What Are Omics Sciences?

1

Mario Vailati-Riboni, Valentino Palombo, and Juan J. Loor

Abstract

The word omics refers to a field of study in biological sciences that ends with *-omics*, such as genomics, transcriptomics, proteomics, or metabolomics. The ending -ome is used to address the objects of study of such fields, such as the genome, proteome, transcriptome, or metabolome, respectively. More specifically genomics is the science that studies the structure, function, evolution, and mapping of genomes and aims at characterization and quantification of genes, which direct the production of proteins with the assistance of enzymes and messenger molecules. Transcriptome is the set of all messenger RNA molecules in one cell, tissue, or organism. It includes the amount or concentration of each RNA molecule in addition to the molecular identities. The term proteome refers to the sum of all the proteins in a cell, tissue, or organism. Proteomics is the science that studies those proteins as related to their biochemical properties and functional roles, and how their quantities, modifications, and structures change during growth and in response to internal and external stimuli. The metabolome represents the collection of all metabolites in a biological cell, tissue, organ, or organism, which are the end products of cellular processes. Metabolomics is the science that studies all chemical processes involving metabolites. More specifically, metabolomics is the study of chemical fingerprints that specific cellular processes establish during their activity; it is the study of all small-molecule metabolite profiles. Overall, the objective of omics sciences is to identify, characterize, and quantify all biological molecules that are involved in the structure, function, and dynamics of a cell, tissue, or organism.

M. Vailati-Riboni, Ph.D. • V. Palombo, Ph.D. • J.J. Loor, Ph.D. (✉)
Division of Nutritional Sciences, Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: vailati2@illinois.edu; valentino.palombo@gmail.com; jloor@illinois.edu

B.N. Ametaj (ed.), *Periparturient Diseases of Dairy Cows*,
DOI 10.1007/978-3-319-43033-1_1

1.1 Introduction

The trademark characteristic of omics technologies is their holistic capability in the context of the cell, tissue, or organism. They are aimed primarily at the universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics) in a specific biologic sample in a non-targeted and non-biased manner. The basic aspect of these approaches is that a complex system can be understood more thoroughly if considered as a whole. The omics approach is suitable for hypothesis-generating experiments, as holistic approaches acquire and analyze all available data to define a hypothesis, which can be further tested, in situations when no hypothesis is known or prescribed due to lack of data. When applied to well-studied scenarios, omics are still applicable to test and prove the connections and interrelationships among the many faces of a complex physiologic state, and to discover missing pieces in the current knowledge.

The first omics technologies were the automated DNA sequencer and the ink-jet DNA synthesizer developed in the early 1990s by Leroy Hood and colleagues as a tool for global gene expression analysis (e.g., transcriptomics) (Hood 2002). Around the same time, Hood's group also introduced the protein sequencer and the protein synthesizer to study protein expression at the cellular level, a process known as "proteomics." Furthermore, the concomitant emergence of metabolomics studies started by Frank Baganz and his group (Oliver et al. 1998) completed the physiologic flow of biological information processing and synthesis, from gene expression, to protein synthesis, and metabolites changes.

1.2 Genomics

Genomics pertains to the study of the complete set of DNA in an organism, including all of its genes, i.e., the "genome." With the advent of next-generation sequencing (NGS) technology the acquisition of genome-scale data has never been easier, expanding our ability to analyze and understand whole genomes and decreasing the existing gap between genotype and phenotype. Genetics and genomics sound alike but they have specific distinctions. Genetics is the study of heredity, or how the characteristics of living organisms are transmitted from one generation to the next via DNA. It involves studies focusing on specific and limited numbers of genes, or part of genes with known function, to understand how these influence particular traits of interest. At present, high-throughput technology and advances in computational biology have changed this paradigm enabling the study of organisms in terms of genome structure, addressing biological questions at a genome-wide scale, i.e., genetics is being progressively "contaminated" with genomics.

The advent of genomics turned genome-wide association studies (GWAS) into the gold standard method to identify candidate regions associated with complex traits of interest (quantitative trait loci—QTL), both in humans and other species (Gondro et al. 2013). Probe-based chips developed by various commercial companies and with a large number of single nucleotide polymorphism (SNP) markers

spread across the genome are currently being used to help uncover associations between genes and traits of interest. Species-specific arrays encompassing 10,000 up to 800,000 SNP are currently available. Such coverage ensures that any QTL will be closely linked with at least one marker. For this reason, GWAS became powerful enough to map causal genes with modest effects, i.e., disease-related quantitative traits. With the large number of genes studied simultaneously, genomic studies, in fact, can overcome the limitations of traditional genetic association approaches, enhancing our understanding of peripartal diseases (Loor 2010).

The interpretation of GWAS results still represents an important challenge. For instance, if a robust association between a phenotype and a list of genes is uncovered, one can have more confidence about the possibility for discovery of novel candidate genes. Despite the power of GWAS for discovery, studies to confirm the role of genes associated with the trait of interest should be performed to confirm functional relationships. To address this issue, gene-based software offers an effective solution in post-GWAS analysis (Capomaccio et al. 2015). Several SNP array data management tools have been developed in recent years and among these PLINK (Purcell et al. 2007), due to its speed and stability, is standard for data management. Currently, the entire GWAS pipeline can be easily executable by *ad hoc* computer programs, which, in the majority of cases, are open-source multiplatform software packages often developed in the R environment (Nicolazzi et al. 2015).

In the context of animal breeding for a given trait or traits, "genomic selection" deserves special mention (Meuwissen et al. 2001). This approach is a form of "marker-assisted selection" in which a large number of genetic markers, covering the whole genome, are used to estimate animal breeding values (EBV), i.e., the genetic value of young animals based on their genotype. This can lower the generation interval and increase the rate of genetic progress in different animal populations, traditionally based on progeny testing (Goddard and Hayes 2007). The continued progress in DNA sequencing efficiency in the near future will allow for sequencing complete genomes of individual animals, hence allowing the selection of animals with favorable QTL's alleles. Clearly, we are at the beginning of an era where individual genome sequencing will allow not only the study of domestication and selection of breeds, but also the understanding of quantitative differences associated with environmental factors, all of which will help guide experimental design for more effective animal disease control (Bai et al. 2012).

1.3 Transcriptomics

The transcriptome is the total RNA (i.e., mRNA, noncoding RNA, rRNA, and tRNA) expressed by a cell or tissue, thus representing a snapshot of cellular metabolism. The transcriptome era started when Schena et al. (1995) developed the "microarray" technology using the ink-jet DNA synthesizer, allowing for the analysis of a predetermined set (from hundreds to thousands) of cellular mRNA on a large scale. However, the recent introduction of high-throughput next-generation DNA sequencing (NGS) technology has revolutionized transcriptomics by allowing RNA

analysis through cDNA sequencing on a massive scale (RNAseq) (Voelkerding et al. 2009). This technology eliminated several challenges posed by microarray technologies, including the limited dynamic range of detection, while providing further knowledge of the qualitative, and not only quantitative, aspects of transcriptome: (1) transcription initiation sites, (2) sense and antisense transcripts, (3) alternative splicing events, and (4) gene fusion.

As it also provides detailed information on the noncoding RNA portion of the total RNA, RNAseq has enabled the understanding of complex regulatory mechanisms (e.g., epigenetics). Since the early twenty-first century, among the various epigenetic mechanisms, microRNA (miRNA), a class of small noncoding RNA (18-25 nucleotides), have received the greatest notoriety. Such attention is well founded because miRNA play a major role in controlling posttranscriptional regulation by preventing translation of mRNA (Romao et al. 2011). Furthermore, miRNA are not only part of the epigenetic machinery, but also are involved in its regulation, underscoring their pivotal role as epigenetic mediators (Poddar et al. 2017). In addition, through RNAseq or miRNA-designed microarrays, the miRNome (the total mRNA expressed by a cell at a given time) also can be analyzed.

1.4 Proteomics

The term "proteome" was defined as the characterization and quantification of all sets of proteins in a cell, organ, or organism at a specific time and was coined by Wasinger et al. (1995). Thus, a proteomic analysis provides the protein inventory of a cell or tissue at a defined time point, facilitating discovery of novel biomarkers, identification and localization of post-translational modifications, and study of protein–protein interactions (Chandramouli and Qian 2009). Powerful techniques have been established to identify and differentially quantify protein species of complex biological samples, and proteomic is being adopted by livestock researchers (Lippolis and Reinhardt 2008; Sauerwein et al. 2014).

The core of modern proteomics is mass spectrometry (MS) (Aebersold and Mann 2003), a technique in which all chemical compounds in a sample are ionized and the resulting charged molecules (ions) are analyzed according to their mass-to-charge (m/z) ratios. For a simple pre-separation of complex protein mixtures before MS analysis one- or two-dimensional polyacrylamide gel electrophoresis (1D-PAGE, 2D-PAGE) is often used. But to further enhance automation in the process and create a streamed pipeline analyses, different types of liquid chromatography (LC or HPLC) are used to complement or substitute gel-based separation techniques.

Identification of the proteins among treatments or conditions is performed by comparison against a database of proteins "digested *in silico*," meaning that the raw data are directly compared with theoretically generated data from protein databases. Reliable quantification of the identified protein also is possible with several MS-based quantification methods including chemical, metabolic, enzymatic labeling, and label-free (May et al. 2011). Proteomic advances made absolute

quantification of proteins possible through the AQUA (Absolute quantification of proteins), QConCat (artificial proteins comprised of concatenated peptides), and protein standard for absolute quantification (PSAQ) approaches (Rivers et al. 2007; Brun et al. 2007).

1.5 Metabolomics

The metabolome consists of the global profiling of metabolites in a biological sample. A metabolomics analysis may be conducted on a variety of biological fluids and tissue types and may utilize a number of different technology platforms. Metabolomics typically uses high-resolution analysis together with statistical tools such as principal component analysis (PCA) and partial least squares (PLS) to derive an integrated picture of the metabolome (Zhang et al. 2012). As one of the most-common spectroscopic analytical techniques, nuclear magnetic resonance (NMR) can uniquely identify and simultaneously quantify a wide range of organic compounds in the micro-molar range, providing unbiased information about metabolite profiles. The wide spectrum of molecules detectable by this approach includes peptides, amino acids, nucleic acids, carbohydrates, organic acids, vitamins, polyphenols, alkaloids, and inorganic species. Application of MS is gaining increased interest in high-throughput metabolomics, often coupled with other techniques such as chromatography (GE-MS, LC-MS, UPLS-MS) or electrophoretic techniques (CE-MS). Due to its high sensitivity and wide range of covered metabolites, MS has become the technique of choice in many metabolomics studies (Zhang et al. 2012).

1.6 Perspectives

Omics technologies have contributed widely to the understanding of the delicate physiologic equilibrium that allows for a successful transition into lactation (Vailati-Riboni et al. 2016). Their application to the study of peripartal disease pathophysiology is spreading across research groups worldwide. Despite this, a reductionist approach focusing on parts and sections of the physiology (rather than considering it as a whole) is still the main approach used by scientists when handling this holistic output. We are still considering single organs as the "system" to study, subsequently inferring the connection with the rest of the organism based on the existing literature. The physiologic and metabolic complexity of these diseases unavoidably requires a systems biology approach, i.e., a way to systematically study the complex interactions in the cow using a method of integration instead of reduction. Only in this way researchers will be able to uncover the underlying links (pathways, regulatory networks, and structural organization) within and between tissues (e.g., adipose and liver; skeletal muscle and adipose; gut microorganisms and epithelia), and to detect new emergent properties that may arise from examining the interactions between all components of a system.

The systems approach in its purest connotation has not yet been applied to the field of dairy science. This is largely due to the fact that when integrating multiple datasets, one tends to generate bare numerical relationships rather than meaningful biological connections among organs. Therefore, as a future frontier, the dairy science community must address the need for "useful" approaches (e.g., modeling, bioinformatics) to integrate knowledge derived from multiple "omics" analyses within and between tissues, focusing both on the classical flow of genetic information (transcriptome, proteome, metabolome), and what lies above it (epigenetic).

References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422(6928):198–207. doi:10.1038/nature01511

Bai YS, Sartor M, Cavalcoli J (2012) Current status and future perspectives for sequencing livestock genomes. J Anim Sci Biotechnol 3:8. doi:10.1186/2049-1891-3-8

Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, Vandenesch F, Garin J (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. Mol Cell Proteomics 6(12):2139–2149. doi:10.1074/mcp.M700163-MCP200

Capomaccio S, Milanesi M, Bomba L, Vajana E, Ajmone-Marsan P (2015) MUGBAS: a species free gene-based programme suite for post-GWAS analysis. Bioinformatics 31(14):2380–2381. doi:10.1093/bioinformatics/btv144

Chandramouli K, Qian PY (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics 2009. doi:10.4061/2009/239204

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124(6):323–330. doi:10.1111/j.1439-0388.2007.00702.x

Gondro C, JVD W, Hayes B (2013) Genome-wide association studies and genomic prediction, Methods of molecular biology, vol 1019. Humana Press, New York

Hood L (2002) A personal view of molecular technology and how it has changed biology. J Proteome Res 1(5):399–409. doi:10.1021/pr020299f

Lippolis JD, Reinhardt TA (2008) Centennial paper: Proteomics in animal science. J Anim Sci 86(9):2430–2441. doi:10.2527/jas.2008-0921

Loor JJ (2010) Genomics of metabolic adaptations in the peripartal cow. Animal 4(7):1110–1139. doi:10.1017/S1751731110000960

May C, Brosseron F, Chartowski P, Schumbrutzki C, Schoenebeck B, Marcus K (2011) Instruments and methods in proteomics. Methods Mol Biol 696:3–26. doi:10.1007/978-1-60761-987-1

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Nicolazzi EL, Biffani S, Biscarini F, Orozco Ter Wengel P, Caprera A, Nazzicari N, Stella A (2015) Software solutions for the livestock genomics SNP array revolution. Anim Genet 46(4):343–353. doi:10.1111/age.12295

Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. Trends Biotechnol 16(9):373–378

Poddar S, Kesharwani D, Datta M (2017) Interplay between the miRNome and the epigenetic machinery: implications in health and disease. J Cell Physiol. doi:10.1002/jcp.25819

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575. doi:10.1086/519795

Rivers J, Simpson DM, Robertson DH, Gaskell SJ, Beynon RJ (2007) Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. Mol Cell Proteomics 6(8):1416–1427. doi:10.1074/mcp.M600456-MCP200

Romao JM, Jin W, Dodson MV, Hausman GJ, Moore SS, Guan LL (2011) MicroRNA regulation in mammalian adipogenesis. Exp Biol Med 236(9):997–1004. doi:10.1258/ebm.2011.011101

Sauerwein H, Bendixen E, Restelli L, Ceciliani F (2014) The adipose tissue in farm animals: a proteomic approach. Curr Protein Pept Sci 15(2):146–155

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270(5235):467–470

Vailati-Riboni M, Elolimy A, Loor JJ (2016) Nutritional systems biology to elucidate adaptations in lactation physiology of dairy cows. In: Kadarmideen HN (ed) Systems biology in animal production and health, vol 2. Springer International Publishing, Cham, pp 97–125. doi:10.1007/978-3-319-43332-5

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55(4):641–658. doi:10.1373/clinchem.2008.112789

Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16(7):1090–1094

Zhang A, Sun H, Wang P, Han Y, Wang X (2012) Modern analytical techniques in metabolomics analysis. Analyst 137(2):293–300. doi:10.1039/c1an15605e