risk of future coronary artery disease in apparently healthy men and women. *125*(12), 1482-1490.
- Murtagh, F., Contreras, P. J. W. I. R. D. M., & Discovery, K. (2012). Algorithms for hierarchical clustering: an overview. *2*(1), 86-97.
- Netter, F. H. (2014). *Atlas of human anatomy, Professional Edition E-Book: including NetterReference. com Access with full downloadable image Bank*: Elsevier health sciences. Pontén, F., Jirstrom, K., Uhlen, M. J. T. J. o. P. A. J. o. t. P. S. o. G. B., & Ireland. (2008). The Human Protein Atlas—a tool for pathology. *216*(4), 387-393.
- Radwan, M., Wielgomas, B., Dziejirska, E., Radwan, P., Ka ł użny, P., Klimowska, A., . . . Jurewicz, J. J. A. j. o. m. s. h. (2018). Urinary bisphenol A levels and male fertility. *12*(6), 2144-2151.
- Rani, B., & Sharma, V. J. A. R. (2017). Transcriptome profiling: methods and applications-A review. *38*(4), 271-281.
- Reinartz, J., Bruyns, E., Lin, J.-Z., Burcham, T., Brenner, S., Bowen, B., . . . Woychik, R. J. B. i. F. G. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *1*(1), 95-104.
- Reinli, K., Block, G. J. N., & cancer. (1996). Phytoestrogen content of foods—a compendium of literature values. *26*(2), 123-148.
- Rochester, J. R. J. R. t. (2013). Bisphenol A and human health: a review of the literature. *42*, 132-155.
- Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., . . . Dang, C. J. N. a. r. (2012). Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *41*(D1), D996-D1008.
- Van der Maaten, L., & Hinton, G. J. J. o. m. l. r. (2008). Visualizing data using t-SNE. *9*(11). Van Vliet, A. H. J. F. m. l. (2010). Next generation sequencing of microbial transcriptomes: challenges and opportunities. *302*(1), 1-7.
- Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. J. S. (1995). Serial analysis of gene expression. *270*(5235), 484-487.
- Wang, Z., Gerstein, M., & Snyder, M. J. N. r. g. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *10*(1), 57-63.
- Watson-Haigh, N. S., Kadarmideen, H. N., & Reverter, A. J. B. (2010). PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *26*(3), 411-413.
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., . . . Stuart, J. M. J. N. g. (2013). The cancer genome atlas pan-cancer analysis project. *45*(10), 1113-1120.
- Werner, T. J. C. o. i. b. (2008). Bioinformatics applications for pathway analysis of microarray data. *19*(1), 50-54.
- Wolf, J. B. J. M. e. r. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA - seq tutorial. *13*(4), 559-572.