Status and prospects of systems biology in grapevine research

José Tomás Matus, Valentino Ruggieri, Francisco José Romero, Marco Moretto, Darren Chern Jan Wong

J.T. Matus (✉)
Institute for Integrative Systems Biology, I²SysBio (Universitat de València - CSIC), 46908 Paterna, Valencia, Spain.
e-mail: tomas.matus@uv.es

V. Ruggieri
Center for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Barcelona, Spain.
IRTA (Institut de Recerca i Tecnologia Agroalimentàries). Barcelona, Spain.
e-mail: valentino.ruggieri@irta.cat

F. J. Romero
Plant Development Unit, Instituto de Bioquímica Vegetal y Fotosíntesis, IBVF (Universidad de Sevilla - CSIC)
Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de Sevilla.
Sevilla, Spain.
e-mail: fran@us.es

M. Moretto
Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach, 38010 San Michele all'Adige, Italy.
e-mail: marco.moretto@fmach.it

D.C.J. Wong
Ecology and Evolution, Research School of Biology, Australian National University, Acton, ACT 2601, Australia.
e-mail: wongdcj@gmail.com

Abstract
The cultivated grapevine, *Vitis vinifera* L., has gathered a vast amount of omics data throughout the last two decades, driving the imperative use of computational resources for its analysis and integration. Molecular systems biology arises from this need allowing to model and predict the emergence of phenotypes or responses in biological systems. Beyond single omics networks, integrative approaches associate the molecular components of an organism and combine them into higher order networks to model dynamic behaviors. Application of network-based methods in multi-omics data is providing additional resources to address important questions regarding grapevine fruit quality and composition. Here we review the recent history of systems biology in this species. We highlight the most relevant aspects of the discipline and describe important integrative studies that have helped in the global understanding of how this species responds to the environment and how it triggers the fruit ripening developmental program. We also highlight the latest resources that are available for the grapevine community to exploit and take advantage of all the omics data that its being generated.

Keywords: Integrative analysis, Regulatory network, Gene co-expression, Multi-omics.

8.1 Introduction

Genes and their products perform complex cellular tasks that are essential for all living organisms. At the molecular level, they are organized as modules forming part of large networks. Within these high-order associations, genes/proteins that are functionally related interact, regulate each other, or form part of a metabolic pathway. The functional characterization of these molecules through forward and reverse genetic analyses has allowed the dissection of their networks and their involvement in diverse cellular processes. In the last decade, however, a massively promoted approach to asset the whole comprehension of a network from a global perspective has been the integration of several types of omics data.

The rise of next generation sequencing (NGS) technologies has led to an expansion in the amount of genomic/transcriptomic data required to be stored and processed. In addition, technologies covering proteomics and other types of omics are rapidly increasing the amount of data being produced. Scientists are now racing to develop efficient data analysis algorithms, user-friendly tools and software applications, and establishing extensive hardware infrastructure for answering different questions of modern life science. It is hypothesized that the larger the amount of omics data being generated for a species the easier for its integration, engendering more robust and reliable analyses.

The grapevine (*Vitis vinifera* L.) has become an appealing species to define as a 'model' system for studying non-climacteric fleshy fruits. The increasing amount of genomics data being continuously generated within the grapevine community, after the grape genome was sequenced and released in 2007, has certainly helped in this nomination. The grape genome, currently on its second assembly (12X.v2) and its third annotation (VCost.v3) comprises to date 33,568 genes (Canaguier et al. 2017). With the

purpose of providing biological meaning to this remarkable amount of data, several initiatives have been introduced for describing genes within their biological context (Grimplet et al. 2009a), including not only *in vivo* functional characterizations but also *in silico* analyses such as co-expression networks and other integrative approaches (reviewed by Wong and Matus 2017).

With the commitment of consenting the efficient exploitation of *Vitis* biological resources and understanding the genetic and molecular basis of all processes in this species, the International Grapevine Genome Program (IGGP; www.vitaceae.org) is currently developing the GrapeIS system. This is an integrated set of interfaces supporting advanced data modeling, rich semantic integration and the next generation of data mining tools linking genotypes to phenotypes (Adam-Blondon et al. 2016). Within the same framework, the recently launched INTEGRAPE consortium (COST Action-mediated) aims to integrate data at different levels to maximize the power of omics and establish a manageable and open data platform. The initiatives mentioned here share the use of FAIR principles that ensure data are Findable, Accessible, Interoperable and Reusable (Wilkinson et al. 2016). The establishment of solid integrative data platforms are compulsory to make available interoperable grapevine datasets and tools. The application of systems biology methods has arisen to fulfil this purpose. Here we provide a brief review of the fundamentals of systems biology and the history of applying integrative omics methods in grapevine research. The best-known programming scripts/packages and web-based resources for the analysis and interpretation of omics-generated data will also be described. Before examining the state of the art, a list of terms commonly used in the field of Systems Biology is presented in Box 8.1.

Box 8.1 Glossary of Terms

ATAC-seq: The technology that applies high-throughput sequencing to assay for transposase-accessible regions in the genome effectively analyzing chromatin accessibility.

Big Data/Data Science: An emerging discipline that combines computer science and statistics to analyze massive amounts of data with the goal of answering specific and practical questions of a phenomenon under study.

ChIP-seq: The technology that couples chromatin immunoprecipitation (ChIP) with high-throughput sequencing to analyze protein-DNA interactions.

Cistromics: The omics technology that analyses the cistrome or the complete set of binding sites of a given transcription factor to the DNA under specific conditions.

Community network: Network built from as few as three input networks, diminishing the limitations of each individual method. Edges supported by a higher number of methods are more reliable.

DAP-seq: The technology that couples *in vitro* expression of affinity-purified transcription factors with high-throughput sequencing of a genomic DNA library in order to analyze protein-DNA interactions.

Epicistromics: The omics technology that studies the epicistrome or the complete set of genomic locations occupied by nucleosomes carrying histones with distinct posttranslational modifications under specific conditions.

Gene co-expression network (GCN): A undirected network typically built from transcriptomic data such as RNA-seq or microarray data where nodes represent genes and edges are drawn between two nodes when the corresponding genes are significantly co-expressed under the analyzed conditions.

High performance computing: The use of supercomputers and parallel computational architectures to massively process information in order to solve complex problems.

High-throughput sequencing (HTS): Techniques that sequence massive amounts of DNA in an automatic and parallel manner. High-throughput in omics is referenced to the use of automation equipment to address biological questions that are otherwise unattainable using conventional methods.

MNase-seq: The technique that applies high-throughput sequencing to the DNA protected by nucleosomes during micrococcal nuclease digestion to effectively identify nucleosome positioning.

Molecular systems biology: An emerging discipline at the intersection between molecular biology, mathematics/statistics and computer science that integrates massive amounts of omics data with the final goal of generating predictive models of biological systems focusing on biomolecular interactions rather than on isolated molecular components.

Network: A model of a system where nodes represent the system components and edges between nodes indicate an interaction between the corresponding components. Networks can be directed or undirected depending on whether or not there exists a directionality in the interactions between the system components. Networks can be weighted when numerical values are associated with edges in order to capture specific features of the corresponding interactions.

Next Generation Sequencing (NGS): A term to describe a collection of genetic sequencing techniques that improve upon the original Sanger sequencing process. This technique utilizes DNA sequencing technologies that are capable of processing multiple sequences in parallel. Also known as massively parallel sequencing, deep sequencing or high-throughput sequencing (HTS).

Omics technologies: Techniques that detect and quantify massive amounts of molecules of a specific type from a sample.

Regulon: Group of non-contiguous genes that are regulated as a unit, generally controlled by the same regulatory gene that expresses a protein acting as a repressor or activator.

RNA-seq: The application of high-throughput sequencing to the cDNA corresponding to the entire set of transcripts in a sample. This technology allows researchers to detect and estimate the abundance of transcripts (coding and non-coding) in a sample, also including alternative splicing variants.

Transcriptional network: A directed network typically built from cistromic data corresponding to multiple transcription factors where nodes represent genes and an edge is drawn from gene_i to gene_j when gene_i codifies for a transcription factor that directly binds to the promoter of gene_j. Weights can be associated with edges to represent if the binding of the transcription factor has an activating, repressing or neutral effect over the transcription of a target gene.

Transcriptomics: The omics technology that focuses on the analysis of the transcriptome or the complete set of transcripts expressed from the genome under specific conditions.

8.2 From elements to relations: Overview of plant systems biology

Systems biology is a computational, mathematical and biology-based interdisciplinary field that focuses on complex interactions within biological systems. Its foundation outcomes from amending the general (Von Bertalanffy 1968) and living (Miller 1978) system theories and aims to elucidate biological phenomena applying a systemic view of interactions between molecular entities instead of describing their individual composition or function (Mesarovic 1968). By addressing the cell as a network of genes, their products and their interactions, the latter defined as network motifs or patterns, it's feasible to study

the structural design principles of living organisms. Distant networks that perform similar tasks (e.g. information processing) all share similar types of recurring patterns of interconnections, thus motifs define universal classes of networks (Milo et al. 2002). From this and other studies, it was suggested that structures of different networks were governed by the same principles. This new paradigm is embodied within the Oltvai and Barabási life's complexity pyramid, now re-updated and revisited by systems biology advancements (Figure 8.1). Here, cell components arrange themselves in persistent patterns and these in turn form modules with discrete cellular functions. Finally, these modules are hierarchically organized, defining the cell's large-scale functional organization.

Historically, reductionist studies in plants have been aimed for identifying the individual components associated with the occurrence of certain phenotypes. Although this approach has been massively adopted in the last 50 years, successfully producing extensive repertoires of plant molecular components, it begun to lose its effectiveness at the beginning of the current century when it became apparent that majority of phenotypes were produced by complex orchestrations involving myriads of molecular components, many of which were redundant among them. This scenario became more apparent with the development of the so-called omics technologies that provide an accurate molecular snapshot of the biological processes under study by detecting and quantifying the repertoire of molecules that are present (Yuan et al. 2008). Hence, research in molecular biology is gradually shifting towards a holistic perspective, integrating the individual 'omics' datasets, to gain biologically meaningful aspects of plant systems (Sheth and Thaker 2014).

The recent development of high-throughput DNA sequencing, genomics and transcriptomics have pushed these methodologies to become so far, the best-established

mature and reliable techniques to characterize molecular systems (Bolger et al. 2018). Specifically, RNA-seq, the high-throughput sequencing of the cDNA corresponding to the entire set of transcripts in a sample, is applied to identify and estimate transcript abundance including different isoforms produced by alternative splicing as well as to analyze differential gene expression between specific conditions (Martin et al. 2013). The main molecular mechanisms controlling gene expression, namely the interactions between transcription factors and DNA (recently named 'the cistrome'), and the different posttranslational modifications of histones associated with the DNA (epicistrome) are routinely characterized using techniques such as ChIP-seq; the combination of chromatin immunoprecipitation with the high-throughput sequencing of the purified DNA (Chen et al. 2017). DAP-seq is a technique based on high-throughput sequencing that studies the cistrome based on the *in vitro* expression of affinity-purified transcription factors (Bartlett et al. 2017). Finally, MNase-seq, DNase-seq and ATAC-seq are techniques used to study nucleosome positioning and chromatin accessibility that have been shown to highly influence gene expression (Pajoro et al. 2014; Sullivan et al. 2015; Pass et al. 2017; Bajic et al. 2018).

Despite the clear methodological and analytical advantages of performing genomics studies compared to other omics, it has been demonstrated that the sole use of genomics and transcriptomics is not sufficient to predict phenotypes from the molecular state of biological processes (Papatheodorou et al. 2015). In this respect, proteomics (the analysis of the proteome or the entire set of proteins), and metabolomics (the study of the metabolome or the complete set of metabolites), are currently under development aiming at providing a more exhaustive molecular description of biological systems (Ramalingam et al. 2015).

At this point, the massive amounts of data generated by omics technologies is being stored in public databases considerably exceeding the analytical capacities of humans, making imperative the use of computational resources to extract relevant information. Currently, this scenario is not exclusive to molecular biology as it pervades science in a more general context by inducing the emergence of the so call Big Data or Data Science. This is a discipline that combines high-performance computing, such as the use of computational clusters, with sophisticated statistical methods, in order to answer specific questions of phenomena under analysis (Carmichael et al. 2018). In molecular biology, this has promoted the development of "Molecular Systems Biology". This emerging discipline lays at the intersection between molecular biology, computer science and mathematics/statistics (Figure 8.2). The main methodology in molecular systems biology pertains to the generation of omics data and their integration with already existing data freely available in public databases. This massive amount of data is integrated and analyzed typically using multivariate statistical methods implemented with high-performance computing. Specifically, molecular systems biology pursuits the development of computational/mathematical models of the interactions among the molecular components of the systems responsible for an observed phenotype rather than focusing on the functioning of the isolated individual components. Here, the ultimate goal relates to the generation of tools that allow to model and predict the emergence of specific phenotypes or responses in biological systems (Sheth and Thaker 2014). Commonly, systems of differential equations are used as the modeling structure to achieve this goal. Nonetheless, network science is emerging as a central paradigm in molecular systems biology as an effective modeling framework (Li et al. 2015).

In the context of network science, a network is a graph whose nodes represent the molecular entities of the system and a directed or undirected edge is drawn between two

nodes to specify the interaction between the corresponding molecular components. A numerical value termed weight can be incorporated in the edges to capture the strength of the represented interaction. Topological studies of a network, such as the analysis of free-scale properties, can identify relevant nodes called hubs that are highly connected in the network and play key roles in network robustness and dynamics. Other topological parameters such as 'node transitivity', 'betweenness' and 'eccentricity' are especially suitable to identify relevant molecular components of the biological system under analysis. Clustering techniques and community analysis are used to unravel the underlying structure of networks and are applicable in molecular systems biology to identify molecular modules that function with a certain level of separation from the rest of the system (Aoki et al. 2007). Finally, network motif analysis or the identification of non-random subgraphs can shed light on the building blocks that occur recurrently in biological systems (Defoort et al. 2018).

Two types of gene networks are intensively used in molecular systems biology; gene co-expression networks and transcriptional networks. Gene co-expression networks are normally constructed based on a compendium of microarray and only recently, RNA-seq data sets. These are undirected networks where nodes represent genes and undirected edges are drawn between nodes to represent co-expression relationships between the corresponding genes. Transcriptional networks are constructed from ChIP-seq data corresponding to sets of different transcription factors binding to the genome. These are directed networks where nodes represent genes and a directed edge is drawn from gene_i to gene_j, where gene_i codifies for a transcription factor that binds to the promoter of gene_j. Transcriptional networks can be further refined by adding RNA-seq data corresponding to mutants or overexpressors of the transcription factors previously analyzed using ChIP-seq. According to this, weights can be associated with edges to

represent an activating, repressing or neutral effect of the binding of the transcription factor to the promoter of the target gene.

8.3 A decade conducting grapevine omics. What´s yet to come

Genomics resources for *Vitis* species have increased promptly within the last fifteen years, beginning with the sequencing of expressed sequence tags (ESTs) (Da Silva et al. 2005; Moser et al. 2005). These resources have permitted to quantitatively assess the grape transcriptome by aiding the development of cDNA and oligonucleotide microarrays (Terrier et al. 2005; Waters et al. 2005). Quantitative data acquisition through microarray analysis permitted large-scale mRNA profiling studies of gene expression to unravel the most important events of berry development and ripening. However, it was not but after the concomitant release of the *V. vinifera* cv. 'Pinot Noir' genome sequence (Jaillon et al. 2007; Velasco et al. 2007) that a burst of new transcriptomic technologies emerged for this species. In the Affymetrix Grape GeneChip Genome Array, approximately one-third of the expected genes are represented. This platform was largely used for tissue-specific mRNA expression profiling in grape berry tissues (Grimplet et al. 2007; Deluc et al. 2007) and responses to abiotic stresses (Tattersall et al. 2007; Cramer et al. 2007) and compatible viral diseases (Vega et al. 2011), where all the produced data were collected and unified in the PLEX database (PLEXdb, http://www.plexdb.org; Wise et al. 2007). The microarray Nimblegen platform was developed soon after (Fasoli et al. 2012; http://ddlab.sci.univr.it/FunctionalGenomics/), with an array representing more than 98% of the genes predicted in the 12xV1 grapevine genome annotation (090918 Vitus vinifera exp HX12 chip, with approximately 29,549 denoted genes). To date, this platform has generated the largest amount of transcriptomic data for this species (1605 experiments

until July 2018). All developed arrays in *Vitis* can be found in ArrayExpress EMBL-EBI; https://www.ebi.ac.uk/arrayexpress/).

Although *in situ* oligonucleotide arrays are still widely used for gene expression profiling in grapevine, a rapid development of new nucleic acid technologies have been largely adopted for genomic, transcriptomic and metagenomic studies in grapevine in the last years (Figure 8.3A). A variety of NGS technologies, including the 454 (Roche) (Margulies et al., 2005), the Genome Analyzer/Hiseq (Illumina Solexa) (Bennett et al. 2005) and the SOLiD (Life Technologies), as well as newer platforms such as Helioscope (Helicos) (Milos 2008), PacBio RS and Sequel (Pacific Bioscience) (Eid et al. 2009), Oxford Nanopore Technologies for single molecular sequencing and Ion Torrent (Life Technologies), based on a semiconductor chip (Rothberg et al. 2011), are available. Thanks to high-throughput and cost-efficient capabilities of these technologies, an unprecedented amount of data has been generated and a huge amount of genomic and transcriptomic data has accumulated exponentially in *Vitis* species (Figure 8.3B-3C).

The combination of high throughput sequencing technologies and the grapevine reference genome (Jaillon et al. 2007) has facilitated comprehensive sequence analysis in diverse grapevine germplasms (Table 1). Cultivars with different agronomic and oenological characteristics have been re-sequenced to identify genetic differences underlying the distinct phenotypes (Da Silva et al. 2014; Di Genova et al. 2014; Cardone et al. 2016; Chin et al. 2016, Minio et al. 2017; Minio et al. 2019; see Chapter 05) and comprehensive inventories of sequence variations were generated (Mercenaro et al. 2017; Zhou et al. 2017; Liang et al. 2019). On the other hand, transcriptome sequencing using NGS technologies has been widely used to detect gene expression in grapevines (see Chapter 08), including fruit (e.g., Zenoni et al. 2010), leaves (e.g., Liu et al. 2012), flowers (e.g., Domingos et al. 2016), in response to different biotic and abiotic stresses (e.g.,

Cheng et al. 2015; Blanco et al. 2015; Amrine et al. 2015; Tillett et al. 2011) or to describe the expression of specific transcription factors (e.g., Sweetman et al. 2012). Other grape researchers have used high-throughput expression to examine the phenotypic plasticity of cv. 'Corvina' berries at various developmental stages (Dal Santo.et al. 2013). Despite its primary objective is to characterize expression profile, RNAseq technologies have been also used to identify differential splicing activity and single nucleotide polymorphisms (Zenoni et al. 2010; Vitulo et al. 2014) as well as identifying and profiling long non-coding RNAs (Vitulo et al. 2014; Harris et al. 2017).

Since grapevine naturally hosts a reservoir of microorganisms that interact with the plant and affect both the qualitative and quantitative scale of wine production (Martins et al., 2013; Zarraonaindia et al., 2015), grape metagenomics studies also are assuming an increasing resonance in the grape scientific community. Recently, high-throughput technologies have been used to characterize bacterial communities of different grapevine plant portions, such as leaves and berries (Leveau and Tech 2010), to assess the microbial communities of soils (Zarraonaindia et al. 2015; Burns et al. 2015; Burns et al. 2016) and to survey the associations involving grapevine microbiota, fermentation and wine chemical composition (Bokulich et al. 2014; Bokulich et al. 2016).

Despite the study of epigenetic marks (e.g. histone post-translational modifications and DNA methylation) are known to influence gene expression and largely affect the phenotype of plants, there are still scarce epigenomic data and related resources available for grapevine. Nonetheless, Fortes and Gallusci (2017) recently proposed this species as an essential perennial woody plant model for such studies due to the impact of epigenetic modifications on agricultural traits, and also because epigenetic marks may serve as an interface between the environment and the genome (reviewed by Fabres et al. 2017). Very recently, Xie et al. (2017) used methylation sensitive amplified

polymorphisms (MSAPs) to find global patterns of DNA methylation and explored the genetic and epigenetic diversity of a single cultivar across 22 vineyards located in six different wine sub-regions.

Proteomics resources have also arisen in the last decade, despite at a much lower rate. While at the beginning most of these studies used two-dimensional gel analysis and focused on berry metabolism coupled to abiotic stress responses (Vincent et al. 2007; Jellouli et al. 2008; Grimplet et al. 2009b), high-resolution techniques have also been applied to grape such as iTRAQ (Lucker et al. 2009), or much more recently, 2DE gels coupled to liquid chromatography with electrospray ionization (LC-ESI-MS/MS; Negri et al. 2015), or nanoLC ESI LTQ-Orbitrap tandem mass spectrometry (Wang et al. 2017; Kambiranda et al. 2018).

Targeted and untargeted metabolome studies have unquestionably increased within grapevine research, benefiting from a variety of tools such as massive high-performance liquid chromatography (HPLC) and gas chromatography (GC) being applied for sample separation while tandem mass spectrometry (MS) and nuclear magnetic resonance (NMR) being developed for the identification and quantification of metabolites. Solid phase and micro solid phase extractions (SPE and SPME), followed by GC-MS methods have been used for volatile composition studies (Savoi et al. 2016; Duchêne et al. 2017). Ultra-High-Performance Liquid Chromatography (UHPLC) coupled to triple quadrupole (QqQ) TQD mass spectrometry analysis was recently used for determining polyphenomic composition (phenylpropanoid-specific omics) and its cultivar-dependent changes in response to drought (Pinasseau et al. 2017). Also, Vondras et al. (2017) recently performed untargeted HPLC-MS to quantify amino acids, sugars, organic acids, and phenylpropanoids to compare the different ripening progressions of berries in a single cluster, while Blanco et al. (2015) and Negri et al. (2017) studied the effect of *Botrytis*

cinerea noble rot infection in the metabolome of ripening berries and postharvest withered berries, respectively, by using reversed-phase HPLC coupled to ESI mass spectrometer.

Despite metabolomics analyses are rapidly increasing in *Vitis*, metabolism must be understood as a dynamic process. Fluxomics recognizes this complexity in metabolic systems and seeks to determine the rates of metabolic reactions (Winter and Krömer 2013). With the purpose of describing how metabolic fluxes determine cellular phenotypes, Soubeyrand et al. (2018) performed targeted metabolomics and enzyme activity measurements in grape cell cultures at different time-points of nitrogen limitation in order to construct a constraint-based model (by comparing maps of metabolic fluxes in the two contrasted situations) to identify the metabolic drivers of anthocyanin accumulation under high carbon-to-nitrogen ratios.

Within the cell's functions, the transport of essential and beneficial nutrients allows all basic processes to be performed efficiently. In grapevines, ion content profiles can reflect the mineral composition of soils and therefore they can describe certain components of a *terroir*. Pii et al. (2017) studied the ionomics profile of berries grown in different areas to try to discriminate their geographical origin. By applying multi elemental inductively-coupled plasma-mass spectrometry (ICP-MS), the authors found that rare earth elements were the best chemical descriptors.

Recent attempts for identifying transcription factor binding landscapes have been initiated and deposited in public repositories, despite no publications have yet been produced. Additional efforts are still needed to map protein-DNA and protein-protein interactions at a large scale. Also, DNAse I hypersensitivity mapping could be useful to identify pioneering transcription factors controlling grape and wine quality traits.

8.4 From single omics to integrative data analysis

Within single omics studies the interactions between molecules can be represented in networks, where nodes (genes, proteins, metabolites, etc.) are connected by edges that convey any type of association (e.g. relying in abundance or expression levels). In the case of gene co-expression networks (GCNs), edges represent similar gene expression behaviors, while in genome-wide transcription factor binding studies (e.g. ChIP-seq) edges represent direct target-regulator relationships. In protein-protein interaction networks, edges describe physically interacting protein pairs identified from techniques such as high-throughput yeast two-hybrid screens.

Beyond single omics networks, integrative approaches associate the molecular components of an organism and combine them into higher order networks to model dynamic behaviors. The principle is based in the fact that despite individual functions of a single network may be undetermined, its biological role can sometimes be inferred through association with other networks. Integrated/combined networks provide a more complete information of a certain biological processes as they include two or more omics' layers. In the case of combining several networks of the same type into a community network, this can also be beneficial to effectively reveal discrepancies between individual networks while stressing common associations across individual networks (Proost and Mutwil 2016). Networks of experimental evidence can be integrated by superimposing the nodes from individual networks. However, an appropriate integrative method requires biological data to be normalized, standardized, modeled and visualized in order to build an integrated model (Figure 8.4). Data modeling requires special attention as this analysis involves generalization and simplification steps with several assumptions (Yuan et al. 2008).

The first task to perform during the integration of different multi-dimensional omics data consists in matching the features within each omics, as they measure diverse types of molecules and the correspondence between them is not always straight forward. For instance, a single gene can produce several transcripts with different alternative splicing. Similarly, a single transcript can give rise to multiple proteins through different posttranslational processes, making it difficult to associate genes, transcripts and proteins when measured by genomics, transcriptomics and proteomics techniques. Moreover, cistromics and epicistromics measure transcription factor binding and occupancy of nucleosomes carrying distinct histone modifications in specific genomics regions. The association of these regions to target or regulated genes is not trivial. This problem can be tackled using different software packages such as *RGmatch* (Furió-Tarí et al. 2016), *PeakAnalyzer* or *PeakAnnotator* (Salmon-Divon et al. 2010).

Additional challenges faced during multi-omics data integration are represented by the heterogeneity of the different data sets. Data from each omics is measured using different units whose typical ranges vary in several orders of magnitude. This can potentially affect data analysis and is typically solved using scaling and normalization techniques. Given the wide spectrum of possible normalization techniques it is necessary to apply as many as possible and asses their performance in order to choose the most appropriate technique for the data sets under study. The R package *Normalyzer* can be applied in this pre-processing of the data (Chawade et al. 2014).

Once data pre-processing is completed and prior to the actual multi-omics integration, some exploratory analyses need to be conducted over the individual data sets. Due to the high dimensionality of omics data typically these analyses consist in techniques able to reduce complexity in order to extract relevant information. Principal Component Analysis (PCA) constitutes the most widely used projection method in this

step. PCA is a multivariate analysis technique whose final goal is to reduce the dimensionality of a large multivariate data set. Here a set of new uncorrelated or orthogonal variables are computed as linear combinations or rotations of the original ones. These new variables are called principal components and they are defined in such a way that they are sorted according to the percentage of explained variability from the original data under the constrain of being orthogonal or uncorrelated. In this way, typically, the first two or three principal components are sufficient to capture most of the variability of the original data and therefore, a projection comprising only these principal components are further considered in the analysis. Graphical representations of the selected principal components are then used to assess the quality of data replicates, uncover problems raised during sample collection (e.g. batch effects) or to unveil underlying structure in the data by applying clustering techniques. Several R packages are available to perform this step such as *factorMineR* (Lê et al. 2008) and *made4* (Culhane et al. 2005), among other methods. For instance, a clear example of data integration in grapevine was conducted by Blanco et al. (2015) by using Multiple Factor Analysis (MFA), where four types of quantitative variables were considered: metabolome data, RNA-Seq data from grape and the fungi *Botrytis cinerea*, and *B. cinerea* biomass measurements.

Finally, multi-omics data integration is carried out. Normally, two different goals exist when integrating different omics. On one hand, researchers may be interested on exploratory analysis to identify the underlying relationship between two omics data sets. On the other hand, researchers may treat one of the omics data set as *response variables* that need to be predicted from another explanatory omics data set (considered as *predictors*). Here we discuss two statistical methods that exemplify these two goals. In both cases the input consists of two numerical matrices, $X_{n\times p}$ and $Y_{n\times q}$, that can be

generated using two different omics technologies that detect and quantify p and q as different molecules from the same set of n samples.

Canonical correlation analysis (CCA) This is an example of exploratory analysis that generates rotations or linear combinations, U and V, of the original data, X and Y, under the constrains of maximizing the correlation cor(Ui,Vi) with i = 1,…, min(p,q) and being uncorrelated or orthogonal. These are called canonical variates. Finally, like in any projection technique, only the two or three first canonical variates are considered to capture most of the correlation between the initial data X and Y. Several R packages are available to carry out this methodology such as *CCA* (González et al. 2008) and *mixOmics* (Rohart et al. 2017).

(Sparse) Partial Least Square regression (s)PLS is an example of a multi-omics integration technique in which researchers aim at predicting one omics data set (or physiological data) from another one. In a similar fashion to CCA, rotations U and V of the original data are performed by maximizing the covariance. Projections retaining only two or three components are then considered to perform linear regression. To assess the predictive power of the developed model, cross-validation is commonly applied. In classical PLS regression all the original variables from X and Y are included in the rotation or linear combination making intractable the extraction of relevant information from the developed model. In order to tackle this, the sparse variant of PLS regression (sPLS; González et al. 2012) was introduced by using penalization terms based on the marginal contribution of each variable to the predictive power of the model in such way that some coefficient shrinks to zero removing the corresponding variable. This efficiently implements a feature selection technique. Graphical representations such as correlation

circle plots, relevance networks and clustered image maps can be generated to facilitate the understanding and interpretation of the constructed model. The R packages *pls* (Mevik et al. 2007) and *mixOmics* (Rohart et al. 2017) implement the necessary functions to apply this methodology.

8.5 Recent experiences in grapevine systems biology

Throughout the last years several attempts for representing large biological data in networks have been conducted for elucidating the multilayered organization of biological processes in grapevine. In this species, integrated network analyses have been mostly adopted to predict gene functions or to contribute in the study of the regulatory mechanisms that control berry composition and development, trigger defense responses to biotic and abiotic stresses or that are influenced by the *terroir* (reviewed by Wong and Matus 2017; Fabres et al. 2017). Some research efforts have defined composite networks of genes and secondary metabolites for characterizing fruit ripening processes in red and white-skinned cultivars (Massonnet et al. 2018; Palumbo et al. 2014; Zamboni et al. 2010), whereas others have constructed gene co-expression networks to describe late stages of ripening (Ghan et al., 2017) or characterize transcriptional regulators related to development, metabolism or stress responses (Loyola et al. 2016; Wong et al. 2016; Sun et al. 2018). Processes involving the rewiring of berry metabolite-transcriptional networks under environmental perturbations such as drought (Savoi et al. 2016; Savoi et al. 2017) and elevated light exposure (du Plessis et al. 2017) have also been described. Proteomic/metabolomic composite networks (Wang et al. 2017) and those integrating genome-wide analyses of promoter regulatory elements (Wong et al. 2017) have also been generated. The integration of all these data in multilayered networks has allowed

building complex maps of molecular regulation and interaction. Some relevant cases will be covered in this section.

9.5.1 Identifying molecular hubs controlling light and cold response pathways

The advent and continued adoption of high-throughput transcriptome profiling platforms in grapevine research has led to the vast expansion of transcriptome datasets representing a wide range of experimental conditions (e.g. specific tissue/organ and its associated developmental series, stress – abiotic and biotic, vineyard management strategies, etc.). Although each dataset has been generated to address specific goals of its overarching study, together, individual datasets can be compiled into large expression databases to mine for novel biological insights including, but not limited to, comparative transcriptomics between grapevine and other plants, gene co-expression network analysis and functional assignment of genes, and the discovery of condition-specific *cis*-regulatory motifs (reviewed in Serin et al. 2016).

Genes involved in the same processes might share similar gene expression dynamics across an extensive collection of experiments. This relation, explained by the 'guilt by association' principle (Wolfe et al. 2005), is fundamental to infer the roles of uncharacterized genes in co-expression networks. Transcription factors (TFs) comprise a suitable case of study for addressing the behavior of modules in GCNs as they exhibit plethora of protein-protein and protein-DNA interactions, shaping complex regulatory networks responsible for most developmental process. Such is the case of ELONGATED HYPOCOTYL 5 (HY5) and HY5 HOMOLOGUE (HYH), two bZIP master photomorphogenic orchestrators involved in developmental processes responsive to light environmental conditions. Loyola et al. (2016) combined microarray and RNA-Seq co-expression data with a genome-wide binding site promoter inspection to identify HY5

and HYH community gene co-expression and *cis*-regulatory sub-networks in grapevine. Search of potential gene targets identified a preferential regulation of photosynthetic-related processes, heat-shock and DNA/protein repair processes, and regulation of the flavonol biosynthetic pathway. This study was crucial for describing the molecular mechanisms explaining the high radiation adaptive mechanisms that grapevines possess (reviewed by Matus, 2016).

Gene co-expression networks have also been integrated with transcription factor binding data to address grape responses to low temperature, in relation to the role of a MYB-like regulator termed AcQUIred tolerance to LOw temperatures (AQUILO; Sun et al. 2018). Here, the authors performed a multispecies GCN, incorporating gene co-expression analysis and *in silico* TFBS data from grape, with co-expression (associated to the heterologous overexpression of AQUILO) and DAP-seq data in Arabidopsis. The relevance of this study came from the finding that AQUILO was tightly associated with the raffinose family of oligosaccharides (RFOs), a connection that was later validated by quantifying these osmoprotectant molecules in cold-treated grape AQUILO-overexpressing calli.

8.5.2 Regulation of phenylpropanoid metabolism

Presently, the most widely adopted methodology to identify candidate transcriptional factors (TFs) involved in secondary metabolism pathways in grapevine involves the inference of function via sequence homology with functionally characterized proteins from model plants (for example, see Hichri et al. 2010, Cavallini et al. 2015 and Matus et al. 2017). However, in the recent years many of these regulators have been prioritized by using gene co-expression network analyses. For example, the putative functions of

134 grapevine R2R3-MYB genes were inferred based on their top 100 co-expressed genes (Wong et al. 2016). This study revealed that GCNs of many R2R3-MYB TFs (46 genes) were enriched with secondary metabolism-related functions. Demonstrating the power of such method is the ability to recover expected relationships between structural pathway genes and their known transcriptional regulators. For example, this was demonstrated with the frequent co-expression of large suites of *STILBENE SYNTHASE* genes (*STSs*) with VviMYB14 and VviMYB15, two R2R3-MYB TFs involved in the regulation of *STS* (Höll et al. 2013). Similar inferences were accounted for VviMYB13, a close homolog of VviMYB14 and VviMYB15, therefore suggested as involved in the regulation of tissue- and stress-specific *STS* expression (Wong et al. 2016). Two recent studies have also used *STS* genes as 'guides' to identify co-expressed TFs in both condition-specific (Wong and Matus 2017) and -independent contexts (Vannozzi et al. 2018). A berry-specific GCN encompassing five red cultivars across four key berry developmental stages revealed novel roles for AP2/ERF and WRKY TFs in the regulation of *STSs*. TFs of the latter two families were not only frequently co-expressed with *STSs* but were also enriched for their respective TF binding sites (TFBS) in the promoters of many *STSs*. Recent studies have now demonstrated that VviWRKY24 and VviWRKY03 are additional players in the regulation of *STSs* at various hierarchies – acting as singular effector or in synergy with VviMYB14 to activate STSs (Vannozzi et al. 2018).

The integration of non-coding RNA network analysis to existing condition-specific GCNs has also been presented to unravel the regulation of phenylpropanoid and flavonoid biosynthesis during berry development and ripening (Wong and Matus 2017). One of the key findings from this initiative was the discovery of long non-coding RNAs (lncRNAs) that were not only strongly correlated with key structural pathway genes but were also located in close proximity to their co-expressed gene). The lncRNA

VIT_210s0042n00100, present in close proximity with all nine *VviSTSs* of chromosome 10 presented consistent co-expression with all of them. Another case represents one predicted lncRNA (VIT_203s0180n00020) that is linked to VviGT2 through strong co-expression and co-location. This gene encodes an enzyme putatively involved in hydroxycinnamic ester biosynthesis and proanthocyanidin galloylation (Khater et al. 2012).

GCN approaches may reveal additional layers and deconvolute the complexities of secondary metabolic pathway regulation in grapevine. Indeed, in a first study of its kind, Zhang et al. (2018) demonstrated that multiple lncRNAs, named LNC1 and LNC2, were involved in the regulation of anthocyanin biosynthesis in fruits of sea buckthorns (*Hippophae sp*.) by serving as endogenous target mimics (eTM) of miR156a and miR828, respectively. Functional studies confirmed that silencing of LNC1 and LNC2, led to the induction and repression of anthocyanin biosynthetic pathway gene expression and anthocyanin levels in fruits, respectively, validating the integrated lncRNA-miRNA-mRNA network prediction.

9.5.3 The fight club goes dry: networks related to grape berry ripening in response to drought

To understand the molecular mechanisms underpinning berry development and ripening at greater detail, recent efforts have focused on understanding the transcriptome dynamics in multiple cultivars across the entire process of berry development and ripening. A study by Massonnet et al. (2018) represented the first monumental study to catalogue the genome-wide transcriptional profile of ten Italian grapevine varieties at four critical stages of berry development, all being cultivated in a single vineyard. In less than a handful of studies, network-based approaches have been applied to identify genes

potentially involved in critical developmental stage transitions. Such cases often complement the findings from the widely-adopted differential expression analysis but are also pivotal in revealing novel genes and relationships that were otherwise unattainable from traditional differential expression methods. For example, berry-specific gene co-expression network analysis encompassing immature-to-mature transitions has been particularly insightful in revealing groups of genes with distinct topological properties that can be classified into 'party', 'date' (see Han et al. 2004 for details), or 'fight-club' hubs (Palumbo et al. 2014). Genes that belong to the 'fight-club' hubs in particular were often negatively correlated with their interacting partners in gene co-expression networks, and those who do, were inferred as biologically relevant 'switches' fulfilling negative regulatory roles in the transition of major developmental phases such as ripening. Although the identity of these major switches was first documented in red grapevine varieties, recent research has now ascertained several common but also reveal variety (red and white-skinned)-specific switch genes (Massonnet et al. 2018). From a total of 271 berry-specific switch genes identified to date, 131 genes were in common in both varieties while 81 and 50 genes were specific to all white and red varieties, respectively. A large proportion of these 'switches' encode for transcription factors (31 genes), followed by genes involved in stress responses (31 genes), carbohydrate metabolism (22 genes), signaling (20 genes), secondary metabolism (20 genes), and cell wall metabolism (18 genes), among others (Massonnet et al. 2018).

Recent works have provided evidence for the involvement of multiple stress regulons – both ABA-dependent and ABA-independent (reviewed in Nakashima et al. 2014) – in the berry ripening program (Savoi et al. 2017). Certain TF families (e.g NAC, bZIP, AP2/ERF) that share co-expression with downstream water deficit stress-responsive genes may be required to orchestrate the balance between the progression of

berry development and stress-associated transcriptional regulation. Further analysis of gene co-expression and gene-metabolite co-response networks of the berry subjected to water deficit stress across critical berry development and ripening phases revealed several distinct modules that were congruently induced by ripening and water deficit stress (Savoi et al. 2016; 2017). Here, metabolome and transcriptome integrated network-based analysis revealed close associations between the expression behaviors of module members (especially the activation of multiple signal transduction pathways) and the dynamics of key central and specialized metabolites involved in the drought response (e.g. proline, branched-chain amino acids, phenylpropanoids, anthocyanins, and free volatile organic compounds). For example, the grapevine homologue of Arabidopsis ERF1, a key regulatory component of the jasmonate and ethylene signaling network (Cheng et al. 2013), whose expression was congruently induced by ripening and water deficit stress, was also identified to be a common berry 'switch' gene. While its precise regulatory role remains to be elucidated, integrated network analysis positioned ERF1 as a putative regulator of proline and anthocyanin accumulation in the berry (Savoi et al. 2017). *VviERF1* was significantly co-expressed with pyrroline-5-carboxylate synthase (*P5CS*) and *VviMYBA2*, the key structural gene of proline biosynthesis and a key regulatory gene of anthocyanin biosynthesis in the berry, respectively and shared significant correlation with various anthocyanin compounds. The presence of potential AP2/ERF TFBS (i.e. DRE and GCC-box) situated within the promoter region of P5CS and MYBA2 further reinforce its involvement as a regulator of berry composition during ripening and water deficit stress.

9.5.4 Non-coding RNA networks within grape-fungi pathosystems

Grapevine diseases caused by biotic agents can be devastating for the wine and table grape industries. Among fungal-related disorders, grape trunk diseases together with downey and powdery mildew are among the most important pathologies, causing significant economic losses in vineyards practically all over the world. The symptoms of downey mildew, caused by *Plasmopara viticola*, are quite detrimental, as for instance, as soon as fruits become infected, berries completely dry out. The *Vitis* sp. – *P. viticola* association is of great interest as this oomycete is an obligate biotroph and relies entirely on the host to complete its life cycle (i.e. needs to keep its host cells alive before sporulation; Grenville-Briggs and van West 2005), and also because North American *Vitis* species are naturally resistant (Polesani et al. 2010). In order to model this complex pathosystem, Brilli et al. (2018) performed a multi-omics and multi-species functional genomic study. The authors sequenced and assembled the draft genome of *P. viticola*, identifying the lost metabolic features responsible for its total dependence on the grape host, and further studied the fungus transcriptome changes occurring during the infection process, identifying a protein triggering immunity in the resistant *V. riparia*. The most striking results from this study arise from the small RNA sequencing (sRNA-Seq) analysis in control and infected plants at different times after the infection, combined with genome-wide degradome (or parallel analysis of RNA ends) analyses in both the plant and the oomycete. As a result, a large number of sRNA-mediated cleavages exclusively occurred in infected tissues, where sRNAs produced by *P. viticola* triggered cleavage of grapevine genes while sRNAs processed from grapevine transcripts targeted the fungus mRNAs, unveiling a bi-directional RNA silencing network mediated by non-coding RNAs shuffling between the pathogen and its host (Brilli et al. 2018). As more pathogen genomes become available, a broader understanding of pathosystems and their dynamics

will be achieved, especially regarding the roles of secreted effectors in interfering plant immune recognition (reviewed by Dalio et al. 2018).

Grape pathogen responses have been recently studied by addressing potential interactions of transcription factors and cis-regulatory element (CRE), and also by constructing gene co-expression networks (GCNs) of plant gene families related with defense. Wong et al. (2017) performed a genome-wide analysis of known plant CREs in all grape predicted protein-coding gene promoters, constructing an integrated CRE-driven network. Numerous CRE-driven modules inferred from using condition-dependent GCNs suggested important roles in pathogen stress responses. For example, *GCC*-core sub-modules were contained in many genes that were highly induced in berries and leaves infected with fungi such as *Botrytis cinerea* and *Erysiphe necator*. Finally, gene co-expression networks of the ATL protein family showed that many of these E3 ubiquitin ligases were induced in grapevine–pathogen interactions including *P. viticola* and necrotrophic fungi (Wong et al. 2018).

8.6 Resources

Next-generation sequencing as well as traditional Sanger sequencing methods are of great significance in unraveling the complexity of plant genomes. These are constantly generating heaps of sequence data to be analyzed, annotated and stored, thus creating a revolutionary demand for resources and tools to manage and handle these necessities (Basantani et al. 2017). Here we present a brief compilation of web resources that are either specific for grape or encompass a variety of species including *Vitis* sp (Table 2).

At least two grape-specific platforms have been effectively used to study the extent of gene regulatory networks: the ViTis Co-expression DataBase (VTCdb; Wong et al. 2013) and VESPUCCI (Moretto et al. 2016). These resources have played an

important role in determining the roles of genes related to photomophogenic responses and secondary metabolism in targeted functional studies (Loyola et al. 2016; Malacarne et al. 2016). Integration of multi-omics datasets (i.e. gene expression, metabolite, and protein profiles), mapping of data onto relevant molecular networks, and the visualization of the dynamic interactions between the various molecular classes are also the first few steps when performing any systems biology experiments. Tools such as Cytoscape (Shannon et al. 2003) have been specially designed for this task and have been largely adopted by the grape research community to visualize and analyze complex networks. In addition, one ongoing Initiatives in grapevine, VitisNet (Grimplet et al. 2009a) serve as a resource for manually curated functional gene annotation and provides a wide range of manually curated pathway-level molecular networks (over 240 categories) as templates for grapevine systems biology experiments.

The increasing release of plant genomes provided unseen opportunities and challenges for comparative genomics resources. Indeed, different genomics multi-species platforms also exist constituting relevant hubs to exploit omics data in grape. For instance, recent examples include the fruitENCODE platform (http://www.epigenome.cuhk.edu.hk/encode.html) that provides a comprehensive repository oriented to shed light on the genetic and epigenetic basis of fruit ripening in climacteric and non-climacteric species. Multi-species GCNs allowing comparative co-expression analysis are also now available for many plants including grapes (**Table 2**). Resources such as ATTED-II (http://atted.jp/) are amongst the most popular, providing the opportunity to query microarray and RNA-seq GCNs using the 'guide' gene approach. ATTED-II also allows assessments of co-expression conservation of co-expressed genes across different plant lineages (Obayashi et al. 2018). The Plant Omics Data Center (PODC; http://plantomics.mind.meiji.ac.jp/podc/) is a NGS-derived gene expression

network repository aimed at integrating large-scale omics resources for a broad range of species (Ohyanagi et al., 2015). Such resources may be used in conjunction with existing grapevine-specific co-expression platforms to build community GCNs or to gain additional insights into the evolutionary context of conserved and/or species-specific co-expressed genes relationship.

Additional multi-species platforms gathering grape's omics and mainly aimed at comparative studies include Ensembl Plants (http://plants.ensembl.org) (Bolser et al., 2016), Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012), PlantGDB (http://www.plantgdb.org/) (Duvick et al., 2008) and AraNet v2 (http://www.inetbio.org/aranet) (Lee et al., 2015). These integrative resources encompassing genome-scale information (genome sequence, gene models, functional annotation, polymorphic loci, expression) offer a variety of sequence analysis tools and web services. Example of integrative platforms also come from other species including both model (Araport, Solgenomics) and non-model (Melonomics, Ginseng Genome Database) plants. A common feature underlying these resources rely on the use of customized instances of JBrowse (Buels et al. 2016), a fast and full-featured genome browser built with JavaScript and HTML5. Thanks to its speed, scalability and versatility this platform supports complex interactive queries on large track sets representing a suitable and solid mean to handle omics data in a genomic context. In addition, a variety of analysis functions can readily be added using the plugin framework (e.g. visualization of whole-genome bisulfite sequencing data, glyphs for variants and GWAS data, small RNA visualization, etc.). Very recently, a JBrowse (v. 1.11.5) was set up to visualize and give access to some omics data in the *Vitis vinifera* 12X.v2 PN40024 assembly (https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/Vitis/data_gff) (Canaguier et al. 2017). The platform hosts 11 annotations tracks, including the different

releases of the grapevine genome annotations (CRIBI v1, CRIBI v2, Genoscope, Cost v3, etc.), automated and manual curated transposable elements annotations and manual curated gene family sets. In addition, 12 tracks highlighting the variants coming from re-sequencing experiments are also present in the platform, which could help in the identification of useful markers for applied research purposes.

8.6.1 VESPUCCI and NES^{2}RA as grape-oriented resources

Exploring shifts in gene expression as response to different experimental conditions has become commonplace whilst transcriptomic experiments are being performed on a daily basis. Public available gene expression datasets, however, conceal most of their true potential since they are meant to answer to a specific biological question and aren't considered in the light of a wider context. Within transcriptomics, we have witnessed a major shift in data production with the advent of high-throughput sequencing technologies. Despite nowadays Illumina sequencing is the *de facto* standard for RNA-seq experiments, microarrays are still extensively used and, more importantly, constitute a wealth of public information available to be explored.

With the advent of systems biology approaches in grapevine research, data integration arises as a leading aspect to take advantage of such rich sources of information (Gligorijević and Pržulj 2015). Different methods have been proposed to carry out the task of effectively integrating gene expression data and can be usually divided in two categories: i) direct integration and ii) meta-analysis. Direct integration (Rung and Brazma 2013) considers the sample-level measurements within each study and merges them into a single data set. The latter (Garrett-Mayer et al. 2008), instead, integrates gene expression analysis combining information from several data sources defining confidence

levels for each study individually (without a general scheme) and is commonly used to integrate conclusions coming from different studies.

One of the platforms used for data integration in transcriptomics is COLOMBOS (Moretto et al. 2016a), originally named as a COLlection Of Microarrays for Bacterial OrganismS, which was developed for three bacterial species (*Escherichia coli*, *Bacillus subtilis*, and *Salmonella enterica* serovar Typhimurium) and later updated with others prokaryotic species and also including RNA-seq technology. The implementation of the COLOMBOS framework to the *Vitis* species led to the development of VESPUCCI (Moretto et al. 2016b) (Vitis Expression Studies Platform Using COLOMBOS Compendia Instances), an integrated gene expression database for grapevine that originally included 1,500 samples at the time of its first release and now has doubled in size including most of publicly available transcriptomic data.

Both VESPUCCI and COLOMBOS fall under the direct integration methodology. Their approach to data integration is unique in the sense of directly combining gene expression information from different technological platforms and experiments, without the need for batch-normalization since it calculates log-ratios for contrasts, i.e. samples being compared that come from the same experiment and platform combination (a 'batch'). This results in crossing out a high proportion of batch-related variation (Luo et al. 2010). While gathering a large amount of data is made easy for model organisms like *E. coli* (due to the abundant number of experiments available), for non-model species the situation is different as only fewer experiments are usually performed. In this case the importance of transcriptomics data integration is even more significant as an adequate magnitude of data is needed to be able to draw valid and general conclusions. In this sense, working with plant species highlighted the need for the authors to significantly rethink some aspects of the data acquisition and annotation process. The creation of a gene

expression compendium using COLOMBOS technology is facilitated by the use of COMMAND (Moretto et al. 2019), a web-based application used to download, collect and manage gene expression data from public databases, but it is still mainly a manual effort. The peculiarity and complexity of plant transcriptomes and experimental designs in plant biology require the ability to manage how probes (for microarray) and short read sequences (for RNA-seq) are mapped and thus assigned to genes. The concept of 'measurable transcript' was also used to account for some technical limitations that prevent the possibility to precisely distinguish among genes with high sequence similarity. In VESPUCCI, data and experiment-related information (meta-data) are collected and curated starting from raw intensities (for microarrays) and raw sequence reads (for RNA-Seq). A robust normalization method and a quality control procedure are performed to allow the direct comparison of gene expression values across different experimental conditions (Engelen et al. 2011). This results in a single coherent gene expression matrix in which each row represents a gene and each column represents a 'sample contrast'. Sample contrasts measure the difference (in log scale) between a test and a reference condition, both which are designed *a priori* by curators during the compendium creation process. The expression data itself is a matrix of log-ratios (base 2), so that positive values represent up-regulation, and negative values represent down-regulation of a gene in the test sample compared to the reference sample. VESPUCCI's main goal is to gather together as many expression data as possible to explore patterns of co-expression across several experimental conditions and to provide a high-quality gene expression database to be used for downstream analysis. The creation of a co-expressed genes cluster (known as module) is performed similarly to a BLAST (Camacho et al. 2009) search in which the users can look for expression values for a given set of conditions but using expression correlation instead of sequence similarity to score the best matches. Modules can be

modified in several ways in order to highlight the behavior of the genes of interest and to analyze (anti)co-expression patterns.

Considering that gene expressions are represented as relative values, it is fundamental to extensively annotate samples with various sorts of meta-data to ensure that valid biological conclusions can be drawn from the exploration of the compendium. One of VESPUCCI's biggest effort and most notable feature is the manual curation and quality check of samples. Each sample has been annotated by curators using controlled vocabularies to ensure both human readability and computational tractability. To completely fulfill the properties of the FAIR (Findable Accessible Interoperable Reusable) principles (Wilkinson et al. 2016), VESPUCCI is undergoing a constant renovation to exploit standards and bio-ontologies for data annotation. Finally, the interface is the other pivotal point towards seamless integration with other services and tools and has been designed to adapt to users' needs, as well as to simplify the implementation of other tools on top of it. One example of such means is the NES^{2}RA algorithm (Asnicar et al. 2018), a mining tool for transcriptomic data used to expand a known local gene network (LGN) by finding new related genes. This method has been applied to the grapevine transcriptomic dataset using VESPUCCI as data source to expand LGNs related to the secondary metabolic pathways for anthocyanin and stilbenoid synthesis and signaling networks related to the hormones abscisic acid and ethylene (Malacarne et al. 2018). Compared to Pearson correlation, NES^{2}RA LGNs show less edges as it removes less significant interactions, due to noisy or redundant information. This allows to reduce the complexity of the network and focus on the network topology and the most likely gene interactions. NES^{2}RA is computationally demanding and relies on the BOINC platform that distributes supercomputation tasks among computers made available by the volunteers participating in the gene@home project.

Besides the importance of having a single point of access to easily check at what is already available in terms of transcriptomic experiments in grapevine and, of course, the possibility to empower data analysis with thousands of integrated samples, the development of VESPUCCI has led to few considerations about the importance of correctly annotating experiments, extrapolable to all types of resources. Building the compendium itself was the most time-consuming step, as curators devoted their time and ongoing effort to describe sample conditions and their key descriptors, after carefully reading the experiment descriptions as well as scientific papers. The importance of early annotation of experiments as soon as (or even before) data are available is also underrated. It is often considered as an annoying request to fulfill before the publication, while it should be treated as an integral part of the experimental design with the same importance as notes and protocols written in lab notebooks have.

8.7 Final remarks

The accuracy of molecular systems biology relies on efficient methods that handle, analyze and visualize large omics data sets. However, it has become evident that the use of a single omics technology is not sufficient to develop predictive models, which in turn is the ultimate goal of this new discipline. Accordingly, the multiple use of technologies such as transcriptomics, cistromics, epicistromics, proteomics and metabolomics, over the same samples or biological conditions has started to be a central methodology in plant molecular systems biology. Multi-omics network modeling has proven to be a successful advance for unraveling the structure of biological processes in plants, as it allows identifying the key components and interactions for system regulation. Conversely, networks frequently require assumptions for data modeling, and since their methods may rely on the existing knowledge regarding the components and interactions of a system,

they can evolve to more exactly represent a biological system. Thus, data should be interpreted carefully while these approaches can be complemented by reductionist methods. Notwithstanding these limitations, the use of these methodologies in grapevine research have provided novel perspectives for interpreting omics data and despite its just starting, it is already challenging the analysis of the large amount of data that its being generated for this species.

References

Adam-Blondon AF, Alaux M, Pommier C, Cantu D, Cheng ZM, Cramer GR, Davies C, Delrot S, Deluc L, Di Gaspero G, Grimplet J, Fennell A, Londo JP, Kersey P, Mattivi F, Naithani S, Neveu P, Nikolski M, Pezzotti M, Reisch BI, Töpfer R, Vivier MA, Ware D, Quesneville H (2016) Towards an open grapevine information system. Hortic Res 3:16056.

Amrine KC, Blanco-Ulate B, Riaz S, Pap D, Jones L, Figueroa-Balderas R, Walker MA, Cantu D. (2015). Comparative transcriptomics of Central Asian *Vitis vinifera* accessions reveals distinct defense strategies against powdery mildew. Hortic Res. 2:15037.

Aoki K, Ogata Y, Shibata, D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol 48: 381 – 390.

Asnicar F, Masera L, Coller E, Gallo C, Sella N, Tolio T, Morettin P, et al. (2018). NES^{2}RA: Network Expansion by Stratified Variable Subsetting and Ranking Aggregation. The International Journal of High Performance Computing Applications 32 (3): 380–92.

Bajic M, Maher KA, Deal RB (2018) Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. Methods Mol Biol 1675: 183 – 201.

Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc 12: 1659 – 1672.

Basantani MK, Gupta Divya, Mehrotra Rajesh, Mehrotra Sandhya, Vaish Swati, Singh Anjali (2017) An Update on Bioinformatics Resources for Plant Genomics Research. Current Plant Biology, 11-12: 33-40.

Bennett S, Barnes C, Cox A et al. (2005) Toward the 1,000 dollars human genome. Pharmacogenomics 6:373–382.

Blanco-Ulate B, Amrine KC, Collins TS, Rivero RM, Vicente AR, Morales-Cruz A, Doyle CL, Ye Z, Allen G, Heymann H, Ebeler SE, Cantu D. (2015). Developmental and Metabolic Plasticity of White-Skinned Grape Berries in Response to Botrytis cinerea during Noble Rot. Plant Physiol. 169(4):2422-43.

Bokulich NA, Thorngate JH, Richardson PM, Mills DA (2014) Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proceedings of the National Academy of Sciences. 111(1):E139–48.

Bokulich NA, Collins TS, Masarweh C, Allen G, Heymann H, Ebeler SE, et al. (2016) Associations among Wine Grape Microbiome, Metabolome, and Fermentation Behavior Suggest Microbial Contribution to Regional Wine Characteristics mBio. 7(3):e00631–16.

Bolger M, Arsova B, Usadel B. (2018) Plant genome and transcriptome annotations: from misconceptions to simple solutions. Briefings in Bioinformatics 19, 3, 437 – 449.

Bolser D, Staines DM, Pritchard E, Kersey P (2016) 'Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data'. Methods Mol Biol, 1374:115-40.

Brilli M, Asquini E, Moser M, Bianchedi PL, Perazzolli M, Si-Ammour A. (2018) A multi-omics study of the grapevine-downy mildew (*Plasmopara viticola*) pathosystem unveils a complex protein coding- and noncoding-based arms race during infection. Sci Rep. 8(1):757.

Buels R, Yao E, Diesh C. M., Hayes R. D., Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) 'Jbrowse: A Dynamic Web Platform for Genome Visualization and Analysis'. Genome Biology 17 66.

Burns, K. N., Kluepfel, D. a., Strauss, S. L., Bokulich, N. a., Cantu, D., & Steenwerth, K. L. (2015). Vineyard soil bacterial diversity and composition revealed by 16S rRNA

genes: Differentiation by geographic features. Soil Biology and Biochemistry, 91, 232–247. https://doi.org/10.1016/j.soilbio.2015.09.002

Burns, K. N., Bokulich, N. A., Cantu, D., Greenhut, R. F., Kluepfel, D. A., O'Geen, A. T., … Steenwerth, K. L. (2016). Vineyard soil bacterial diversity and composition revealed by 16S rRNA genes: Differentiation by vineyard management. Soil Biology and Biochemistry, 103, 337–348. https://doi.org/10.1016/j.soilbio.2016.09.007

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and Applications. BMC Bioinformatics 10: 421.

Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, Choisne N, Mohellibi N, Guichard C, Rombauts S, Le Clainche I, Bérard A, Chauveau A, Bounon R, Rustenholz C, Morgante M, Le Paslier MC, Brunel D, Adam-Blondon AF (2017) A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). Genom Data 14:56-62.

Cardone M. F., D'Addabbo P., Alkan C., Bergamini C., Catacchio C. R., Anaclerio F., et al. (2016). Inter-varietal structural variation in grapevine genomes. Plant J 88 648–661.

Carmichael I, Marron JS (2018) Data Science vs. Statistics: two cultures? Japanese Journal of Statistics and Data Science. 1:1, 117–138.

Cavallini E, Matus JT, Finezzo L, Zenoni S, Loyola R, Guzzo F, Schlechter R, Ageorges A, Arce-Johnson P, Tornielli GB (2015) The phenylpropanoid pathway is controlled at different branches by a set of R2R3-MYB C2 repressors in grapevine. Plant Physiol 167(4):1448-70.

Chawade A, Alexandersson E, Levander F (2014) Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets, J Proteome Res 13: 3114 – 3120.

Chen X, Bhadauria V, Ma B (2017) ChIP-Seq: A Powerful Tool for Studying Protein-DNA Interactions in Plants. Curr Issues Mol Biol 27: 171 – 180.

Cheng C, Jiao C, Singer SD, Gao M, Xu X, Zhou Y, et al. (2015) Gibberellin-induced changes in the transcriptome of grapevine (*Vitis labrusca* × *V. vinifera*) cv. Kyoho flowers. BMC Genomics 16:128.

Cheng MC, Liao PM, Kuo WW, Lin TP (2013) The Arabidopsis ETHYLENE RESPONSE FACTOR1 regulates abiotic stress-responsive gene expression by binding to different cis-acting elements in response to different stress signals. Plant Physiol. 162(3):1566-82.

Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods.13(12):1050-1054.

Cramer G, Ergul A, Grimplet J, Tillett R, Tattersall E, et al. (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7: 111–134.

Culhane AC, Thioulouse J, Perrière G, Higgins DG (2005) MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 21: 2789 – 2790.

Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L, et al. (2014). The high polyphenol content of grapevine cultivar tannat berries is conferred primarily by genes that are not shared with the reference genome. Plant Cell 25: 4777–4788.

Da Silva FG, Iandolino A, Al-Kayal F, Bohlmann MC, Cushman MA, et al. (2005) Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple Vitis species and development of a compendium of gene expression during berry development. Plant Physiology 139: 574–597.

Dal Santo S, Tornielli GB, Zenoni S, Fasoli M, Farina L, Anesi A, et al. (2013) The plasticity of the grapevine berry transcriptome. Genome Biol. 14, r54.

Dalio RJD, Herlihy J, Oliveira TS, McDowell JM, Machado M (2018) Effector Biology in Focus: A Primer for Computational Prediction and Functional Characterization. Mol Plant Microbe Interact 31(1):22-33.

Defoort J, Van de Peer Y, Vermeirssen V. (2018) Function, dynamics and evolution of network motif modules in integrated gene regulatory networks of worm and plant. Nucleic Acids Res 10.1093/nar/gky468.

Deluc LG, Grimplet J, Wheatley MD, Tillett RL, Quilici DR, et al. (2007) Transcriptomic and metabolite analyses of cabernet sauvignon grape berry development. BMC Genomics 8: 429.

Di Genova A., Almeida A. M., Muñoz-Espinoza C., Vizoso P., Travisany D., Moraga C., et al. (2014) Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. BMC Plant Biol. 14:7 10.1186/1471-2229-14-7.

Domingos S, Fino J, Paulo OS, Oliveira CM, Goulao LF (2016) Molecular candidates for early-stage flower-to-fruit transition in stenospermocarpic table grape (Vitis vinifera L.) inflorescences ascribed by differential transcriptome and metabolome profiles. Plant Science 244:40-56.

du Plessis K, Young PR, Eyéghé-Bickong HA, Vivier MA (2017) The Transcriptional Responses and Metabolic Consequences of Acclimation to Elevated Light Exposure in Grapevine Berries. Front Plant Sci. 8:1261.

Duchêne, E., Butterlin, G., Claudel, P., Dumas, V., Jaegli, N., Hugueney, P., Arnold, G. and Merdinoglu, D (2017) Genetic determinism of the'Muscat' flavour in grapevine (Vitis vinifera L.) cultivars. Acta Hortic 1157, 87-92.

Duvick, Jon et al. (2008) “PlantGDB: A Resource for Comparative Plant Genomics.” Nucleic Acids Research 36.Database issue: D959–D965.

Eid J, Fehr A, Gray J et al. (2009) Real-time DNA sequencing from single polymerase molecules. Science 323(5910):133–138.

Engelen K, Fu Q, Meysman P, Sánchez-Rodríguez A, De Smet R, Lemmens K, Fierro AC, Marchal K. (2011). COLOMBOS: Access Port for Cross-Platform Bacterial Expression Compendia. PLOS ONE 6 (7): e20938.

Fabres PJ, Collins C, Cavagnaro TR, Rodríguez López CM (2017) A Concise Review on Multi-Omics Data Integration for Terroir Analysis in Vitis vinifera. Front Plant Sci. 8:1065.

Fasoli M, Dal Santo S, Zenoni S, Tornielli GB, Farina L, Zamboni A, Porceddu A, Venturini L, Bicego M, Murino V, Ferrarini A, Delledonne M, Pezzotti M (2012) The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24(9):3489-505.

Fortes AM, Gallusci P (2017) Plant Stress Responses and Phenotypic Plasticity in the Epigenomics Era: Perspectives on the Grapevine Scenario, a Model for Perennial Crop Plants. Front Plant Sci 8:82.

Furió-Tarí P, Consea A, Tarazona S (2016) RGmatch: matching genomic regions to proximal genes in omics data integration. BMC Bioinformatics 17, 427.

Garrett-Mayer E, Parmigiani G, Zhong X, Cope L, Gabrielson E. (2008). Cross-Study Validation and Combined Analysis of Gene Expression Microarray Data. *Biostatistics (Oxford, England)* 9 (2): 333–54.

Ghan R, Petereit J, Tillett RL, Schlauch KA, Toubiana D, Fait A, Cramer GR (2017) The common transcriptional subnetworks of the grape berry skin in the late stages of ripening. BMC Plant Biol 17(1):94.

Gligorijević V, Nataša P. (2015). Methods for Biological Data Integration: Perspectives and Challenges. Journal of The Royal Society Interface 12 (112): 20150571.

González I, Déjean S, Martin P, Baccini A. (2008) CCA: An R Package to Extend Canonical Correlation Analysis. Journal of Statistical Software 23, 12.

González I, Cao KA, Davis MJ, Déjean S (2012) Visualising associations between paired 'omic' data sets. BioData Mining, 5:19, 1–23.

Grenville-Briggs LJ, van West P (2005) The biotrophic stages of oomycete-plant interactions. Adv Appl Microbiol 57:217-43.

Grimplet J, Cramer GR, Dickerson JA, Mathiason K, Van Hemert J, Fennell AY (2009a) VitisNet: "Omics" integration through grapevine molecular networks. PLoS One 4(12):e8365.

Grimplet J, Deluc LG, Tillett RL, Wheatley MD, Schlauch KA, et al. (2007) Tissue-specific mRNA expression profiling in grape berry tissues. BMC Genomics 8: 187.

Grimplet J, Wheatley MD, Jouira HB, Deluc LG, Cramer GR, Cushman JC (2009b) Proteomic and selected metabolite analysis of grape berry tissues under well-watered and water-deficit stress conditions. Proteomics 9(9):2503-28.

Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430(6995):88-93.

Harris ZN, Kovacs LG, Londo JP (2017) RNA-seq-based genome annotation and identification of long-noncoding RNAs in the grapevine cultivar 'Riesling'. BMC Genomics 18(1):937.

Hichri I, Heppel SC, Pillet J, Léon C, Czemmel S, Delrot S, Lauvergeat V, Bogs J (2010) The basic helix-loop-helix transcription factor MYC1 is involved in the

regulation of the flavonoid biosynthesis pathway in grapevine. Mol Plant 3(3):509-23.

Höll J, Vannozzi A, Czemmel S, D'Onofrio C, Walker AR, Rausch T, Lucchin M, Boss PK, Dry IB, Bogs J (2013) The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in Vitis vinifera. Plant Cell 25(10):4135-49.

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467.

Jellouli N, Jouira BH, Skouri H, Ghorbel A, Gourgouri A, et al. (2008) Proteomic analysis of Tunisian grapevine cultivar Razegui under salt stress. J Plant Physiol 165: 471–481.

Kambiranda D, Basha SM, Singh R, Snowden J, Mercer R (2018) Proteome Profile of American Hybrid Grape cv. Blanc du Bois during Ripening Reveals Proteins Associated with Flavor Volatiles and Ethylene Production. Proteomics 18(8):e1700305.

Khater F, Fournand D, Vialet S, Meudec E, Cheynier V, Terrier N (2012) Identification and functional characterization of cDNAs coding for hydroxybenzoate/hydroxycinnamate glucosyltransferases co-expressed with genes related to proanthocyanidin biosynthesis. J Exp Bot 63(3):1201-14.

Lê S, Josse J, Husson, F (2008) FactoMineR: an R package for multivariate analysis, Journal of Statistical Software 25, 1.

Lee T, Yang S, Kim E, Ko Y, Hwang S, Shin J, Shim JE, Shim H, Kim H, Kim C, Lee I (2015) 'Aranet V2: An Improved Database of Co-Functional Gene Networks for the Study of Arabidopsis Thaliana and 27 Other Nonmodel Plant Species', Nucleic Acids Res 43, D996-1002.

Leveau JH, Tech JJ (2010) Grapevine microbiomics: bacterial diversity on grape leaves and berries revealed by high-throughput sequence analysis of 16S rRNA amplicons. International Symposium on Biological Control of Postharvest Diseases: Challenges and Opportunities 905 31–42.

Li Y, Pearl SA, Jackson SA (2015) Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis. Trends Plant Sci 20, 664 – 675.

Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C, Nick P, Du F, Fan P, Mao R, Zhu Y, Deng W, Yang M, Huang H, Liu Y, Ding Y, Liu X, Jiang J, Zhu Y, Li S, He X, Chen W, Dong Y. (2019). Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses. Nat Commun. 10(1):1190.

Liu G, Wang J, Cramer G, Dai Z, Duan W, Xu H, et al. (2012) Transcriptomic analysis of grape (Vitis viniferaL.) leaves during and after recovery from heat stress. BMC Plant Biology 12:174.

Loyola R, Herrera D, Mas A, Wong DCJ, Höll J, Cavallini E, et al. (2016) The photomorphogenic factors UV-B RECEPTOR 1, ELONGATED HYPOCOTYL 5, and HY5 HOMOLOGUE are part of the UV-B signalling pathway in grapevine and mediate flavonol accumulation in response to the environment. J. Exp. Bot 67(18):5429-5445.

Lucker J, Laszczak M, Smith D, Lund ST (2009) Generation of a predicted protein database from EST data and application to I-TRAQ analyses in grape (Vitis vinifera cv. Cabernet Sauvignon) berries at ripening initiation. BMC Genomics 10: 50.

Luo J, Schumacher M, Scherer A, Sanoudou D, Megherbi D, Davison T, Shi T, et al. (2010). A Comparison of Batch Effect Removal Methods for Enhancement of Prediction Performance Using MAQC-II Microarray Gene Expression Data. The Pharmacogenomics Journal 10 (4): 278–91.

Malacarne, G., Coller, E., Czemmel, S., Vrhovsek, U., Engelen, K., Goremykin, V., et al. (2016) The grapevine VvibZIPC22 transcription factor is involved in the regulation of flavonoid biosynthesis. J. Exp. Bot. 67, 3509-3522.

Margulies M, Egholm M, Altman WE et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380.

Martin LBB, Zhangjun F, Giovannoni JJ, Rose JKC (2013) Catalyzing plant science research with RNA-seq. Front Plant Sci, 4, 66.

Martins G, Lauga B, Miot-Sertier C, Mercier A, Lonvaud A, Soulas ML, et al. (2013) Characterization of epiphytic bacterial communities from grapes, leaves, bark and soil of grapevine plants grown, and their relations. Plos One 8(8):e73013.

Matus JT, Cavallini E, Loyola R, Höll J, Finezzo L, Dal Santo S, Vialet S, Commisso M, Roman F, Schubert A, Alcalde JA, Bogs J, Ageorges A, Tornielli GB, Arce-Johnson P. (2017) A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus. Plant J 91(2):220-236.

Matus JT (2016) Transcriptomic and Metabolomic Networks in the Grape Berry Illustrate That it Takes More Than Flavonoids to Fight Against Ultraviolet Radiation. Front Plant Sci 7:1337.

Mercenaro, Luca et al. (2017) Sequence Polymorphisms and Structural Variations among Four Grapevine (Vitis Vinifera L.) Cultivars Representing Sardinian Agriculture. Frontiers in Plant Science 8 1279.

Mesarovic MD (1968) Systems theory and biology - view of a theoretician. In Systems Theory and Biology, Mesarovic MD (ed) pp 59–87. Springer: Verlag New York, USA.

Mevik B, Wehrens R (2007) The pls Package: Principal Component and Partial Least Squares Regression in R, Journal of Statistical Software, 18, 2.

Miller JG (1978) Living Systems. Mcgraw-Hill, New York, USA. 1102 p.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824-7.

Milos P (2008) Helicos BioSciences. Pharmacogenomics 9:477–480.

Minio A, Massonnet M, Figueroa-Balderas R, Castro A, Cantu D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. G3 (Bethesda). pii: g3.400030.2019.

Minio A, Lin J, Gaut BS, Cantu D. (2017). How Single Molecule Real-Time Sequencing and Haplotype Phasing Have Enabled Reference-Grade Diploid Genome Assembly of Wine Grapes. Front Plant Sci. 8:826.

Moretto, Marco, Paolo Sonego, Nicolas Dierckxsens, Matteo Brilli, Luca Bianco, Daniela Ledezma-Tejeida, Socorro Gama-Castro, et al. (2016a). COLOMBOS v3.0: Leveraging Gene Expression Compendia for Cross-Species Analyses. Nucleic Acids Research 44 (D1): D620-623.

Moretto, M., Sonego, P., Pilati, S., Malacarne, G., Costantini, L., Grzeskowiak, L., et al. (2016b) VESPUCCI: Exploring patterns of gene expression in grapevine. Front. Plant Sci. 7, 633.

Moretto, Marco, Paolo Sonego, Ana B. Villaseñor-Altamirano, Kristof Engelen. (2019). First step toward gene expression data integration: transcriptomic data acquisition with COMMAND>. BMC Bioinformatics 20 (1): 54.

Moser C, Segala C, Fontana P, Salakhudtinov I, Gatto P, et al. (2005) Comparative analysis of expressed sequence tags from different organs of Vitis vinifera L. Funct Integr Genomics 5: 208–217.

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci 5:170.

Negri AS, Prinsi B, Failla O, Scienza A, Espen L (2015) Proteomic and metabolic traits of grape exocarp to explain different anthocyanin concentrations of the cultivars. Front Plant Sci. 6:603.

Negri S, Lovato A, Boscaini F, Salvetti E, Torriani S, Commisso M, Danzi R, Ugliano M, Polverari A, Tornielli GB, Guzzo F (2017) The Induction of Noble Rot (Botrytis cinerea) Infection during Postharvest Withering Changes the Metabolome of Grapevine Berries (Vitis vinifera L., cv. Garganega). Front Plant Sci 8:1002.

Obayashi T, Aoki Y, Tadaka S, Kagaya Y, Kinoshita K (2018) ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. Plant Cell Physiol 59(1):e3.

Obayashi T, Nishida K, Kasahara K, Kinoshita K (2011) ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. Plant Cell Physiol 52, 213-219.

Ohyanagi H, Takano T, Terashima S, Kobayashi M, Kanno M, Morimoto K, Kanegae H, Sasaki Y, Saito M, Asano S, Ozaki S, Kudo T, Yokoyama K, Aya K, Suwabe K, Suzuki G, Aoki K, Kubo Y, Watanabe M, Matsuoka M, Yano K (2015) Plant Omics Data Center: An Integrated Web Repository for Interspecies Gene Expression Networks with Nlp-Based Curation. Plant Cell Physiol 56, e9.

Palumbo MC, Zenoni S, Fasoli M, Massonnet M, Farina L, Castiglione F, et al. (2014) Integrated network analysis identifies fight-club nodes as a class of hubs encompassing key putative switch genes that induce major transcriptome reprogramming during grapevine development. Plant Cell 4617-4635.

Papatheodorou I, Oellrich A, Smedley D (2015). Linking gene expression to phenotypes via pathway information. Journal of Biomedical Semantics 6, 17.

Pass DA, Sornay E, Marchbank A, Crawford MR, Paszkiewicz K, Kent NA, Murray JAH (2017) Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in Arabidopsis. PLoS Genet 13, e1006988.

Pii Y, Zamboni A, Dal Santo S, Pezzotti M, Varanini Z, Pandolfini T (2017) Prospect on Ionomic Signatures for the Classification of Grapevine Berries According to Their Geographical Origin. Front Plant Sci. 8:640.

Pinasseau L, Vallverdú-Queralt A, Verbaere A, Roques M, Meudec E, Le Cunff L, Péros JP, Ageorges A, Sommerer N, Boulet JC, Terrier N, Cheynier V (2017) Cultivar Diversity of Grape Skin Polyphenol Composition and Changes in Response to Drought Investigated by LC-MS Based Metabolomics. Front Plant Sci. 8:1826.

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Lovato A, Pezzotti M, Delledonne M, Polverari A (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (Vitis vinifera) and a resistant (V. riparia) grapevine species. BMC Genomics 11:117.

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Polverari A (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (Vitis vinifera) and a resistant (V. riparia) grapevine species. BMC Genomics, 11, 117.

Proost S, Mutwil M (2016) Tools of the trade: Studying molecular networks in plants. Curr. Opin. Plant Biol. 30, 130-140.

Ramalingam A, Kudapa H, Pazhamala LT, Weckwerth W, Varshney RK (2015) Proteomics and Metabolomics: Two Emerging Areas for Legume Improvement, Front Plant Sci 6, 1116.

Roach MJ, Johnson DL, Bohlmann J, van Vuuren HJJ, Jones SJM, Pretorius IS, Schmidt SA, Borneman AR. (2018). Population sequencing reveals clonal diversity

and ancestral inbreeding in the grapevine cultivar Chardonnay. PLoS Genet. 14(11):e1007807.

Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: An R package for omics feature selection and multiple data integration. PLos Comput Biol 13, e1005752.

Rothberg JM, Hinz W, Rearick TM et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356):348–352.

Rung, Johan, Alvis Brazma. (2013). Reuse of Public Genome-Wide Gene Expression Data. Nature Reviews. Genetics 14 (2): 89–99.

Salmon-Divon M, Dvinge H, Tommoja K, Bertone P (2010) PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics 11, 415.

Savoi S, Wong DC, Arapitsas P, Miculan M, Bucchetti B, Peterlunger E, Fait A, Mattivi F, Castellarin SD (2016) Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (Vitis vinifera L.). BMC Plant Biol 16:67.

Savoi S, Wong DCJ, Degu A, Herrera JC, Bucchetti B, Peterlunger E, Fait A, Mattivi F, Castellarin SD (2017) Multi-Omics and Integrated Network Analyses Reveal New Insights into the Systems Relationships between Metabolites, Structural Genes, and Transcriptional Regulators in Developing Grape Berries (Vitis vinifera L.) Exposed to Water Deficit. Front Plant Sci 8:1124.

Savoi, S., Wong, D. C. J., Arapitsas, P., Miculan, M., Bucchetti, B., Peterlunger, E., et al. (2016) Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (Vitis vinifera L.). BMC Plant Biol 16, 67.

Serin EA, Nijveen H, Hilhorst HW, Ligterink W (2016) Learning from Co-expression Networks: Possibilities and Challenges. Front Plant Sci 7:444.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498-504.

Sheth BP, Thaker VS (2014) Plant Systems Biology: Insights, Advances and Challenges. Planta 240, 33-54.

Soubeyrand E, Colombié S, Beauvoit B, Dai Z, Cluzet S, Hilbert G, Renaud C, Maneta-Peyret L, Dieuaide-Noubhani M, Mérillon JM, Gibon Y, Delrot S, Gomès E (2018) Constraint-Based Modeling Highlights Cell Energy, Redox Status and α-Ketoglutarate Availability as Metabolic Drivers for Anthocyanin Accumulation in Grape Cells Under Nitrogen Limitation. Front Plant Sci 9:421.

Sullivan AM, Kerry LB, Sandstrom R, Stamatoyannpoulos JA, Queitsch C (2015) DNase I hypersensitivity mapping, genomic footprinting, and transcription factor networks in plants. Current Plant Biology 3–4: 40-47.

Sweetman C, Wong DC, Ford CM, Drew DP (2012) Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13:691.

Tattersall E, Grimplet J, Deluc L, Wheatley M, Vincent D, et al. (2007) Transcript abundance profiles reveal larger and more complex responses of grapevine to chilling compared to osmotic and salinity stress. Funct Integr Genomics 7: 317–333.

Terrier N, Glissant D, Grimplet J, Barrieu F, Abbal P, et al. (2005) Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (Vitis vinifera L.) development. Planta 222: 832–847.

Tillett RL, Ergül A, Albion RL, Schlauch KA, Cramer GR, Cushman JC (2011) Identification of tissue-specific, abiotic stress-responsive gene expression patterns in wine grape (Vitis vinifera L.) based on curation and mining of large-scale EST data sets. BMC Plant Biology 11:86.

Vannozzi A, Wong DCJ, Höll J, Hmmam I, Matus JT, Bogs J, Ziegler T, Dry I, Barcaccia G, Lucchin M (2018) Combinatorial Regulation of Stilbene Synthase Genes by WRKY and MYB Transcription Factors in Grapevine (Vitis vinifera L.). Plant Cell Physiol 59(5):1043-1059.

Vega A, Gutiérrez RA, Peña-Neira A, Cramer GR, Arce-Johnson P (2011) Compatible GLRaV-3 viral infections affect berry ripening decreasing sugar accumulation and anthocyanin biosynthesis in Vitis vinifera. Plant Mol Biol 77(3):261-74.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2: 1326.

Vincent D, Ergul A, Bohlman MC, Tattersall EA, Tillett RL, et al. (2007) Proteomic analysis reveals differences between Vitis vinifera L. cv. Chardonnay and cv. Cabernet Sauvignon and their responses to water deficit and salinity. J Exp Bot 58: 1873–1892.

Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M, Vannozzi A, Bonghi C, Lucchin M, Valle G (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol 14:99.

Von Bertalanffy L (1968) General System Theory: Foundations, Development, Applications. George Braziller (ed), New York, USA. 295 p.

Vondras AM, Commisso M, Guzzo F, Deluc LG (2017) Metabolite Profiling Reveals Developmental Inequalities in Pinot Noir Berry Tissues Late in Ripening. Front Plant Sci 8:1108.

Wang L, Sun X, Weiszmann J, Weckwerth W (2017) System-Level and Granger Network Analysis of Integrated Proteomic and Metabolomic Dynamics Identifies

Key Points of Grape Berry Development at the Interface of Primary and Secondary Metabolism. Front Plant Sci 8:1066.

Wang L, Sun X, Weiszmann J, Weckwerth W (2017) System-level and granger network analysis of integrated proteomic and metabolomic dynamics identifies key points of grape berry development at the interface of primary and secondary metabolism. Front. Plant Sci 8:1066.

Waters DL, Holton TA, Ablett EM, Lee LS, Henry RJ (2005) cDNA microarray analysis of developing grape (Vitis vinifera cv. Shiraz) berry skin. Funct Integr Genomics 5: 40–58.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018.

Winter G, Krömer JO (2013) Fluxomics – connecting 'omics analysis and phenotypes". Environmental Microbiology 15 (7): 1901–1916.

Wolfe CJ, Kohane IS, Butte AJ (2005) Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics 6, 227.

Wong DCJ, Matus JT (2017) Constructing Integrated Networks for Identifying New Secondary Metabolic Pathway Regulators in Grapevine: Recent Applications and Future Opportunities. Front Plant Sci 8:505.

Wong DCJ, Zhang L, Merlin I, Castellarin SD, Gambetta GA (2018) Structure and transcriptional regulation of the major intrinsic protein gene family in grapevine. BMC Genomics 19(1):248.

Wong DCJ, Lopez-Gutierrez R, Gambetta GA, Castellarin SD (2017) Genome-wide analysis of cis-regulatory element structure and discovery of motif-driven gene co-expression networks in grapevine. DNA Res 24(3):311-326.

Wong DCJ, Schlechter R, Vannozzi A, Höll J, Hmmam I, Bogs J et al. (2016) A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. DNA Res 23, 451-466.

Wong DCJ, Sweetman C, Drew DP, Ford CM (2013) VTCdb: a gene co-expression database for the crop species Vitis vinifera (grapevine). BMC Genomics 14, 882.

Xie H, Konate M, Sai N, Tesfamicael KG, Cavagnaro T, Gilliham M, Breen J, Metcalfe A, Stephen JR, De Bei R, Collins C, Lopez CMR (2017) Global DNA Methylation Patterns Can Play a Role in Defining Terroir in Grapevine (Vitis vinifera cv. Shiraz). Front Plant Sci 8:1860.

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart N (2008) Plant system biology comes of age. Trends in Plant Science 13, 4, 165-171.

Zamboni A, Di Carli M, Guzzo F, Stocchero M, Zenoni S, Ferrarini A et al. (2010) Identification of putative stage-specific grapevine berry biomarkers and omics data integration into networks. Plant Physiol 154, 1439-59.

Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, et al. (2015) The soil microbiome influences grapevine-associated microbiota. MBio 6(2):e02527–14.

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G et al. (2010) Characterization of transcriptional complexity during berry development in Vitis vinifera using RNA-Seq. Plant Physiology 152(4):1787–95.

Zhang G, Chen D, Zhang T, Duan A, Zhang J, He C (2018) Transcriptomic and functional analyses unveil the role of long non-coding RNAs in anthocyanin biosynthesis during sea buckthorn fruit ripening. DNA Res. doi: 10.1093/dnares/dsy017.

Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. (2017). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. Proc Natl Acad Sci USA. 114(44):11715-11720.

Tables and Figures

Table 8.1. Number of SRA experiments (No. of SRA) and Gbp of data produced (Gbp of data) for grapevine cultivars according to the type of the library source (genomic or transcriptomic).

	GENOMIC		TRANSCRIPTOMIC	
CULTIVAR	**No of SRA**	**Gbp of data**	**No of SRA**	**Gbp of data**
Cabernet Sauvignon	6	166.59	393	805.27
Barossa Shiraz	197	68.22		
Pinot noir	15	44.16	115	341.59
Chardonnay	95	2544.81	34	48.67
Merlot	2	0.00	74	277.67
Carmenere	63	147.27		
Muscat table			54	508.32
Pinot Meunier	4	31.89	48	137.02
Thompson Seedless	3	10.63	49	174.05
Sangiovese	3	0.01	47	61.49
Sauvignon blanc	2	0.01	35	199.50
Tempranillo			36	143.73
Riesling	2	34.24	31	51.64
Cabernet Franc	2	0.01	28	85.89
Tocai friulano			30	35.50
Barbera	1	0.01	19	47.36
Kyoho			20	176.67
Semillon	3	8.61	16	40.41
Vermentino	4	12.84	12	39.80
Gaglioppo			15	50.45
Garganega	2	0.01	12	38.85
Primitivo di Manduria	2	18.16	12	42.64
Tannat	2	65.20	11	79.96
Carignan			12	39.14
Glera			12	42.68
Koshu			12	31.01
Moscatel Galego			12	50.10
Moscato bianco			12	43.17
Muscat Hamburg	3	0.96	9	13.42
All other cultivars	1224	4557.52	1344	4340.34
ND	958	3905.94	1269	3983.17

ND: information not available in SRA archive

Table 8.2. Online resources useful for gene network mining in grapevine. For multi-species DB, only grapevine-specific features are highlighted. GRN = Genome-wide transcriptional regulatory interaction network. Vis.: visualization

DB name	Type	Species	Datatypes	Features	Query examples:	Webcite
ATTED-II	GCN	multi-species (9)	Co-expression (Microarray, RNA-seq)	Grape GCN were constructed using RNA-seq and Nimblegen arrays. Similarity metric = MR	1. 'Guide' gene lists. 2. Comparative analysis of CEG rankings across multiple species.	http://atted.jp/
CoNekT	GCN	multi-species (7)	Co-expression (RNA-seq)	Grape GCN were constructed using RNA-seq. Similarity metric = HRR.	1. 'Guide' gene lists. 2. Comparative analysis of CEG rankings across multiple species.	http://conekt.mpimp-golm.mpg.de/pub/
AraNet/ AraNetv2	Integrated (CFN)	multi-species (29)	19 datatypes (e.g. co-expression, domain co-occurrence, genomic neighborhood of orthologs, protein-protein interactions, phylogenetic profile).	Grapevine gene function were inferred using any combination of the associated datatypes. Some include orthology-based projections from model plant species (i.e. Arabidopsis).	1. 'Guide' gene lists.	http://www.inetbio.org/aranet
PODC	Integrated (CFN)	multi-species (11)	Co-expression (RNA-seq), natural language processing-based curation	Grape GCN were constructed using RNA-seq. Similarity metric = PCC and Distance in Correspondence Analysis (DCA)	1. 'Guide' gene lists.	http://plantomics.mind.meiji.ac.jp/podc/
COP	GCN	multi-species (8)	Co-expression (Microarray)	Grape GCN were constructed using Affymetrix microarray data. Similarity metric = Cosine correlation (CC). Not recommended for grapevine, but fine for Arabidopsis.	1. 'Guide' gene lists.	http://webs2.kazusa.or.jp/kagiana/cop0911/
PLANEX	GCN	multi-species (8)	Co-expression (Microarray)	Grape GCN were constructed using Affymetrix microarray data. Similarity metric = PCC. Not recommended for grapevine, but fine for Arabidopsis.	1. 'Guide' gene lists.	http://planex.plantbioinformatics.org/
ePlant	Vis.	Grape	Gene expression	Interactive grapevine gene atlas expression browser.	1. 'Guide' gene lists.	http://bar.utoronto.ca/efp_grape/cgi-bin/efpWeb.cgi

PlantReg Map	GRN	multi-species (132 species)	CHIP-seq, DAP-seq, PBM, literature curation	**Grapevine TF binding sites were inferred using orthology-based projections from model plant species (i.e. Arabidopsis). Genome-wide TFBS analysis of grapevine promoters.**	1. 'Guide' gene lists to query downstream target genes of input gene (i.e. TFs) 2. 'Guide' gene lists to query upstream regulators (TFs) of input genes	http://plantregmap.cbi.pku.edu.cn/network.php
VitisNet	Vis.	Grape	Manually-curated molecular networks encompassing 247 distinct biological processes.	**Allows the visualization of multi-omics datasets (i.e. genes, proteins or metabolites) simultaneously on these molecular networks.**	1. Downloaded networks can be imported into Cytoscape for further multi-omics datasets visualization	https://www.sdstate.edu/vitisnet-molecular-networks-grapevine
STRING	Integrated (CFN)	multi-species (2,031 species plants & animals)	8 datatypes (e.g. gene neighborhood, gene co-occurrence, textmining, co-expression, protein homology)	**Grapevine gene function were inferred using any combination of the associated datatypes. Some include orthology-based projections from model plant and non-plant species. Similar to AraNet.**	1. 'Guide' gene lists.	https://string-db.org/
VTCdb	GCN	Grape	Co-expression (Microarray, RNA-seq)	Grape GCN were constructed using RNA-seq and Nimblegen arrays. Similarity metric = MR, HRR, PCC.	1. 'Guide' gene lists. 2. Biological processes of interests	http://vtcdb.adelaide.edu.au/Home.aspx
Vespucci	GCN	Grape	Co-expression (Microarray, RNA-seq)	Grape GCN were constructed using RNA-seq and multiple microarray platforms. Similarity metric = PCC. Includes an exploratory tool to analyze expression of genes across 1,608 manually-curated (vocabulary-controlled) experimental conditions.	1. 'Guide' gene lists.	http://vespucci.colombos.fmach.it/
grape_sRNA_atlas	miRNA	Grape	miRNA (RNA-seq)	Grape miRNA-target (gene) networks were constructed using a comprehensive miRNA catalogue (both known and novel) and *in silico* target prediction analysis. miRNA expression browser available.	1. miRNA query	https://mpss.danforthcenter.org/dbs/index.php?SITE=grape_sRNA_atlas
BIOWINE	miRNA	Grape	miRNA (RNA-seq)	Grape miRNA-target (gene) networks were constructed using *in silico* target prediction analysis.	1. miRNA query 2. Biological processes of interests	https://alpha.dmi.unict.it/biowine/

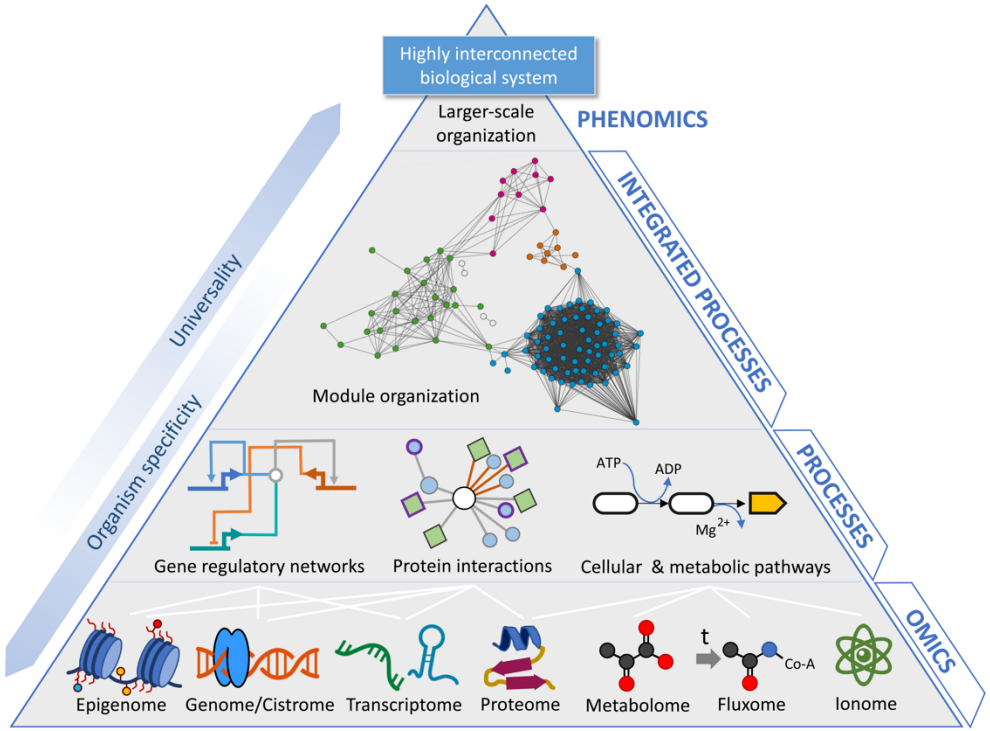

Figure 8.1. The Oltvai and Barabási's pyramid of life reviewed by systems biology approaches. The complexity of a biological system can be represented by several layers of functional organization. Starting from the cell's building blocks; the life biomolecules, these are responsible for the genetic information to be stored, processed and finally executed in several developmental programs or in response to the environment. Genes and their epigenetic marks, transcripts, proteins and their modifications, metabolites and their fluxes and even ions can be collectively characterized and quantified through omics. The huge amount of data acquired from these technologies can only be handled with intensive bioinformatics. At the second level, biomolecules form gene-regulatory and protein-interacting motifs and subcellular signaling /metabolic pathways, all of them with the inherent capacity of impacting each other. As these biological processes are tightly connected (e.g. a set of genes, proteins and metabolites being activated in response to a pathogen) they are organized in functional modules. Complex biological processes can be studied from a 'multi-omics' perspective thanks to the recent improvements in genome-wide techniques and systems biology methods. Modules

can be studied by integrative systems biology tools but can be further organized in higher hierarchical multidimensional structures. Larger-scale modules are also dynamic in time and translate into phenotypes. In recent efforts modeling algorithms have been applied to largely annotate phenotypes (i.e. 'phenomics'). Computational biology has supported an adequate data management, efficient data analysis, and user-friendly software applications to study biological systems at each of these levels. Although the individual components are unique to a given organism, the topologic properties of networks are surprisingly similar (Adapted from Oltvai and Barabási, 2002).

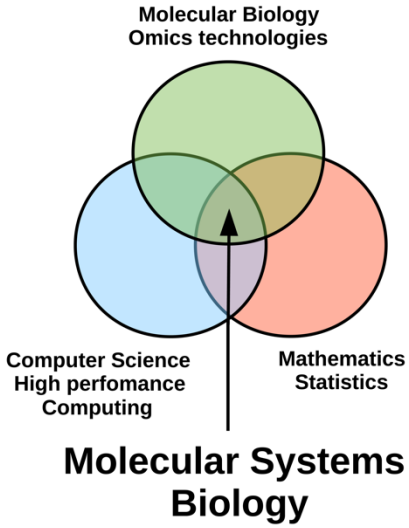

Figure 8.2. Schematic representation of Molecular Systems Biology as a discipline resulting from the overlapping of computational, mathematical and biological explorations.

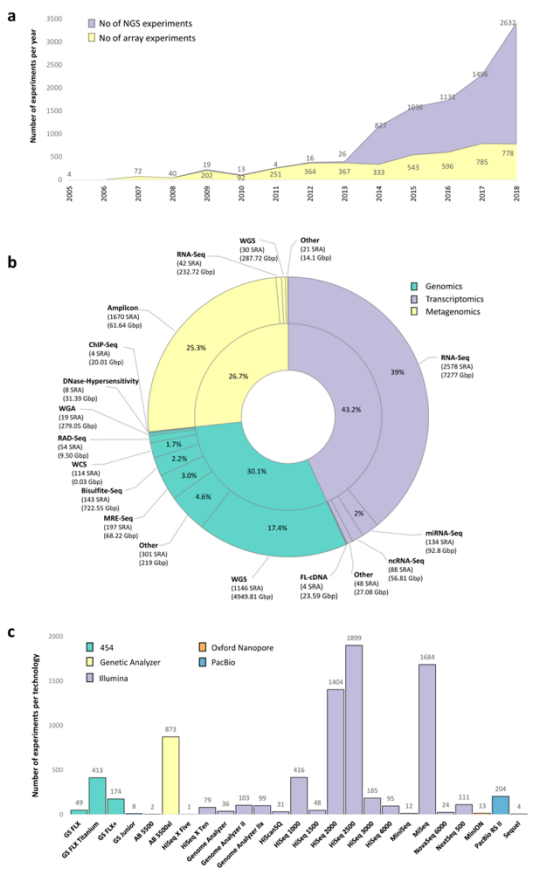

Figure 8.3. Next-generation sequencing and array data available for grapevine. Next-generation sequencing and oligonucleotide array have represented two relevant genome-scale methodologies for grapevine studies. The data presented were retrieved from the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) and Gene Expression Omnibus (GEO) NCBI repositories (https://www.ncbi.nlm.nih.gov/gds/) as of July 2018, by using a keyword search "Vitis" or "Grapevine". **a)** timeline of grapevine experiments performed since 2005 according to the methodology used (*in situ* oligonucleotide array or NGS). **b)** Number and distribution of grapevine experiments from high-throughput sequencing

technologies. The inner circle represents the distribution according to the library layer (Genomics, Transcriptomics, Metagenomics) while the outer circle is according to the library strategy used (e.g. RNA-seq, Chip-seq, etc). For each outer section the number of experiments (SRA) and the Giga base pair of data (Gbp) were also reported. **c)** distribution of the NGS platforms used, including Roche 454 GS System, Illumina Genome Analyzer, Applied Biosystems SOLiD System, Helicos Heliscope, Pacific Biosciences SMRT.

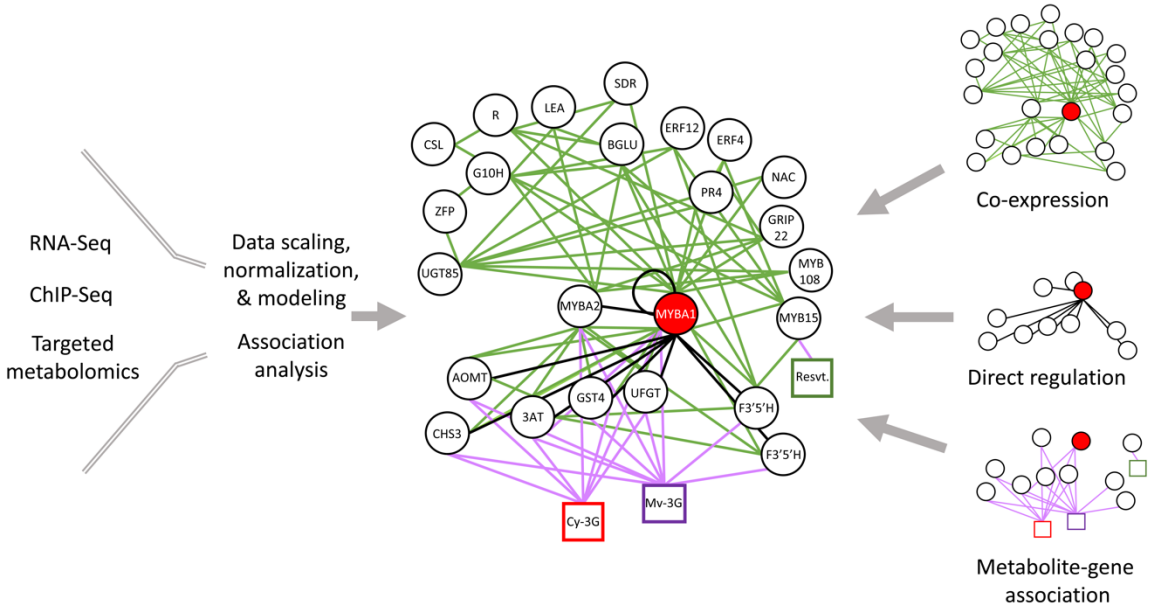

Figure 8.4. Methods for building integrative network models. Different omics technologies generate data with diverging formats (e.g. numerical scales) and therefore are considered as multidimensional. A hypothetical regulatory network for the berry color locus was used to illustrate how gene co-expression, transcription factor binding and metabolic data can be integrated to generate a composite network. These can be generated by applying scaling and normalization algorithms to all omics datasets (at the left) or by superposing independently-produced networks (on the right). The main anthocyanin regulator MYBA1 is centered in the network. Its co-expressed genes were taken from previous gene GCN analyses (Wong

et al. 2016). Direct regulation examples are taken from experimental evidence (e.g. Matus et al., 2017). Cyanidin or malvidin-related derivatives (di or tri-hydroxylated anthocyanins) are represented by ‘Cy-3G’ and ‘Mv-3G’, respectively. Resvt: the stilbene resveratrol.