We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,300
Open access books available

129,000
International authors and editors

155M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 7

Machine Learning and Rule Mining Techniques in the Study of Gene Inactivation and RNA Interference

Saurav Mallik, Ujjwal Maulik, Namrata Tomar, Tapas Bhadra, Anirban Mukhopadhyay and Ayan Mukherji

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.83470

Abstract

RNA interference (RNAi) and gene inactivation are extensively used biological terms in biomedical research. Two categories of small ribonucleic acid (RNA) molecules, *viz*., microRNA (miRNA) and small interfering RNA (siRNA) are central to the RNAi. There are various kinds of algorithms developed related to RNAi and gene silencing. In this book chapter, we provided a comprehensive review of various machine learning and association rule mining algorithms developed to handle different biological problems such as detection of gene signature, biomarker, gene module, potentially disordered protein, differentially methylated region and many more. We also provided a comparative study of different well-known classifiers along with other used methods. In addition, we demonstrated the brief biological information regarding the immense biological challenges for gene activation as well as their advantages, disadvantages and possible therapeutic strategies. Finally, our study helps the bioinformaticians to understand the overall immense idea in different research dimensions including several learning algorithms for the benevolent of the disease discovery.

Keywords: machine learning, association rule mining, RNAi, gene silencing, multi-omics data

1. Introduction

RNAi [1] is an innate biological process in which RNA molecules inhibit gene expression or translation [2] by suppressing targeted mRNA molecules. Since the discovery of RNAi by

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

Andrew Fire and Craig Mello, it has become evident that RNAi has immense potential in suppression of desired genes [3]. The first evidence that double-stranded RNA (dsRNA) could achieve efficient gene silencing through RNAi came from studies on the nematode *Caenorhabditis elegans* [4] and *Drosophila melanogaster* [5], which lead toward understanding the biochemical nature of the RNAi pathway. Two types of small ribonucleic acid (RNA) molecules—microRNA (miRNA) and small interfering RNA (siRNA)—are central to RNAi [6]. To compare two types of elicit RNAi, the siRNA must be fully complementary to its target mRNA, whereas, miRNA only needs to be partially complementary to its target mRNA. In organisms like *C. elegans* and *D. melanogaster*, RNAi can be induced by introducing long dsRNA complementary to the target mRNA to be degraded, however, in mammalian cells and organisms, introducing dsRNA longer than 30 bp activates a potent antiviral response. To solve this limitation, siRNAs are used to induce RNAi in mammalian cells and organisms [7–9].

The discovery of both siRNA and miRNA provides a new therapeutic approach [10, 11] for the treatment of diseases by targeting genes that have undesired mutated or overexpression of normal genes. The RNAi Process is as following. SiRNAs that induce the degradation of specific endogenous is very common phenomenon in eukaryotic cells to inhibit protein production at post transcriptional level [12]. The RNAi process is initiated by short dsRNAs, 21–25 nucleotides that lead to the sequence specific inhibition of their homologous mRNAs. These siRNAs are normally produced in cells from cleavage of longer dsRNA precursors by Dicer that is a ribonuclease III family member. The cleaved parts are incorporated into a multicomponent nuclease complex known as the RNA-induced silencing complexes (RISC), which contain the splicing protein Argonaute-2 (Ago-2) [13]. The ssRNA derived from the short dsRNA acts as a antisense strand directing the complex to the specific target mRNA; in where a RISC-associated endoribonuclease cleavages the target mRNA [14]. Therapeutic approaches based on siRNA involve the introduction of a synthetic siRNA into the target cells to elicit RNAi, thereby inhibiting the expression of a specific messenger RNA (mRNA) to produce a gene silencing effect [15]. RNAi is beneficial in accelerating cures in medicine, especially when a disease is thought to due to a defective gene [16]. For historical perspective, the first application of RNAi therapy was in age-related macular degeneration (AMD) by using siRNAs to suppress the vascular endothelial growth factor (VEGF) pathway that causes abnormal growth of blood vessels behind the retina, carried out directly to the patient's eye [17]. RNAi techniques have been used against the spread of tumor growth and increasing its sensitization toward drug treatment, RNAi technology will be beneficial to selectively affect cancer cells without damaging normal cells as the RNAi therapy against cancer cells is used for directly targeting the oncogenes; and therefore, found to stop progression and invasion of the tumor cells [18, 19] and also increase the sensitization of tumor against drug, as mentioned earlier [20]. As RNAi can silence disease-associated genes in tissue culture and animal models, the development of RNAi-based reagents for therapeutic applications involves technological enhancements that improve siRNA stability and delivery *in vivo* [21], while minimizing off-target and nonspecific effects.

A number of different approaches have been developed for the *in vivo* delivery of siRNA, among which, rapid infusion by hydrodynamic injection of siRNA achieves the best delivery in rodents [22]. However, this way, the delivery is restricted to highly vascularized tissues,

such as the liver [23] and also, it is currently not a viable method for delivery in human clinical studies. Lipid-based *in vivo* applications have been devised [24], which have been used extensively for cell culture experiments, with some issues, like the cationic nature of the lipids used in cell culture leads to aggregation when used in animals and results in rapid serum clearance and lung accumulation. Even then, there are an increasing number of reports citing success with lipid-mediated delivery of siRNAs *in vivo*. To improve the delivery of siRNA into human liver cells [25] without transfection agents, lipophilic siRNAs were conjugated with derivatives of cholesterol, lithocholic acid, or lauric acid, where the lipid moieties were covalently linked to the 5′-ends of the RNAs using phosphoramidite chemistry [26]. These could down-regulate the expression of a LacZ expression construct. By conjugating cholesterol to the 3′-end of the sense strand of siRNA by means of a pyrrolidine linker, the pharmacological properties of siRNA molecules was improved by Soutschek et al. [27]. Advantages of cholesterol attachments are evident as being more resistant to nuclease degradation, more stable in the blood by increasing binding to human serum albumin and increased uptake of siRNA molecules by the liver. Intravascular delivery of siRNA molecules is a very simple technique, which was used to protect mice from fulminant hepatitis using siRNAs against Fas receptors by Song et al. [28], who administer Fas siRNA by intravenous injection into mice over a 24-hour period. The authors could show the persist effects for 10 days and protected mice against experimentally induced liver fibrosis. Local delivery of siRNAs have also been tried into the eye to target the VEGF pathway and shown that it could be therapeutically beneficial in neovascularization-related eye diseases. SiRNA topical gels have also been used to deliver them to cells in dermatological applications and cervical cancer treatment [29]. Gene gun method was used for an intradermal administration of nucleic acids to enhance cancer vaccine potency [30]. The other technique is an electroporation, which has been used to deliver siRNAs into the brain [31] and muscles of rodents. Injecting viral vectors for the *in vivo* delivery of siRNA directly have been tried, where an adeno-associated virus (AAV) associated shRNA vector injected directly into the midbrain neurons of adult mice to silence of the tyrosine hydroxylase gene near the site of injection for several weeks. However, there exist an alternative to injection, called as an *ex vivo* approach to generate human immunodeficiency virus (HIV)-1-resistant lymphocytes and macrophages [32]. It was accomplished through using a lentiviral vector, an anti-rev siRNA construct into CD34(+) hematopoietic progenitor cells. The siRNA-transduced progenitor cells were allowed to mature into macrophages *in vitro* and T-cells *in vivo*, [33].

Many machine learning, bio-statistical [34] and association rule mining methods [35] are available that have been developed to solve different problems related to gene silencing and disease discovery. In this book chapter, we provided a comprehensive survey of different machine learning and association rule mining algorithms developed for tackling various biological challenges such as detection of gene signature, biomarker, gene module, potentially disordered protein detection, differentially methylated region, multi-omics data integration, etc. We also described a comparative study of different well-known classifiers along with other used methods for the study. Meanwhile, many gene module discovery based approaches are also developed that employs several machine learning, deep learning and soft computing approaches. In addition, many multi-objective algorithms are also developed to find optimal multi-omics genetic signatures for the respective disease. Furthermore, we demonstrated the

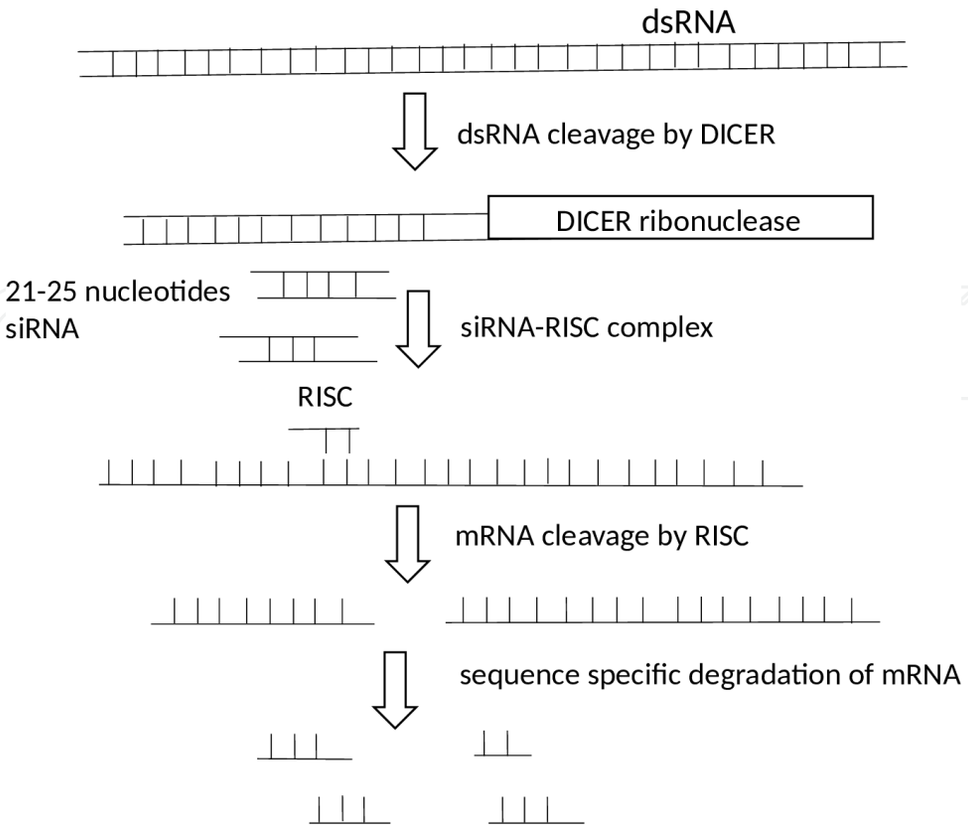

Figure 1. Flowchart of the RNAi mechanism [37, 38].

brief biological information regarding the immense biological challenges for gene activation and their advantages, disadvantages and possible therapeutic strategies. There are certain challenges exist, such as off-target effects, cytotoxicity, need for efficient delivery methods, their clinical implementation need efficient delivery vehicles and siRNA activity, itself, non-specific gene silencing, activation of innate immune system, the lack of efficient in vivo delivery systems still remain to be handled. Apart from these challenges, the development of efficient tissue-specific and differentiation dependent expression of siRNA is essential for transgenic and therapeutic approaches. However, there are successful *in vitro* and *in vivo* experiments for raising hopes in treating human diseases with RNAi [36]. Moreover, our study is useful for the researchers to understand the central idea about RNAi and gene silencing, along with the current machine/deep learning and association rule mining algorithms related to these (**Figure 1**).

2. Fundamental concepts

In this section, some basic symbols of the graph mining, pattern recognition, [39] and information theory are described. A graph is an ordered pair G = (V, E) comprising of a set of vertices denoted as V and a set of edges denoted as E. To avoid ambiguity, the graph is described here precisely as undirected and simple. Let, $Q = (N, E)$ be an unweighted as well as undirected

graph, and H be a (hypograph) of it, ($H \subseteq N$). Further, suppose, the density of H, denoted by $Ds(H)$, be defined as $Ds(H) = \frac{|IE(H)|}{|H|}$, where $IE(H)$ depicts the induced edge-set of H, and $|H|$ refers to the cardinality of H. Suppose, the highest density of the graph H, referred to as $Ds^*(H)$, is illustrated as follows: $Ds^*(H) = \max_{H \subseteq V}\{DS(H)\}$. Now, if $Q = (N, E)$ is a weighted graph, $Ds(H)$ will be $\frac{\sum_{e \in IE(H)} wt_e}{|H|}$, where $IE(H)$ symbolizes the induced edge-set of H, and wt_e denotes the weight of the edge $e \in IE(H)$. Entropy of a random variable evaluates the amount of uncertainty corresponding to the variable [40]. The entropy of a discrete variable A, referred to as $EP(A)$, is defined in the following: $EP(A) = -\sum_{a \in A} p(a) \log_b p(a)$, where $p(a)$ refers to the probability mass function of A, and the value of b, in general, is considered as 2. Mutual information [41] between two random variables estimates the quantity of information that they combinedly share, i.e., the mutual dependency between them. When mutual information is zero, this signifies that these two variables are entirely independent to each other; whereas when mutual information is higher, it signifies that these two variables are extremely dependent on each other.

Topological Overlap Measure (TOM) and other related measures: Ravasz et al. [42] proposed a new measure Topological Overlap Measure (TOM) that provided the similarity between two nodes belonging to a network depending upon nearest neighbor concept. Furthermore, various modified versions of TOM such as weighted TOM (wTOM) [43], generalized TOM (GTOM) [44] are present in the literature. In the course of computing the wTOM, Pearson correlation coefficient scores are first evaluated for all pairs of vertices, and then a soft thresholding power (say, $\beta >= 1$) is utilized from the correlation coefficient matrix through scale free topology. After that, weighted adjacency matrix is calculated using the coefficient matrix using the calculated power β. Then wTOM is computed from the weighted adjacency matrix. In the same way, the GTOM can also be defined just like TOM except it counts the number of m-step neighbors while calculating TOM measure between two vertices. Now, for calculating GTOM of order 0 (i.e., GTOM0), the adjacency score becomes the score of GTOM0. But, for determining the GTOM with higher order than zero (i.e., GTOM1, GTOM2, GTOM3,...), it follows the same procedure of TOM calculation, but counts up to $\ddot{d}$-th neighbors for each vertex ($\ddot{d} = 1, 2, 3, \ldots$). Notably, GTOM1, GTOM2 and other higher order GTOM work only on binary matrix. So, before using those measures, the weighted adjacency matrix is translated into binary matrix in which the greater adjacency value than a specified cutoff (e.g., 70% score of the distance between the minimum and maximum adjacency values is converted into 1, and the lower value than the cutoff is transferred into 0.

In data mining, hierarchical clustering is one of the most popular cluster analyses in forming a hierarchy of clusters. There exist two types of strategies: agglomerative and divisive [45]. As is already known, agglomerative hierarchical clustering does not need any input parameters except the similarity matrix. Thus, there is no extra burden of utilizing cluster initialization as it simply merges two closest clusters at each iteration and continues till a singleton cluster is found. Divisive hierarchical clustering also follows the same style but in a reverse order. This is the major benefit of performing hierarchical clustering over the traditional K-means clustering algorithm, which is sensitive to initialization.

Association rule mining (ARM) [46] is a popular method for generating interesting relationships among different items (*viz*., genes). Suppose, $GST = \{g_1, g_2, \ldots, g_n\}$ be a item set (gene set) and $SST = \{s_1, s_2, \ldots, s_m\}$ be sample set (*viz*., transaction set). Therefore, an association rule can be stated as $A \Rightarrow C$, where $A, C \subseteq GST$ and $A \cap C = \phi$. Notably, A and C symbolize as antecedent and consequent, respectively. An association rule can be described as the cause-effect relationships of the corresponding item sets in the transactions of a transactional data-profile in a big shopping market. A set of bought items may fall into a transaction. In a similar fashion, many genes may occur together in a sample (transaction) of a gene expression profile or similar profile. Many of these genes may be up-regulated or down-regulated, whereas the remaining genes will be non-differentially expressed.

3. Machine learning and rule mining approaches for gene inactivation

Currently, omics data analysis is one of the widely popular research domains. It can be categorized into two major types, single-omics data analysis, and multi-omics data analysis. In earlier, single-omics data processing such as gene expression data processing was highly popular. In those days, basically microarray gene expression data was popular. Now, the microarray data becomes obsolete while RNAseq, next-generation sequencing (NGS) and whole exome sequencing (WES) data become popular. However, the major aim of the single omics data analysis was to identify genetic marker as well as gene module identification. In current era, multi-omics data integration is now a big challenge to any researcher since it consists of various kind of profiles that are either proportional or inversely proportional to each other. Different kinds of regression analysis (logistic regression, sglasso [47, 48], flasso [47], etc.) are popular to integrate the multi-omics data. In case of the multi-omics data, the aim is to determine either single (or, combinatorial) gene marker, or gene signature, or multi-biomolecular closed bio-circuit. There are many machine learning and association rule mining methods available that have been developed to solve different problems related to gene silencing and disease discovery (**Table 1** for tools and **Table 2** for their application). For this regard, Bandyopadhyay et al. provided a comprehensive survey of various statistical tests for determining differentially expressed transcripts from microarray or other related datasets [69]. Then a rank based weighted association rule mining, RANWAR is developed to identify weighted interesting genomic rules applicable to any kind of genomic or epigenomic data [9]. A new technique of gene-based association rule mining approach was developed in [62]. Next, another statistics-based association rule mining technique "StatBicRM" had been proposed that utilized statistical test and Binary Inclusion maximal algorithm (BiMax) to find classification-based genetic rules [46]. Reverently, further enhancement of "StatBicRM" algorithm was performed and a new method of combinatorial marker discovery had been developed whose central concept was based upon the inverse relationship between the gene expression and methylation pattern [50]. In addition, mutual information based feature selection strategy had been incorporated into the statistical methodology, and a new method of identifying epigenetic biomarkers through maximal relevance and minimal redundancy based feature (gene) selection method from bi-omics dataset was proposed [63]. A new method of

Method name	Reference	Type	Brief description
Multi-view gene modules using hypograph mining	Bhadra et al. [49]	Gene-module detection	Module detection from multi-view data using the statistical test and mutual information based dense subgraph.
RANWAR	Mallik et al. [9]	Rank based genomic rule mining	Rank based weighted association rule mining to identify interesting genomic rules applicable to any genomic/epigenomic data.
Combinatorial marker discovery by integrating multiple profiles	Bandyopadhyay et al. [50]	Combinatorial marker discovery	Integrating gene expression and methylation profiles, and identifying combinatorial gene markers.
DTFP-growth	Mallik et al. [51]	Gene based ARM	Multiple-threshold based ARM integrating gene expression, methylation and protein-protein interaction profiles.
StatBicRM	Maulik et al. [46]	Statistical biclustering-based rule mining	Statistical biclustering-based rule mining and analyzing the gene expression and methylation data profiles using it.
sglasso	Augugliaro [47, 48]	Regression method	Sglasso tool develops the structured graphical lasso estimator for the weighted l1-penalized RCON(V, E) model.
flasso	Augugliaro [47, 52]	Regression method	Implements the weight l1-penlized factorial dynamic Gaussian graphical model.
MVDA	Serra et al. [53]	Multi-view genomic profile integration	Works to conjoin the those kinds of data at the levels of the outcomes of every single view clustering iteration.
Machine learning for epigenetics and future medical applications	Holder et al. [54]	Machine learning and deep learning approaches	Active learning and imbalanced class learning are utilized to solve the shortcoming with machine learning for building better feature selection and solving the imbalance data problem.
A machine learning approach to integrate big data for precision medicine	Lee et al. [55]	Molecular marker discovery	The robust molecular markers that might be useful for targeted treatment of the acute myeloid leukemia are identified.
Deep learning based multi-omics integration robustly predicts survival	Chaudhary et al. [56]	Deep learning based multi-omics integration method	A deep learning method is used to integrate multi-omics data and to perform survival study on hepatocellular carcinoma.
Deep learning for genomics: a concise overview	Yue et al. [57]	Deep learning applications on genomic data	The strengths of various deep learning methodologies are demonstrated that are applicable on any kind of genomic profile.
intNMF	Chalise and Fridley [58]	Integrative clustering method	Integrative clustering of several high dimensional profiles and subtype classification by non-negative matrix factorization (NMF).
Multi-modal data analysis for heterogeneous data	Yang and Michailidis [59]	Module detection for heterogeneous data	The multi-modal profile analysis is conducted for heterogeneous data depending upon NMF.
Comparative study and evaluation of the integrative techniques for the multilevel omics data	Pucher et al. [60]	Integrative method for multilevel omics profiles	The comparative study of three integrative methods (*viz.*, NMF, sparse canonical correlation analysis (sCCA) and logic data mining MicroArray Logic Analyzer

Method name	Reference	Type	Brief description
			(MALA)) is conducted on simulated data and real omics profile.
WeCoMXP	Mallik and Bandyopadhyay [61]	Weighted connectivity (similarity) measure	*WeCoMXP* is developed integrating co-expression, co-methylation and protein-protein interactions, and useful for determining the similarity between any two molecules.
Tumor prediction using integrated analysis of expression and methylation	Mallik et al. [62]	Rule-based classifier	Integrated analysis of gene expression and DNA methylation and classification rule mining for tumor/cancer prediction.
Epigenetic gene marker discovery through feature selection	Mallik et al. [63]	Gene based ARM	Epigenetic gene marker discovery using maximal relevance and minimal redundancy based feature selection.

Table 1. The machine learning and rule mining methods related to gene inactivation and RNAi.

Method name	Reference	Type	Brief description
TF-MiRNA-gene network based modules for cytosine variants	Sen et al. [64]	Module detection	TF-MiRNA-gene network based module detection for 5hmC and 5mC brain samples between human and rhesus.
IDPT	Mallik et al. [65]	Intrinsically disordered protein finding	Potential intrinsically disordered protein identification through transcriptomic analysis of genes for epigenetic data.
Integration of DNA methylation data and gene expression data	Singh et al. [66]	Finding differentially methylated regions	Differentially methylated regions are determined and further statistical analysis is performed.
Application of machine-learning algorithms for gene expression regulation	Cheng and Worzel [67]	Applications of machine learning methods on gene regulation	The machine learning strategies on gene regulation are reviewed, and their functional links mediated by histone modifications and transcription factors are demonstrated.
Application of machine-learning techniques on histone methylation	Xu et al. [68]	Predictive model of gene expression by epigenetic factors by regression	A new model is developed to predict the gene expression using the function of histone modification levels through multi-linear regression multivariate adaptive regression splines.

Table 2. Applications of machine learning and rule mining methods related to gene inactivation.

identifying multi-view gene-module identification was also proposed that applied the integrated methodology of statistical method and dense subgraph mining [49]. Detection of strongly connected genetic modules in multi-omics regulatory networks is an important study for the integrated study analysis of the network-based architecture. Many profiles belonging to the multi-omics datasets basically consist of a massive amount of genes, many of them are noisy and redundant. Such kind of noisy and redundant genes (or, features) are irrelevant while obtaining knowledge from the data. Furthermore, it is computationally absurd to utilize any clustering technique on such type of huge sized data profiles to get the dense genetic

clusters. In many times, researchers face problems while calculating and subsequently accumulating the similarity matrix of such massive dimensions consisting of all the mutual dependency information between all the possible gene-pairs equivalent to every such profile. So, managing the high dimensionality of the underlying profile is a critical challenge to the researchers. To overcome the "curse of dimensionality" problem, the job of feature selection is basically treated as one of the most important preprocessing works to remove such noisy and redundant genes, which in turn decreases the total elapsed time. The main purpose of the feature selection is to find an optimal subset of features depending on some optimization conditions by which efficient knowledge discovery can be performed [70]. Depending on the availability of the class labels, the feature selection process can be organized into two types: supervised and unsupervised [71]. Unsupervised feature selection does not need the class label information while choosing the minimized feature subset [72], whereas supervised feature selection selects a subset of favorable features by utilizing the knowledge of class labels into the feature selection procedure. In the case of supervised feature selection, significant test [73], mutual information [74], are some broadly used measures to evaluate the excellence of the candidate features. In the territory of biological rematches, a statistical test is generally treated as one of the important tools for obtaining the significant genes for the big sized datasets, and therefore aids in decreasing the size of the dataset. There are different types of statistical tests such as t-test, significant analysis of microarrays, empirical Bayes test, etc. in the literature.

The significant genes therefore provide a weighted graph in which the nodes refer to the significant genes and the weighted edges signify the association between the related two nodes. Recently, graph data can be obtained in different rising fields of studies for forming the complicated structures *viz*., biological networks, chemical compounds, social networks, protein structures, etc. With the increasing stipulate on the analysis of large sized structured data, graph mining has become one of the most demanding topics of research for identifying the critical relationships among various entities included in the large graphs [75]. In the recent era, analyzing multi-omics dataset is one of the emerging topics of research where different profiles denoting several directions are applied to carry out different important tasks *viz*., marker determination, classification, and clustering. For this regard, many research works have been performed in the following directions *viz*., marker identification [76], classification [77], clustering [78], etc. Recently, Bhadra et al. [49] have developed a new algorithm handling an integrated study comprising of statistical method and normalized mutual information oriented hypo-graph mining to find the multi-omics co-similar genetic modules present in multi-omics datasets. Formerly, various statistical (*viz*., correlation, regression oriented) and/or weight-based techniques (*viz*., [79]) are matured for multi-omics data integration, but not for multi-omics genetic-module detecting. Furthermore, some multi-view data integration mechanism employs various soft-computing methods such as clustering, non-matrix factorization, etc. Recently, Serra et al. [53] proposed a framework for combining different data profiles of multi-view datasets by integrating several clustering results done on each profile through non-matrix factorization. Pucher et al. [60] provided a comprehensive review and comparative study of the three integrative methods (*viz*., non-negative matrix factorization (NMF), sparse canonical correlation analysis (sCCA) and logic data mining MicroArray Logic Analyzer (MALA)) on simulated data as well as real omics profile. In addition, there are many deep

C4.5 classifier	K-nearest neighbors (KNN) classifier	Naive Bayes classifier	Support vector machines (SVM) classifier	Artificial neural networks (ANN) classifier
• Can use both discrete and continuous values.	• Can use only continuous values.	• Can use both discrete and continuous values.	• Can use only continuous values.	• Can use both discrete and continuous values.
• Handles noise.	• Sensitive to noisy features.	• Sensitive to noisy features.	• Is less effective when data contains noisy features.	• Handles noisy features.
• Classes need not be linearly separable.	• Classes need not be linearly separable.	• Classes need not be linearly separable.	• Works well even if data is not linearly separable in the input feature space.	• Works fine even if data is not linearly separable in the input feature space.
• Faces the problem of overfitting.	• Overcomes the problem of overfitting.	• Faces the problem of overfitting.	• Overcomes the problem of overfitting.	• Overcomes the problem of overfitting.
• Needs large searching time.	• Requires higher searching time for a larger data.	• Enormous Computational efficiency.	• Needs higher searching time for a larger data.	• Needs high processing time if neural network is huge.

Table 3. Comparison of different classifiers.

learning techniques that were also developed to handle biological data. Chaudhary et al. [56] proposed a deep learning based methodology to integrate multi-omics data and robustly perform survival study on hepatocellular carcinoma. Furthermore, there are many interesting applications of the above machine learning and deep learning techniques. For example, Xu et al. [68] developed a new model using the regression to predict the gene expression using the function of histone modifications/variants levels through the consecutive regression methods (*viz*., multi-linear regression as well as multivariate adaptive regression splines). Mallik et al. [65] performed a comprehensive analysis to identify potential intrinsically disordered proteins through the transcriptomic analysis of genes for the expression and methylation data. To find differentially methylated regions is also an area of interest. Comparison of different classifiers used in many tools related to RNAi and gene inactivation is described in **Table 3**.

4. Biological challenges for gene inactivation

There are certain challenges exist, such as off-target effects, cytotoxicity, need for efficient delivery methods, their clinical implementation need efficient delivery vehicles and siRNA activity, itself, non-specific gene silencing, activation of innate immune system, the lack of efficient *in vivo* delivery systems still remain to be handled [80]. The effective delivery of RNAi therapeutics *in vivo* is one of the important challenge and have to consider several parameters

http://dx.doi.org/10.5772/intechopen.83470

for an efficient silencing, particle sizing, duration of the RNAi effect, its stability and modification, the delivery system and clearing off-target effects [81]. Apart from these challenges, the development of efficient tissue-specific and differentiation dependent expression of siRNA is essential for transgenic and therapeutic approaches. Bioactive drugs have been shown to perturb the naturally running system as these can clog/saturate the biochemical pathways. Since siRNA/shRNA relies on the endogenous microRNA machinery, thereby high doses of ectopic RNA have the risk of saturating all component of the miRNA pathway components. This was observed in the work by Grimm et al. [82] observed fatality association with high doses of liver-directed AAV-encoded shRNAs in mice, where high doses killed the recipient mice within 2 months. The length threshold of siRNAs seems to vary among cell types and it is an important consideration as dsRNA would induce innate immune responses that would eventually lead to cell death in mammalian. However, dsRNA less than 30 nucleotides have been shown good enough for no induction of cellular toxicity in mammalian and longer dsRNA is known to rapidly induce interferon responses. This suggests the careful risk assessment strategies when using longer and more potent Dicer substrates siRNAs. Moreover, correct RNAi targets are must, though ideal specificity of RNAi targets has not been shown. However if RNAi is going to silence off-targets, it can alter the gene function, which is clearly undesirable, therefore, care should be taken before-hand not to suppress the off-targets. If one third of siRNA are chosen randomly that it results in a toxic phenotype [83]. Comparison of siRNA and miRNA is described **Table 4**. However, there are successful *in vitro* and *in vivo* experiments for raising hopes in treating human disease with RNAi. The epigenetic network is one of the complex regulatory networks where epigenetic mechanisms such as DNA

siRNA	miRNA
• Must be fully complementary to its target mRNA.	• Can be partially complementary to its target mRNA.
• 21–23 nucleotide RNA duplex, notably endogenous siRNAs' origin is more polemic.	• 19–25 nucleotide RNA duplex, derived from gene units.
• dsRNA (30–100 nucleotides), before Dicer processing.	• Precursor miRNA (70–100 nucleotides) with interspersed mismatches and hairpin structure, prior to Dicer processing.
• One mRNA target.	• Can have multiple targets (>100 at the same time).
• For gene regulation, endonucleolytic cleavage of mRNA occurs.	• For gene regulation, translational repression degradation of mRNA occurs.
• Used as a therapeutic agent.	• Utility as a drug target therapeutic agent Diagnostic and biomarker tool.
• siRNAs shut down gene expression at a post-transcriptional level through mRNA degradation.	• MiRNAs silence their target genes mainly and most of the times through translational repression.
• Occurs in plants and lower animals, occurrence in mammals is questionable.	• Occurrence in plants and animals.
• Rarely found as an evolutionary conserved.	• Evolutionary conserved most of the time in the related organism.

Table 4. Comparison of siRNA and miRNA.

Disadvantages	Advantages and possible therapeutic strategies
• RNAi-based therapeutics has led to trigger several off-target (unintended) effects and hence shown host innate immune responses.	• Strategies for selective internalization and with endogenous mechanism without disrupting the natural pathway should be used to achieve maximal benefit from RNAi-based therapeutics.
• Pol III expressed shRNAs delivered in an AAV delivered in mice tail vein through injection was lethal due to acute liver failure.	• Levels of ectopic expression of therapeutic shRNAs should be carefully controlled (low yet effective) to avoid off target effects.
• Using naked siRNA has poor cellular uptake, it activates toll-like receptors and does not target to specific cell types.	• Naked siRNA are comparatively stable and non-immunogenic.
• Viral vectors for shRNA, expensive to create and cause immune reactions.	• High affinity toward infecting target cells, expression can long last.
• Lack of efficient delivery systems is the most critical challenge for the therapeutic applications of small RNAs.	• Identify the critical problem from the literature and allow researchers to publish failed ideas.

Table 5. Disadvantages, advantages of RNAi and possible therapeutic strategies.

methylation and modifications to histone proteins regulate gene expression and high-order DNA structure [84]. Epigenetics is basically a study of heritable changes in phenotypes where the DNA sequences are not changed anymore. DNA methylation [85] is an epigenetic factor that represents the inclusion of a methyl group (–CH3) to the fifth position of a cytosine pyrimidine ring or to the sixth nitrogen position of an adenine purine ring in genomic DNA. DNA methylation generally decreases belong to the gene expression level. In this connection, copy number variation (*CNV*) [86] is another latest domain of research in genomics. It is basically an event where the repetition of different portions of the genome continuously happens, and an alteration on the number of repeats in the genome is recognized between individual to individual in the human population. Copy number variation is a category of structural changes, especially, it is a type of either duplication or deletion event which generally influences a reasonable number of base pairs. It has been realized from recent researches that around two-thirds of the total human genome is made up of repeats. In the case of mammals, copy number alteration provides a significant contribution on producing the necessary deviation in both the population and disease phenotype. Cancer forms by various types of somatic genetic changes including copy-number alternations which affect the activity of the critical genes regulating the growth of the cell. Disadvantages and advantages of RNAi, and possible overcome strategies are demonstrated briefly in **Table 5**.

5. Conclusion

RNAi and gene inactivation are well-known research topics in the research of biomedical field. MiRNA and siRNA are closely associated with RNAi. Various categories of algorithms associated with RNAi and gene silencing have been developed in last 2 decades. In this book

chapter, we provided a comprehensive review of various machine/deep learning as well as association rule mining algorithms that have been developed for handling different biological problems such as gene signature detection, multi-omics data integration, single/combinatorial biomarker identification, gene module detection, potentially disordered protein detection, differentially methylated region finding, and many more. Thereafter, a comparative study of several well-known classifiers along with other used approaches for the study has been included. In addition, we provided a brief biological description of the immense biological challenges for the gene activation along with their advantages, disadvantages and possible therapeutic strategies. Finally, this chapter helps the bioinformaticians to understand the central idea of RNAi and gene silencing along with their peripheral machine/deep learning and association rule mining algorithms for the benevolent of the disease discovery as well as possible therapeutic values.

Author details

Saurav Mallik[1,2]*, Ujjwal Maulik[3], Namrata Tomar[4], Tapas Bhadra[5], Anirban Mukhopadhyay[6] and Ayan Mukherji[7]

*Address all correspondence to: sauravmtech2@gmail.com

1 Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India

2 Department of BioStatistics, University of Miami, Miami, Florida, USA

3 Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

4 Department of Bio Medical Engineering, Medical College of Wisconsin, Milwaukee, WI, USA

5 Department of Computer Science and Engineering, Aliah University, Kolkata, India

6 Department of Computer Science and Engineering, University of Kalyani, Kalyani, West Bengal, India

7 Department of Computer Science and Engineering, Mallabhum Institute of Technology, Bishnupur, West Bengal, India

References

[1] Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, et al. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003;**115**(2):199-208

[2] van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhes JP. Gene co-expression analysis for functional classification and gene-disease predictions. Briefings in Bioinformatics. 2018;**19**(4):575-592. DOI: 10.1093/bib/bbw139

[3] Sen GL, Blau HM. A brief history of RNAi: The silence of the genes. The FASEB Journal. 2006;**20**:1293-1299. DOI: 10.1096/fj.06-6014rev

[4] Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes & Development. 2001;**15**(2):188-200. Available from: www.genesdev.org/cgi/doi/10.1101/gad.862301

[5] Fuchs U, Damm-Welk C, Borkhardt A. Silencing of disease-related genes by small interfering RNAs. Current Molecular Medicine. 2004;**4**(5):507-517. DOI: 10.2174/1566524043360492

[6] Zhang Y, Zhang YF, Bryant J, Charles A, Boado RJ, Pardridge WM. Intravenous RNA interference gene therapy targeting the human epidermal growth factor receptor prolongs survival in intracranial brain cancer. Clinical Cancer Research. 2004;**10**(11):3667-3677. DOI: 10.1158/1078-0432.CCR-03-0740

[7] Chu Y, Corey DR. RNA sequencing: Platform selection, experimental design, and data interpretation. Nucleic Acid Therapeutics. 2012;**22**(4):271-274. DOI: 10.1089/nat.2012.0367

[8] Maher CA, Kumar-Sinha C, Cao X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;**458**(7234):97-101. DOI: 10.1038/nature07638

[9] Mallik S, Mukhopadhyay A, Maulik U. RANWAR: Rank-based weighted association rule mining from gene expression and methylation data. IEEE Transactions on Nanobioscience. 2015;**14**(1):59-66. DOI: 10.1109/TNB.2014.2359494

[10] Aqil M, Naqvi AR, Mallik S, Bandyopadhyay S, Maulik U, Jameel S. The HIV Nef protein modulates cellular and exosomal miRNA profiles in human monocytic cells. Journal of Extracellular Vesicles. 2014;**3**:23129. DOI: 10.3402/jev.v3.23129

[11] Aqil M, Mallik S, Bandyopadhyay S, Maulik U, Jameel S. Transcriptomic analysis of mRNAs in human monocytic cells expressing the HIV-1 Nef protein and their exosomes. BioMed Research International. 2015;**2015**:1-10. Article Id: 492395. DOI: 10.1155/2015/492395

[12] Kaymaz BT, Kosova B. Advances in therapeutic approaches using RNA-interference as a gene silencing tool. Advanced Techniques in Biology and Medicine. 2013;**1**(2):1-9. DOI: 10.4172/atbm.1000108

[13] Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature. 2000;**404**:293-296. DOI: 10.1038/35005107

[14] Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell. 2004;**116**:281-297. DOI: 10.1016/S0092-8674(04)00045-5

[15] Behlke MA. Progress towards *in vivo* use of siRNAs. Molecular Therapy. 2006;**13**:644-670. DOI: 10.1016/j.ymthe.2006.01.001

[16] Brummelkamp TR, Nijman SM, Dirac AM, Bernards R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB. Nature. 2003;**424**:797-801. DOI: 10.1038/nature01811

[17] Slimane-Hadj R, Lepelletier Y, Lopez N, Garbay C, Raynaud F. Short interfering RNA (siRNA), a novel therapeutic tool acting on angiogenesis. Biochimie. 2007;**89**:1234-1244. DOI: 10.1016/j.biochi.2007.06.012

[18] Park S, Chapuis N, Tamburini J, et al. Role of the PI3K/AKT and mTOR signaling pathways in acute myeloid leukemia. Haematologica. 2010;**95**(5):819-828. DOI: 10.3324/haematol.2009.013797

[19] Zhou J, Ching YQ, Chng WJ. Aberrant nuclear factor-kappa B activity in acute myeloid leukemia: From molecular pathogenesis to therapeutic target. Oncotarget. 2015;**6**(8):5490-5500. DOI: 10.18632/oncotarget.3545

[20] Qiuwei P, Rong C, Xinyuan L, Cheng Q. A novel strategy for cancer gene therapy: RNAi. Chinese Science Bulletin. 2006;**51**(10):1145-1151. DOI: 10.1007/s11434-006-1145-x

[21] Ge Q, Filip L, Bai A, Nguyen T, Eisen HN, Chen J. Inhibition of influenza virus production in virus-infected mice by RNA interference. Proceedings of the National Academy of Sciences of the United States of America. 2004;**101**(23):8676-8681. DOI: 10.1073/pnas.0402486101

[22] Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003;**115**(2):209-216. DOI: 10.1016/S0092-8674(03)00801-8

[23] Liu F, Song Y, Liu D. Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. Gene Therapy. 1999;**6**(7):1258-1266. DOI: 10.1038/sj.gt.3300947

[24] Sioud M, Sorensen DR. Systemic delivery of synthetic SiRNAs. Methods in Molecular Biology. 2004;**252**:515-522. DOI: 10.1385/1-59259-746-7:515

[25] Kim B, Tang Q, Biswas PS, Xu J, Schiffelers RM, Xie FY, et al. Inhibition of ocular angiogenesis by SiRNA targeting vascular endothelial growth factor pathway genes: Therapeutic strategy for herpetic stromal keratitis. The American Journal of Pathology. 2004;**165**(6):2177-2185. DOI: 10.1016/S0002-9440(10)63267-1

[26] Lorenz C, Hadwiger P, John M, Vornlocher HP, Unverzagt C. Steroid and lipid conjugates of SiRNAs to enhance cellular uptake and gene silencing in liver cells. Bioorganic & Medicinal Chemistry Letters. 2004;**14**(19):4975-4977. DOI: 10.1016/j.bmcl.2004.07.018

[27] Soutschek J, Akinc A, Bramlage B, Charisse K, Constien R, Donoghue M, et al. Therapeutic silencing of an endogenous gene by systemic administration of modified SiRNAs. Nature. 2004;**432**(7014):173-178. DOI: 10.1038/nature03121

[28] Song E, Lee SK, Wang J, Ince N, Ouyang N, Min J, et al. RNA interference targeting fas protects mice from fulminant hepatitis. Nature Medicine. 2003;**9**(3):347-351. DOI: 10.1038/nm828

[29] Jiang M, Rubbi CP, Milner J. Gel-based application of SiRNA to human epithelial cancer cells induces RNAi-dependent apoptosis. Oligonucleotides. 2004;**14**(4):239-248. DOI: 10.1089/oli.2004.14.239

[30] https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/?datasetSearch=TCGA, [Accessed: 15th May, 2018]

[31] Akaneya Y, Jiang B, Tsumoto T. RNAi-induced gene silencing by local electroporation in targeting brain region. Journal of Neurophysiology. 2005;**93**(1):594-602. DOI: 10.1152/jn.00161.2004

[32] Banerjea A, Li MJ, Bauer G, Remling L, Lee NS, Rossi J, et al. Inhibition of HIV-1 by lentiviral vector-transduced SiRNAs in T lymphocytes differentiated in SCID-Hu mice and CD34+ progenitor cell-derived macrophages. Molecular Therapy. 2003;**8**(1):62-71. DOI: 10.1016/S1525-0016(03)00140-0

[33] Golzio M, Mazzolini L, Moller P, Rols MP, Teissie J. Inhibition of gene expression in mice muscle by *in vivo* electrically mediated SiRNA delivery. Gene Therapy. 2005;**12**(3):246-251. DOI: 10.1038/sj.gt.3302405

[34] Vapnik V. The Nature of Statistical Laerning Theory. 2nd ed. New York, USA: Springer; 2013. DOI: 10.1007/978-1-4757-3264-1

[35] Agrawal R, Imielinski T, and Swami A. Mining association rules between sets of items in large databases. In: Proceeding SIGMOD '93 Proceedings of the 1993 ACM SIGMOD international conference on Management of data; Washington, DC, USA. May 25–28, 1993. DOI: 10.1145/170035.170072

[36] Kim DH, Rossi JJ. Strategies for silencing human disease using RNA interference. Nature Reviews. Genetics. 2007;**8**:173-184. DOI: 10.1038/nrg2006

[37] Agrawal A, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar R K, Mukherjee SK. RNA interference: Biology, mechanism, and applications. Microbiology and Molecular Biology Reviews. 2003;**67**(4):657-685. DOI: 10.1128/MMBR.67.4.657–685.200

[38] Tijsterman M, Plasterk RHA. Dicers at RISC: The mechanism of RNAi. Cell. 2004:**117**(1); 1-3. DOI: 10.1016/S0092-8674(04)00293-4

[39] Fukunaga K. Introduction to Statistical Pattern Recognition. 2nd ed. San Diego, USA: Academic Press; 2013. eBook ISBN: 9780080478654

[40] Cover TM, Thomas JA. Elements of Information Theory. 2nd ed. New York, USA: John Wiley & Sons; 2012. ISBN: 0-471-20061-1

[41] Battiti R. Using mutual information for selecting features in supervised neural net learning. IEEE Transactions on Neural Networks. 1994;**5**(4):537-550. DOI: 10.1109/72.298224

[42] Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. Science. 2002;**297**(5586):1551-1555. DOI: 10.1126/science.1073374

[43] Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;**9**(1):559. DOI: 10.1186/1471-2105-9-559

[44] Yip A, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics. 2007;**8**(1):22. DOI: 10.1186/1471-2105-8-22

[45] Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967;**32**(3):241-254. DOI: 10.1007/BF02289588

[46] Maulik U et al. Analyzing gene expression and methylation data profiles using StatBicRM: Statistical Biclustering-based rule mining. PLoS One. 2015;**10**(4):e0119448. DOI: 10.1371/journal.pone.0119448

[47] Wit EC, Mineo AM, Augugliaro L, Abbruzzo A. Cyclic coordinate for penalized Gaussian Graphical Models with symmetry restrictions. In: Proceeding of COMPSTAT 2014—21th International Conference on Computational Statistics; Geneva. August 19–24, 2014. Available from: http://hdl.handle.net/10447/96091

[48] Hojsgaard S, Lauritzen SL. Graphical gaussian models with edge and vertex symmetries. Journal of the Royal Statistical Society. Series B. 2008;**70**(5):1005-1027. DOI: 10.1111/j.1467-9868.2008.00666.x

[49] Bhadra T, Mallik S, Bandyopadhyay S. Identification of multi-view gene modules using mutual information based hypograph mining. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017:1-12 (accepted). DOI: 10.1109/TSMC.2017.2726553

[50] Bandyopadhyay S, Mallik S. Integrating multiple data sources for combinatorial marker discovery: A study in tumorigenesis. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2018;**15**(2):673-687. DOI: 10.1109/TCBB.2016.2636207

[51] Mallik S, Bhadra T, Mukherji A. DTFP-growth: Dynamic threshold based FP-growth rule mining algorithm through integrating gene expression, methylation and protein-protein interaction profiles. IEEE Transactions on Nanobioscience. 2018;(99):1-10. DOI: 10.1109/TNB.2018.2803021

[52] Wit EC, Abbruzzo A. Factorial graphical models for dynamic networks. Networking Science. 2015;**3**(1):37-57. DOI: 10.1017/nws.2015.2

[53] Serra A, Fratello M, Fortino V, Raiconi G, Tagliaferri R, Greco D. MVDA: A multi-view genomic data integration methodology. BMC Bioinformatics. 2015;**16**(1):261. DOI: 10.1186/s12859-015-0680-3

[54] Holder LB, Haque MM, Skinner MK. Machine learning for epigenetics and future medical applications. Cancer Research. 2017;**12**(7):505-514. DOI: 10.1080/15592294.2017.1329068

[55] Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nature Communications. 2018;**9**(42):1-13. DOI: 10.1038/s41467-017-02465-5

[56] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep LearningBased multi-Omics integration robustly predicts survival in liver cancer. Nature Communications. 2018;**24**(6):1248-1259. DOI: 10.1158/1078-0432.CCR-17-0853

[57] Yue T, Wang H. Deep learning for genomics: A concise overview. In: Nature Communications, Book Chapter, Handbook of Deep Learning Applications. Smart Innovation, Systems and Technologies 136, Springer Nature Switzerland AG. 2019. Available from: https://arxiv.org/abs/1802.00810

[58] Chalise P, Fridley BL. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. PLoS One. 2017;**12**(5):e0176278. DOI: 10.1371/journal.pone.0176278

[59] Yang Z, Michailidis G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. Bioinformatics. 2016;**32**(1):1-8. DOI: 10.1093/bioinformatics/btv544.

[60] Pucher BM, Zeleznik OA, Thallinger GG. Comparison and evaluation of integrative methods for the analysis of multilevel omics data: A study based on simulated and experimental cancer data. 2018. DOI: 10.1093/bib/bby027

[61] Mallik S, Bandyopadhyay S. WeCoMXP: Weighted connectivity measure integrating Co-methylation, Co-expression and protein-protein interactions for gene-module detection. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2018:1-11 (accepted). DOI: 10.1109/TCBB.2018.2868348

[62] Mallik S et al. Integrated statistical and rule-mining techniques for DNA methylation and gene expression data analysis. Journal of Artificial Intelligence and Soft Computing Research. 2013;**3**(2):101-115. DOI: 10.2478/jaiscr-2014-0008

[63] Mallik S et al. Identifying epigenetic biomarkers using maximal relevance and minimal redundancy based feature selection for multi-Omics data. IEEE Transactions on Nanobioscience. 2017;**16**(1):3-10. DOI: 10.1109/TNB.2017.2650217

[64] Sen S et al. Detecting TF-MiRNA-gene network based modules for 5hmC and 5mC brain samples: A intra- and inter-species case-study between human and rhesus. BMC Genetics. 2018;**19**(9):1-22. DOI: 10.1186/s12863-017-0574-7

[65] Mallik S, Sen S, Maulik U. IDPT: Insights into potential intrinsically disordered proteins through Transcriptomic analysis of genes for prostate carcinoma epigenetic data. Gene. 2016;**586**(1):87-96. DOI: 10.1016/j.gene.2016.03.056

[66] Singh A, Rahman R, Hasija Y. Integration of DNA methylation data and gene expression data for prostate adenocarcinoma: A proof of concept. Current Bioinformatics. 2017;**12** (999):423-430. DOI: 10.2174/1574893612666170328171106

[67] Cheng C, Worzel WP. Application of machine-learning methods to understand gene expression regulation, book chapter. In: Genetic Programming Theory and Practice XII. Springer; 2015:1-15. DOI: 10.1007/978-3-319-16030-6_1

[68] Xu X, Hoang S, Mayo MW, Bekiranov S. Application of machine learning methods to histone methylation ChIP-Seq data reveals H4R3me2 globally represses gene expression. BMC Bioinformatics. 2010;**11**:396. DOI: 10.1186/1471-2105-11-396

[69] Bandyopadhyay S, Mallik S, Mukhopadhyay A. A survey and comparative study of statistical tests for identifying differential expression from microarray data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2013;**11**(1):95-115. DOI: 10.1109/TCBB.2013.147

[70] Xiao C, Chaovalitwongse WA. Optimization models for feature selection of decomposed nearest neighbor. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2016; **46**(2):177-184. DOI: 10.1109/TSMC.2015.2429637

[71] Bandyopadhyay S, Bhadra T, Maulik U, Mitra P. Integration of dense subgraph finding with feature clustering for unsupervised feature selection. Pattern Recognition Letters. 2014;**40**:104-112. DOI: 10.1016/j.patrec.2013.12.008

[72] Wang S, Zhu W. Sparse graph embedding unsupervised feature selection. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2016;**48**(3):329-341. DOI: 10.1109/TSMC.2016.2605132

[73] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism. 2012;**10**(2):486-489. DOI: 10.5812/ijem.3505

[74] Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;**27**(8):1226-1238. DOI: 10.1109/TPAMI.2005.159

[75] Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. The Morgan Kaufmann Series in Data Management Systems; 2006. DOI: 10.1145/565117.565130

[76] Mallik S, Maulik U. MiRNA-TF-gene network analysis through ranking of biomolecules for multi-informative uterine leiomyoma dataset. Journal of Biomedical Informatics. 2015; **57**:308-319. DOI: 10.1016/j.jbi.2015.08.014

[77] Henry VJ, Bandrowski AE, Pepin AS, et al. OMICtools: An informative directory for multi-omic data analysis, Database. Oxford; 2014;**2014**:bau069. DOI: 10.1093/database/bau069

[78] Wang D, Gu J. Integrative clustering methods of multi-omics data for molecule-based cancer classifications. Quantitative Biology. 2016;**4**(1):58-67. DOI: 10.1007/s40484-016-0063-4

[79] Cao KAL, Gonzlez I, Djean S. integrOmics: An R package to unravel relationships between two omics datasets. Bioinformatics. 2009;**25**(21):2855-2856. DOI: 10.1093/bioinformatics/btp515

[80] Aagaard L, Rossi JJ. RNAi therapeutics: Principles, prospects and challenges. Advanced Drug Delivery Reviews. 2008;**59**(2–3):75-86. DOI: 10.1016/j.addr.2007.03.005

[81] Gao K, Huang L. Achieving efficient RNAi therapy: Progress and challenges. Acta Pharmaceutica Sinica B. 2013;**3**(4):213-225. DOI: 10.1016/j.apsb.2013.06.005

[82] Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, et al. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature. 2006; **441**(7092):537-541. DOI: 10.1038/nature04791

[83] Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, et al. Off-target effects by siRNA can induce toxic phenotype. RNA. 2006;**12**(7):1188-1196. DOI: 10.1261/rna.28106

[84] Gropman AL, Batshaw ML. Epigenetics, copy number variation, and other molecular mechanisms underlying neurodevelopmental disabilities: New insights and diagnostic approaches. Journal of Developmental & Behavioral Pediatrics. 2010;**31**(7):582-591. DOI: 10.1097/DBP.0b013e3181ee384e

[85] Bhattacharjee S, Renganaath K, Mehrotra R, Mehrotra S. Combinatorial control of gene expression. BioMed Research International. 2013;**2013**:1-11. Article Id: 407263. DOI: 10.1155/2013/407263

[86] Sharp AJ, Locke DP, Mcgrath SD, et al. Segmental duplications and copy-number variation in the human genome. The American Journal of Human Genetics. 2005;**77**(1):78-88. DOI: 10.1086/431652

IntechOpen