Physiol. Res. 73: 687-702, 2024 *https://doi.org/10.33549/physiolres.935389*

REVIEW

Reproductomics: Exploring the Applications and Advancements of Computational Tools

Pallav SENGUPTA[1]*, Sulagna DUTTA[2]*, Fong Fong LIEW[3], Antony V. SAMROT[4], Sujoy DASGUPTA[5], Muhammad Ali RAJPUT[1], Petr SLAMA[6], Adriana KOLESAROVA[7,8], Shubhadeep ROYCHOUDHURY[9]

[1]Department of Biomedical Sciences, College of Medicine, Gulf Medical University, Ajman, UAE, [2]Basic Medical Sciences Department, College of Medicine, Ajman University, Ajman, UAE, [3]Department of Preclinical Sciences, Faculty of Dentistry, MAHSA University, Jenjarom, Malaysia, [4]Faculty of Medicine, MAHSA University, Jenjarom, Malaysia, [5]Genome Fertility Centre, Kolkata, India, [6]Laboratory of Animal Immunology and Biotechnology, Department of Animal Morphology, Physiology and Genetics, Faculty of AgriSciences, Mendel University in Brno, Brno, Czech Republic, [7]Institute of Applied Biology, Faculty of Biotechnology and Food Sciences, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic, [8]AgroBioTech Research Centre, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic, [9]Department of Life Science and Bioinformatics, Assam University, Silchar, India

** These authors contributed equally to this work.*

Received April 17, 2024
Accepted June 25, 2024

Summary

Over recent decades, advancements in omics technologies, such as proteomics, genomics, epigenomics, metabolomics, transcriptomics, and microbiomics, have significantly enhanced our understanding of the molecular mechanisms underlying various physiological and pathological processes. Nonetheless, the analysis and interpretation of vast omics data concerning reproductive diseases are complicated by the cyclic regulation of hormones and multiple other factors, which, in conjunction with a genetic makeup of an individual, lead to diverse biological responses. Reproductomics investigates the interplay between a hormonal regulation of an individual, environmental factors, genetic predisposition (DNA composition and epigenome), health effects, and resulting biological outcomes. It is a rapidly emerging field that utilizes computational tools to analyze and interpret reproductive data, with the aim of improving reproductive health outcomes. It is time to explore the applications of reproductomics in understanding the molecular mechanisms underlying infertility, identification of potential biomarkers for diagnosis and treatment, and in improving assisted reproductive technologies (ARTs). Reproductomics tools include machine learning algorithms for predicting fertility outcomes, gene editing technologies for correcting genetic abnormalities, and single cell sequencing techniques for analyzing gene expression patterns at the individual cell level. However, there are several challenges, limitations and ethical issues involved with the use of reproductomics, such as the applications of gene editing technologies and their potential impact on future generations are discussed. The review comprehensively covers the applications and advancements of reproductomics, highlighting its potential to improve reproductive health outcomes and deepen our understanding of reproductive molecular mechanisms.

Key words

Genetic abnormalities • Human reproduction • Male infertility • Integrative *in-silico* analysis • Interactomics

Corresponding author

S. Roychoudhury, Department of Life Science and Bioinformatics, Assam University, Silchar, India. E-mail: shubhadeep1@gmail.com

Introduction

Omics technologies, including genomics, transcriptomics, epigenomics, proteomics, metabolomics, and microbiomics, have recently improved our understanding of the molecular underpinnings of diverse physiological and pathological processes [1]. Omics methods enable comprehensive examination of various events and interactions, from structural elements to biological processes, in an integrated way. This allows simultaneous analysis of multiple biological components, including epigenetic markers, genes, mRNA, proteins, and metabolites, within a single experiment.

A mere decade ago, investigating multifactorial human reproductive conditions posed significant difficulties [2]. The advent of omics technologies has broadened our understanding in this domain, enabling a more comprehensive perspective on complex biological systems. A key advantage of these approaches is the abundant data they offer, which can be acquired with minimal investment and exertion [1]. Nevertheless, this advantage occasionally transforms into an impediment, necessitating powerful tools for distilling biologically significant conclusions from the immense quantities of data generated. The complexity of the human body further complicates the comprehension and analysis of omics data. For example, the field of reproductomics focuses on the interplay between periodically regulated hormones and various other factors, which, when combined with genetic composition of an individual, result in distinct biological reactions.

The vast amount of data generated by these high-throughput techniques remains considerably underutilized, posing a formidable challenge for biomedical research, as the information necessitates meticulous processing and analysis *via* robust bioinformatic applications. The crux of the issue is that a data management bottleneck has been reached, wherein data volumes vastly surpass our ability to thoroughly analyze and interpret them. In reality, numerous gene expression datasets, amounting to millions, can be found in public repositories such as the Gene Expression Omnibus (GEO) and ArrayExpress [3]. However, despite the abundance of available data, only a small number of researchers fully utilize this information to uncover new insights. Rather, they tend to concentrate on a restricted subset of data in order to draw comparisons with their own results [4].

Thus, in this review, we comprehensively present and discuss a diverse range of computational strategies employed to examine the vast and intricate reproductomics data generated in contemporary research. We highlight well-recognized databases and analytical instruments pertinent to this field and offer a demonstrative example of their application within the realm of reproductomics. The rationale for this review is grounded in the ever-increasing volume and intricacy of omics datasets, which demand employment of effective computational techniques to facilitate meaningful interpretation and derive insightful conclusions. By acquainting readers with established databases and tools and illustrating their use in reproductomics, this review endeavors to enhance comprehension and stimulate the adoption of computational methodologies across various research disciplines related to omics.

Integrative *in-silico* analysis: A unified approach to combining diverse studies to analogous research questions

In the realm of omics data analysis, *in-silico* data mining provides a method for amalgamating disparate studies with analogous research questions (e.g., endometrial receptivity transcriptomics and endometriosis-associated genes) [5-7]. Despite copious endometrial transcriptome data within reproductomics, only three published studies so far have employed *in-silico* data mining [5-7]. Bhagwat and colleagues developed the Human Gene Expression Endometrial Receptivity Database (HGEx-ERdb), which includes data on 19,285 endometrial genes and highlights 179 genes associated with receptivity [7]. Zhang *et al.* analyzed raw microarray data from three prior studies, identifying 148 potential receptive endometrium biomarkers [6,8-10]. Tapia *et al.* integrated gene lists from seven preceding studies, listing 61 endometrial receptivity biomarkers [5]. A mere nine genes overlap among these three studies, underscores omics analyses' limitations [11] and the necessity for further research.

Regarding endometriosis, data mining elucidates certain genes' pathogenic roles. Wang *et al.* developed a decision tree to analyze clinical data from 178 endometriosis patients, determining ultrasound-guided aspiration treatment efficacy thresholds based on endometrioma quantity and size [12]. Mathew *et al.* mined public endometrial microarray data, implicating transcription factor FOXD3 in endometrial hormonal

regulation and endometriosis-associated gene expression [13]. Using a unique approach, Liu and Zhao conducted text mining on 19,904 PubMed articles, identifying 1531 endometriosis-related genes, with 121 showing significant associations and enrichment across multiple biological processes [14].

Correlation analysis in reproductomics presents challenges in both execution and interpretation, particularly when examining epigenomic modifications such as DNA methylation, which profoundly influences gene expression and underlying biological processes [15]. DNA methylation, a dynamic process, plays a critical role in regulating gene expression and functional alterations within hormone-dependent endometrial tissue [16]. Recent investigations have explored endometrial methylome variations throughout the menstrual cycle, revealing the significance of epigenetic regulation in distinct hormonal environments within the human endometrium [17].

Saare *et al.* analyzed endometrial DNA methylome signatures in healthy women and endometriosis patients, revealing minimal differences between the groups, suggesting that epigenetic alterations are not responsible for the aberrant expression of genes implicated in endometriosis pathogenesis [18]. Kukushkina *et al.* posited that transcriptomic fluctuations during the implantation window may arise from global DNA methylation pattern changes, establishing a link between methylation and gene expression activation/repression [19]. However, non-linear associations between the epigenome and transcriptome have been proposed [20,21], further complicating the understanding of reproductive processes [22]. Consequently, elucidating the precise correlation between DNA methylation and gene expression necessitates additional investigation.

Meta-analysis is an advanced computational strategy in omics research that facilitates the identification of patterns across studies, thereby increasing statistical power and enhancing the reliability of findings [23]. Transcriptome analysis of endometrial receptivity represents a primary focus within reproductomics. Nevertheless, discrepancies in experimental design, endometrial sampling, sample selection, data processing pipelines, and data presentation standards have impeded meta-analyses in this field [11]. Ideally, meta-analyses necessitate raw expression dataset examination; however, data inaccessibility and variations in gene transcript counts and technological platforms render proper dataset integration challenging.

To overcome these limitations, Altmäe *et al.* employed a robust rank aggregation method designed to compare distinct gene lists and identify common overlapping genes [24]. Through this methodology, the authors analyzed differentially expressed gene lists from nine studies, comprising 96 endometrial biopsies from healthy women, to generate an updated meta-signature of endometrial receptivity biomarkers [25]. Their meta-analysis identified 57 potential biomarkers, with SPP1, PAEP, GPX3, GADD45A, MAOA, CLDN4, IL15, CD55, DP44, ANXA4, and S100P meriting further attention [5-7,26,27]. As the inaugural transcriptome-based meta-analysis in reproductomics, this research could inspire innovative analytic approaches.

Another pivotal meta-analysis in reproductomics involved Genome-Wide Association Studies (GWAS) for endometriosis [28]. Rahmioglu *et al.* assessed the consistency and heterogeneity of results from eight prior studies, demonstrating remarkable congruence and minimal population-based heterogeneity within endometriosis GWAS [28].

Reproductomics has significantly contributed to understanding the molecular mechanisms underlying polycystic ovary syndrome (PCOS). Studies have identified several microRNAs (miRNAs) that are dysregulated in PCOS, which can serve as potential biomarkers for diagnosis and targets for therapeutic intervention. For instance, miRNA-409 has been shown to play a role in the pathogenesis of PCOS, affecting ovarian function and insulin resistance [29,30].

Research utilizing reproductomics tools has identified crucial pathways and genetic markers associated with premature ovarian insufficiency (POI). Mesenchymal stem cell-derived extracellular vesicles (MSC-EVs) have emerged as a promising therapeutic approach for POI, showing potential in restoring ovarian function and improving fertility outcomes [30].

Genomic and transcriptomic analyses have shed light on the etiology of uterine fibroids. Studies have revealed alterations in gene expression and signaling pathways that contribute to fibroid development and growth. miRNAs have also been implicated in the regulation of genes involved in the proliferation and apoptosis of fibroid cells [29].

Reproductomics has been pivotal in identifying biomarkers for early detection and treatment targets for ovarian cancer. Differential expression of miRNAs and

other non-coding RNAs has been linked to ovarian cancer pathogenesis, providing insights into tumor biology and potential avenues for therapeutic intervention [29].

Hypoxia has been identified as a critical factor influencing various reproductive diseases, including ovarian cancer and uterine fibroids. Reproductomics studies have shown that hypoxia-regulated genes and pathways play significant roles in the progression of these diseases, offering potential targets for therapeutic strategies [30].

Largely, reproductomics continues to unveil the complex molecular landscapes of reproductive diseases, aiding in the development of diagnostic tools and novel treatments aimed at improving reproductive health outcomes.

Decoding cellular behavior through inter-disciplinary systems biology

Systems biology, an interdisciplinary approach to deciphering omics data, amalgamates genomics, epigenomics, transcriptomics, proteomics, and metabolomics to generate computational models that interpret cellular, tissue, or organism behavior [11]. This holistic methodology surpasses traditional reductionist strategies, which inadequately describe molecular intricacies across entire systems [31]. Systems biology tackles the challenge of merging data from various disciplines like mathematics, chemistry, physics, biology, statistics, and computational engineering [32]. Utilizing high-throughput omics platforms, systems biology endeavors to provide comprehensive interpretations of biological information, prioritizing complex interactions over isolated characteristics of biological systems [11]. Applications in reproductomics include endometrial-receptivity investigation, placental analysis, sperm examination, and preterm birth analysis [33-36]. Ghosh *et al.* conducted a seminal study employing a systems biology approach to identify key molecules in blastocyst implantation through endometrial omics analysis [37]. Their investigation utilized expression networks to characterize the transcriptomic profile of endometrium in rhesus monkeys, concluding that embryo presence modulates the endometrial transcriptomic network, facilitating blastocyst implantation [37]. Another study combined proteomics, transcriptomics, and interactomics data to develop a dynamic interaction map for the human sperm flagella microtubulome [36]. Through an integrated genomic workflow, 116 gene products associated with the human sperm microtubulome were identified, and a disease-interaction network revealed novel factors potentially linked to sperm motility and male fertility, including CUL3 and DCDC2C [36].

Systems medicine: Advancing predictive, preventive, and personalized healthcare

Systems medicine, an application of systems biology in biomedical research, aims to achieve predictive, preventive, and personalized medicine [38]. While gaining traction, it requires advancements in analytical and computational technologies to integrate into clinical practice [39]. Efforts include detecting personalized biomarkers in reproductive biology to predict treatment outcomes. Kyrgiou *et al.* developed an artificial neural network-based system for managing cervical issues, showing improved diagnostic accuracy compared to traditional tests. In Assisted Reproductive Technologies (ARTs), research focuses on enhancing the techniques for increased transferable embryo yields and better embryo selection [39]. This emerging reproductomics field holds potential for developing personalized treatment protocols for fertility, infertility, and gynecologic complications [40].

Interactomics: Networked molecular interactions

An approach to systems biology, or ‘interactomics,’ examines global molecular interactions at various levels using networks. In reproductomics, studies have explored the molecular interactome between embryos and endometrium to understand the implantation process [32,40,41]. Haouzi *et al.* compared gene expression in stimulated endometrium and blastocyst trophectoderm cells during implantation in women receiving *in vitro* Fertilization (IVF), identifying key interacting molecules [41]. Another study analyzed the transcriptome of mural trophectoderm cells, comparing it to human embryonic stem cell-derived trophoblasts, and identified molecules involved in implantation [42]. Altmäe *et al.* conducted the first true interactome study of molecular networks between human embryos and endometrium, emphasizing cell adhesion molecules and cytokine-cytokine receptor interactions in implantation [33]. Interactome analysis has also been applied to male reproduction, identifying a unique interaction network involving the amyloid precursor protein (APP) and key molecules in sperm-oocyte interaction [43,44].

Essential computational tools in reproductomics

Omics data storage: Solutions for efficient research insights

The substantial volume of data generated through omics research necessitates storage solutions to facilitate valuable insights for investigators. Consequently, there has been a growing advancement in the establishment of both public and private repositories to house the findings from omics investigations. Prime instances of frequently updated repositories for human genetic variant information, ascertained by genome-wide association studies, include the genome variation databases GWASdb, GWAS Central, and the GWAS Catalog developed by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) [45,46]. They aggregate sets of traits/conditions linked with Single Nucleotide Polymorphisms (SNPs), which represent the most prevalent form of genetic variations (Fig. 1).

Regarding methylome data, MethBase from public bisulfite-sequencing datasets offer methylome information of well-studied species with varying methylation levels [47]. DiseaseMeth version 2.0 compiles aberrant DNA methylation in human diseases and disease-gene associations based on high-throughput methylome experiments [48]. For microRNA (miRNA) expression data, miRBase and microRNA.org provide miRNA sequences, annotation, target prediction, and expression profiles [49,50]. The GEO, established by the NCBI, is a public functional genomics database containing >4000 expression datasets as of June 2017 [51]. ArrayExpress is another recommended repository for functional genomics data [52].

For proteome data, repositories include PeptideAtlas, PRIDE, Proteomics DB, Human Proteome Map, and Plasma Proteome Database [53-57]. Phenopedia offers a disease-centered view of genetic association studies, while DiseaseConnect analyzes and visualizes diseases with shared molecular mechanisms [58,59].

OMICtools is a manually curated metadatabase providing an overview of accessible omics data tools [60]. In reproductomics, databases include Follicle Online15 [61], Spermatogenesis Online database [62], GermlncRNA [63], ReproGenomics Viewer [64], Gametogenesis Epigenetic Database (GED) [65], Ovarian Kaleidoscope Database, Endometrial Database, and HGEx-ERdb [66-68].

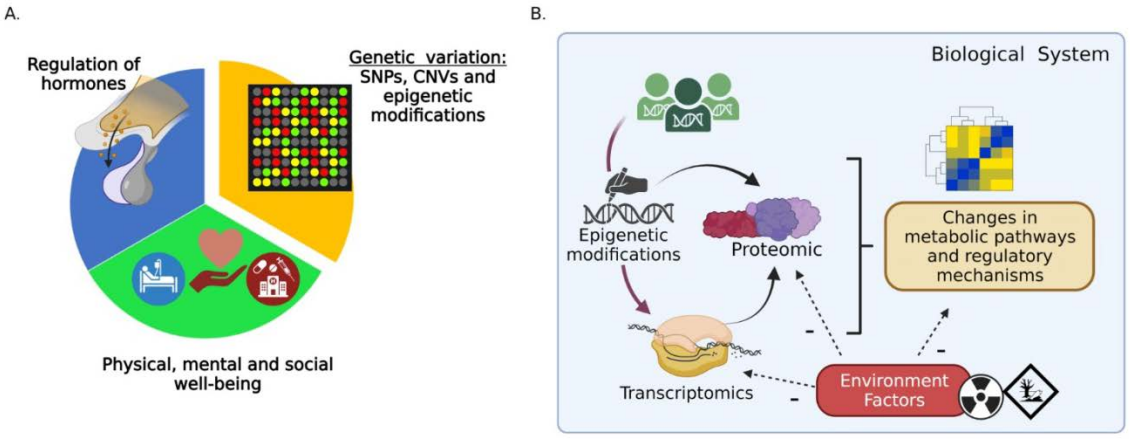

Fig. 1. (**A**) Reproductomics is the study of interactions between hormonal regulation of an individual and their environment, genetic vulnerability such as single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and epigenetic modifications, health impacts, and biological outcomes. (**B**) Bioinformatic analysis tools bring together/integrate the many omics levels in a systems biology approach. The importance of omics fields in biological processes for the research and understanding of biological systems is demonstrated. The genome is largely constant across cells and tissues, but the epigenome has low/moderate temporal variability and may impact both the transcriptome and the proteome. The transcriptome is very variable and is translated into the proteome tissue and physiological state specific, altering the metabolome in a tissue-specific way. Multiple variables influence the model, including proteome-influenced differential splicing, posttranslational protein modification, transcription factor binding, receptor ligand binding, and environmentally driven components.

This comprehensive assessment presents an extensive perspective on the transcriptomic landscape of the human endometrial milieu, encompassing details such as expression presence or the comparative expression magnitude of any previously characterized gene implicated in endometrial expression [7]. GEneSTATION aims to consolidate diverse omics data, primarily functional genomics, across placental mammals

(including humans) to enhance understanding of molecular mechanisms in gestation. The repository contains evolutionary, organismal, and molecular information for each gene, as well as links to other public databases, regardless of their pregnancy disorder focus [69]. Additionally, the database for preeclampsia, dbPEC serves as a repository for published literature and correlated genes pertaining to preeclampsia [70]. Uzun and colleagues recently developed a database focused on preeclampsia, a major factor in maternal and fetal death [70]. The Preterm Birth Database (dbPTB) compiles data on genomic variations linked to premature birth from public databases, expression array archives, and curated literature. This resource aids in better understanding the genetics behind preterm birth, whose cause is still not fully understood [71].

Omics data analysis with computational instruments

Irrespective of data origin, it is crucial to extract comprehensive biomedical information from public databases. While most omics high-throughput technology companies provide specific analytical software, several open-access resources enable information integration from these powerful tools. The TM4 microarray software suite represents a freely available tool for visualizing, analyzing, and integrating microarray data in an accessible environment [72]. The emergence of Next-Generation Sequencing (NGS) techniques, such as RNA-seq, necessitated specialized software like Chipster, a user-friendly platform for RNA-seq and microarray data analysis [73], and ReadXplorer, a free Java-based software for extensive genomic and transcriptomic NGS data analysis [74].

Omics technologies have led to increased reliance on computational systems, necessitating basic programming skills. R, Bioconductor packages, and Python are versatile languages for analyzing high-throughput data. Established in 2001, Bioconductor is an open-source project using R for cost-effective genomic data analysis [75,76]. R proves useful in bioinformatics and systems biology for omics data processing, statistical testing, and further analysis [31]. Python is popular in biomedicine, with libraries and toolkits extending its functionality to biological fields such as sequence analysis or omics data interpretation [77]. Omics data pipeline analysis generally involves preprocessing, normalization, exploratory analysis, statistical tests, and enrichment analysis, with additional steps based on data type and knowledge extraction methods (Fig. 2).

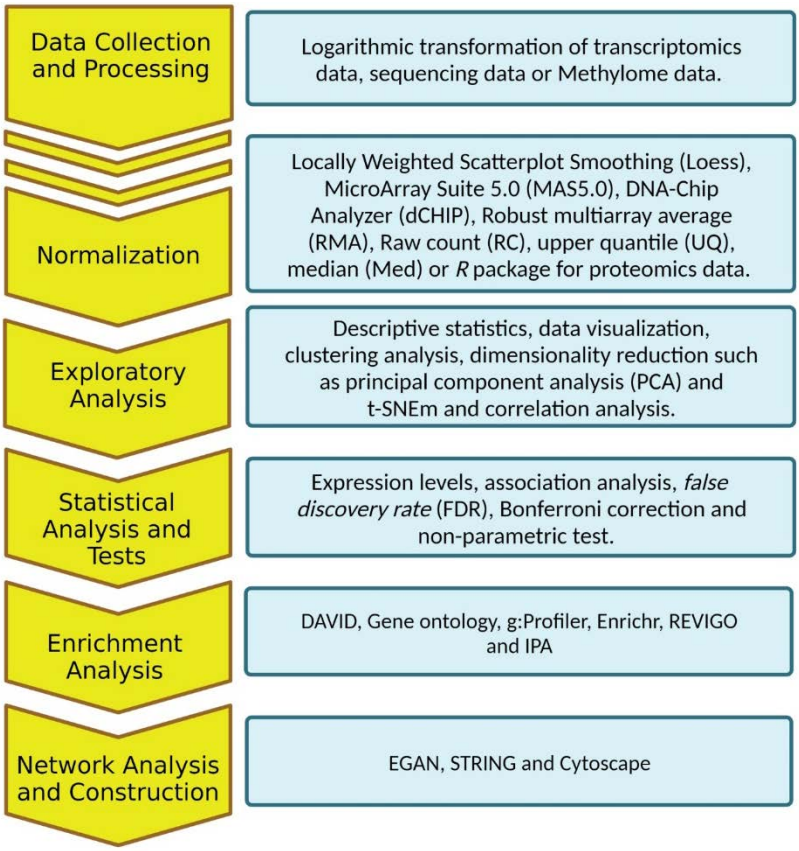

Fig. 2. Analysis pipeline as well as the instruments and software that are often used by systems biologists. After the data has been collected, data is subjected to various preliminary processing and normalization operations. These include an exploratory analysis, which is carried out in order to find prospective trends among the data; statistical tests, which allow for the detection of differentially expressed molecules in different situations; and, finally, an enrichment analysis, along with some network analysis and construction that are applied to the data in order to further elucidate biologically related issues.

Preprocessing: Reproductomics data analysis

Biological data exhibit substantial variability, exacerbated in complex contexts, necessitating preprocessing to eliminate technical outliers [78]. Preprocessing is contingent on researcher-defined criteria, informed by technology and data specifics. Comprehensive descriptions of these processes in scientific literature facilitate experimental reproducibility.

Various preprocessing methodologies are employed for omics data, dependent on data type. Fundamentally, these approaches mitigate potential confounders influencing result reliability. Logarithmic transformation is a prevalent method for homogenizing data by standardizing expression rate scales [79], facilitating transcriptomics data interpretation from microarrays [31]. Additional techniques include background correction, summarization, quality control, missing value estimation, background subtraction [80], and numerical feature discretization (e.g., hormone levels) [81].

Sequencing data (e.g., RNA-seq) analysis presents challenges due to numerous technological biases requiring identification and removal for accurate, reproducible results. Typical preprocessing stages include quality control, filtering, alignment, post-mapping quality control, counting, and pipeline execution using specialized programming languages [82-84].

Methylome data preprocessing involves addressing technical biases and batch effects by removing failed/SNP probes, thereby ensuring reliable, high-quality data [85]. Continued research is needed for optimal analysis of this intricate data. Proteomics data preprocessing, particularly for mass spectrometry-derived data, lacks standard methods and sequence; however, prevalent subtasks involve alignment, filtering, peak detection, and quantification [86].

Eliminating irrelevant discrepancies: Dataset normalization for comparability

The primary objective of the normalization step is to eliminate any discrepancies in the dataset that hold no biological relevance, ensuring comparability [87]. High-throughput omics technology data can be influenced by factors such as sample preparation, manufacturing processes, and molecular biases linked to certain analytical agents like fluorescent dyes [87]. Additionally, high-throughput technology data is often relative in nature, such as peptide abundance measured *via* mass spectrometry, necessitating the application of normalization methods to remove non-biological confounding factors impacting expression values of specific molecules like genes, proteins, miRNA, and metabolites [88].

Selecting an appropriate normalization method is crucial, as it depends on the platform type and analysis objective, which may significantly influence subsequent results [4]. Various methods, including quantiles, loess, dCHIP, and MAS5.0, were created to address systematic inherent errors, gaining popularity in microarray data analysis [4], while techniques like quantiles are increasingly employed in RNA-seq data analysis [5]. Among these, the robust multiarray average (RMA] is a popular and comprehensive microarray data analysis method, streamlining background adjustment, quantile normalization, log transformation, and probe set value summarization [6].

Normalization is crucial in RNA-seq data analysis, leading to the development of various techniques to tackle biases, such as RC, UQ, and Med [89]. Recently, machine learning-based methods like RSEM and Sailfish have emerged [89]. Yang *et al.* proposed an integrated approach for improved detection of Differentially Expressed Genes (DEGs) [90], but the optimal normalization method is still uncertain.

Concerning proteome data, normalization procedures are extensively employed to account for variations due to differential protein amounts across samples or protein degradation. Consequently, several algorithms encompassing both preprocessing and normalization steps have been established [86].

An appropriate normalization method is generally considered to be one that minimally alters the data. Some researchers suggest performing analyses with multiple normalization techniques and evaluating the results obtained [91]. Owing to the challenge of selecting the most suitable method, tools like the R package for proteomics data, Normalyzer (http://quantitativeproteomics.org/normalyzer), have been developed to address this issue [92].

Exploratory data analysis: Unveiling hidden patterns

Discovering latent patterns in data can elucidate unanticipated relationships among variables [93]. Multivariate analysis strives to simplify extensive datasets into fewer variables that capture the data's inherent structure using exploratory methods. Unsupervised techniques like hierarchical clustering, PCA, k-means clustering, consensus clustering, non-

negative matrix factorization, and mixture models are chiefly utilized in omics data exploration [94].

Clustering-techniques divide data objects into distinct groups based on similarity, emphasizing intra-group likeness and minimizing inter-group similarity [95]. Various metrics, such as Pearson's correlation coefficient, Jaccard index, Euclidean distance, and Manhattan distance, can be used to gauge similarity depending on the data's properties [79]. A detailed review of clustering methods can also be found [95]. Omics high-throughput data's high dimensionality necessitates dimensionality reduction techniques that project data onto a lower-dimensional basis [93]. PCA is a flexible unsupervised data reduction method that visualizes data structures by projecting high-dimensional data into a lower-dimensional space, clustering similar objects together [96].

Statistical tests

Statistical tests help identify molecules with specific behavior when comparing different situations, allowing the discovery of differences among experimental groups [97]. Early approaches relied on simple methods, testing associations based on data nature and dimensionality [93], and selecting molecules based on P-values, with values below 0.01 or 0.05 considered significant. High-throughput omics technologies produce high-dimensionality data, which can be biased and yield poor interpretability [98], necessitating multiple testing correction strategies [93,99]. Common methods include false discovery rate (FDR) control [93], Bonferroni correction [79], and nonparametric tests like rank product [100]. Omics technologies are rapidly evolving, requiring further development of specialized analytic tools and software to handle complex data.

Decoding biological processes through molecular enrichment analysis

Upon identifying differentially behaving molecules, subsequent enrichment analysis is necessary to ascertain associated biological processes [101,102]. Traditionally, this reductionist approach focused on individual molecules, utilizing tools like Entrez Gene, UniProt, GeneCards, or the Human Metabolome Database. High-throughput omics data, however, renders this method arduous and biased due to the vast number of molecules [79]. Consequently, several integrative tools emerged, including DAVID [103], Gene Ontology [104], g:Profiler [105], Enrichr [106], and REVIGO [107], streamlining the analysis process. Additionally, commercial platforms like Ingenuity Pathway Analysis (IPA) and Pathway Studio contribute significantly to enrichment analysis [108].

Integrative network analysis: Unraveling interactions in systems biology

Network representation, a prevalent approach in systems biology, encapsulates interactions by linking elements (nodes) *via* connections (edges) and is apt for integrating transcriptomics and proteomics data to delineate protein-protein and protein-DNA interactions [65]. This method effectively demonstrates biological data concerning gene regulation and cellular and signaling pathways, as quantitative information extracted from nodes allows for system importance inference [109]. Centrality indicators, including degree, betweenness, closeness, and Eigenvector centralities, facilitate the identification of crucial system nodes, or “hubs” [110]. Network and pathway-based data visualization and interpretation yield more insightful information than alternative representations [111], leading to the development of several computational tools for network analysis.

EGAN, a free Java desktop application, enables interactive visualization and interpretation of high-throughput assay results and automates metadata coincidence calculations [112]. Application of EGAN to Sertoli cells' inflammasome studies revealed DEGs involved in autophagy activation and innate immunity pathways, implicating Sertoli cells in male infertility pathogenesis *via* inflammatory cytokine induction [113].

STRING, a database comprising known and predicted protein-protein interactions from various sources [114], aids in gene function elucidation within a complex network. STRING has been employed to investigate DEGs in the endometrium during the implantation window, comparing natural and controlled ovarian stimulation cycles [115].

Cytoscape, an open-source platform for network visualization and analysis, offers a user-friendly interface and expanded functionality through numerous community-developed plugins [116]. Its versatility enables molecular interaction exploration related to reproduction, as evidenced by its application in herbal medicine-based anovulatory infertility network pharmacology [117], azoospermia protein-protein interaction network analysis [118], and a pioneering systems biology approach to human implantation processes.

Classifiers and predictors: Decoding normalcy and pathology in systems biology

In systems biology, identifying molecular classifiers or predictors distinguishes normal and pathological samples or forecasts data trends [119]. Numerous data mining and machine learning algorithms facilitate this process, warranting comprehensive explanation. The data classification process entails two primary steps: model construction and classification. The learning phase constructs classifiers using a training set comprising data tuples and associated class labels, employing classification algorithms. Accuracy estimation follows, determining classification rules' applicability to new data tuples. Popular classifiers include Support Vector Machines (SVM), random forests, Relevance Vector Machines (RVM), and logistic curves. Software such as WEKA facilitates execution of these algorithms [120]. Although primarily applied to gene expression and genomic profiling, metabolic pathways are increasingly considered [121]. Liu *et al.* developed a random forest-based method distinguishing pathogenic and normal SNPs [122]. Validated classifiers and predictors enable targeted therapies and personalized medicine development, applicable to reproductive research [123].

Conclusions and Perspectives

The nascent domain of reproductomics signifies a fundamental transformation in our comprehension of reproductive biology and its manifold processes. The methodical amalgamation of cutting-edge computational instruments and omics methodologies has given rise to an abundance of unparalleled revelations pertaining to the molecular mechanisms regulating reproduction, thus empowering the discernment of new therapeutic targets and diagnostic biomarkers. Additionally, the confluence of these methodologies has enabled a more exhaustive understanding of the intricate interplay among genetic, epigenetic, and environmental determinants in reproductive health and disease offering a cohesive perspective.

Contemporary computational strategies in reproductomics are propelling advancements in reproductive medicine and technology, with the resultant data poised to revolutionize ARTs. Although it is imperative to corroborate computational analyses with wet-lab experiments, systems biology presents an invaluable means of grappling with the intricacies inherent to human reproduction. A crucial aspect of the omics approach entails the appropriate analysis of the complex, high-throughput data produced. To surmount the obstacles and limitations inherent to omics data, numerous intelligent tools, including *in-silico* analyses, have been developed. Nonetheless, an amalgamation of consensus approaches and innovative techniques remains necessary.

Despite the extraordinary progress accomplished hitherto, numerous pivotal challenges and prospects for further investigation persist. Primarily, the creation of sturdy, high-resolution computational models adept at capturing the spatiotemporal dynamics of reproductive processes at both cellular and organismal echelons is of supreme significance. These models will require the integration of multi-omics data, encompassing genomics, transcriptomics, proteomics, and metabolomics, in conjunction with sophisticated machine learning and artificial intelligence algorithms to decipher intricate biological networks and their regulatory constituents.

Secondarily, the continuous advancement of single-cell omics techniques promises to provide a more detailed depiction of cellular heterogeneity and its implications for reproductive function. The employment of single-cell methodologies will promote the identification of hitherto uncharacterized cellular subpopulations and their functional contributions to reproduction, as well as explicate the molecular foundations of cellular plasticity and differentiation during gametogenesis, embryogenesis, and tissue regeneration.

Tertiarily, the flourishing domain of epigenomics proffers a plethora of opportunities for reproductomics research. Future inquiries should strive to demarcate the complex patterns of epigenetic regulation, encompassing DNA methylation, histone modifications, and non-coding RNA-mediated mechanisms, which modulate reproductive processes and contribute to the transgenerational transmission of epigenetic information. This knowledge will not only enhance our understanding of the causes of reproductive disorders but also guide the development of innovative therapeutic strategies targeting epigenetic dysregulation.

Lastly, a central objective of reproductomics research should be the translation of these scientific advancements into clinically actionable interventions. Establishing interdisciplinary collaborations among basic researchers, computational scientists, and clinicians will be crucial in this regard. Such partnerships will foster the development and validation of precision medicine approaches, including pharmacogenomics, gene editing, and regenerative medicine, tailored to the unique

molecular profiles of individual patients and their reproductive health requirements.

In summary, the future of reproductomics is filled with opportunities for significant advancements in our understanding of reproductive biology and the development of novel therapeutic and diagnostic approaches. The continued enhancement and integration of computational tools and omics technologies will unquestionably revolutionize the field and lay the groundwork for the emergence of a new epoch in reproductive medicine, one that emphasizes precision, personalization, and a profound grasp of the molecular intricacies that underlie human reproduction.

Conflict of Interest

There is no conflict of interest.

Acknowledgements

The work was supported by Erasmus+ KA220-HED Cooperation Partnerships in higher education "Fostering Internationalisation in Higher Education by BioFood Virtual Labs" (2021-1-SK01-KA220-HED-000032062).

Abbreviations

ART, Assisted Reproductive Technology; dbPTB, Preterm Birth Database; CNV, Copy Number Variation; DEG, Differentially Expressed Gene; EMBL-EBI, European Molecular Biology Laboratory-European Bioinformatics Institute; GED, Gametogenesis Epigenetic Database; GEO, Gene Expression Omnibus; GWAS, Genome-Wide Association Studies; HGEx-ERdb, Human Gene Expression Endometrial Receptivity Database; IPA, Ingenuity Pathway Analysis; IVF, *in vitro* Fertilization; miRNA, microRNA; MSC-EV, Mesenchymal Stem Cell-Derived Extracellular Vesicle; NGS, Next-Generation Sequencing; POI, Premature Ovarian Insufficiency; RVM, Relevance Vector Machines; SNP, Single Nucleotide Polymorphism; SVM, Support Vector Machines

References

1. Dai X, Shen L. Advances and trends in omics technology development. Front Med (Lausanne) 2022;9:911861. https://doi.org/10.3389/fmed.2022.911861
2. Miroslava R, Katarína Š, Ivana Š, Peter U. Omics applications in reproductive medicine. In: *Biotechnology in Healthcare Volume 1*. BARH D (ed.), Academic Press, Elsevier, London, 2022, pp 103-123. https://doi.org/10.1016/B978-0-323-89837-9.00002-4
3. Chicco D. Gene expression from GEO: An R package to facilitate data reading from Gene Expression Omnibus (GEO). Methods Mol Biol 2022;2401:187-194. https://doi.org/10.1007/978-1-0716-1839-4_12
4. Baker M. Gene data to hit milestone. Nature 2012;487:282-283. https://doi.org/10.1038/487282a
5. Tapia A, Vilos C, Marín JC, Croxatto HB, Devoto L. Bioinformatic detection of E47, E2F1 and SREBP1 transcription factors as potential regulators of genes associated to acquisition of endometrial receptivity. Reprod Biol Endocrinol 2011;9:1-14. https://doi.org/10.1186/1477-7827-9-14
6. Zhang D, Sun C, Ma C, Dai H, Zhang W. Data mining of spatial-temporal expression of genes in the human endometrium during the window of implantation. Reprod Sci 2012;19:1085-1098. https://doi.org/10.1177/1933719112442248
7. Bhagwat SR, Chandrashekar DS, Kakar R, Davuluri S, Bajpai AK, Nayak S, Bhutada S, Acharya K, Sachdeva G. Endometrial receptivity: a revisit to functional genomics studies on human endometrium and creation of HGEx-ERdb. PLoS One 2013;8:e58419. https://doi.org/10.1371/journal.pone.0058419
8. Talbi S, Hamilton A, Vo K, Tulac S, Overgaard MT, Dosiou C, Le Shay N, Nezhat CN, Kempson R, Lessey BA, Nayak NR, Giudice LC. Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. Endocrinol 2006;147:1097-1121. https://doi.org/10.1210/en.2005-1076
9. Burney RO, Talbi S, Hamilton AE, Vo KC, Nyegaard M, Nezhat CR, Lessey BA, Giudice LC. Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. Endocrinol 2007;148:3814-3826. https://doi.org/10.1210/en.2006-1692

10. Hever A, Roth RB, Hevezi P, Marin ME, Acosta JA, Acosta H, Rojas J, ET AL. Human endometriosis is associated with plasma cells and overexpression of B lymphocyte stimulator. Proc Natl Acad Sci U S A 2007;104:12451-12456. https://doi.org/10.1073/pnas.0703451104
11. Altmäe S, Esteban FJ, Stavreus-Evers A, Simon C, Giudice L, Lessey BA, Horcajadas JA, ET AL. Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium. Hum Reprod Update 2014;20:12-28. https://doi.org/10.1093/humupd/dmt048
12. Wang Y-F, Chang M-Y, Chiang R-D, Hwang L-J, Lee C-M, Wang Y-H. Mining medical data: a case study of endometriosis. J Med Sys 2013;37:1-7. https://doi.org/10.1007/s10916-012-9899-y
13. Mathew D, Drury J, Valentijn A, Vasieva O, Hapangama D. In silico, in vitro and in vivo analysis identifies a potential role for steroid hormone regulation of FOXD3 in endometriosis-associated genes. Hum Reprod 2016;31:345-354. https://doi.org/10.1093/humrep/dev307
14. Liu J-L, Zhao M. A PubMed-wide study of endometriosis. Genomics 2016;108:151-157. https://doi.org/10.1016/j.ygeno.2016.10.003
15. Vincent ZL, Farquhar CM, Mitchell MD, Ponnampalam AP. Expression and regulation of DNA methyltransferases in human endometrium. Fertil Steril 2011;95:1522-5.e1. https://doi.org/10.1016/j.fertnstert.2010.09.030
16. Caplakova V, Babusikova E, Blahovcova E, Balharek T, Zelieskova M, Hatok J. DNA methylation machinery in the endometrium and endometrial cancer. Anticancer Res 2016;36:4407-4420. https://doi.org/10.21873/anticanres.10984
17. Houshdaran S, Zelenko Z, Irwin JC, Giudice LC. Human endometrial DNA methylome is cycle-dependent and is associated with gene expression regulation. Mol Endocrinol 2014;28:1118-1135. https://doi.org/10.1210/me.2013-1340
18. Saare M, Modhukur V, Suhorutshenko M, Rajashekar B, Rekker K, Sõritsa D, Karro H, ET AL. The influence of menstrual cycle and endometriosis on endometrial methylome. Clin Epigenetics 2016;8:2. https://doi.org/10.1186/s13148-015-0168-z
19. Kukushkina V, Modhukur V, Suhorutšenko M, Peters M, Mägi R, Rahmioglu N, Velthut-Meikas A, Altmäe S, Esteban FJ, Vilo J, Zondervan K, Salumets A, Laisk-Podar T. DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. Sci Rep 2017;7:3916. https://doi.org/10.1038/s41598-017-03682-0
20. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 2014;15:R37. https://doi.org/10.1186/gb-2014-15-2-r37
21. Wan J, Oliver VF, Wang G, Zhu H, Zack DJ, Merbs SL, Qian J. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics 2015;16:49. https://doi.org/10.1186/s12864-015-1271-4
22. Bunkar N, Pathak N, Lohiya NK, Mishra PK. Epigenetics: A key paradigm in reproductive health. Clin Exp Reprod Med 2016;43:59-81. https://doi.org/10.5653/cerm.2016.43.2.59
23. Walker E, Hernandez AV, Kattan MW. Meta-analysis: Its strengths and limitations. Cleveland Clin J Med 2008;75:431-439. https://doi.org/10.3949/ccjm.75.6.431
24. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 2012;28:573-580. https://doi.org/10.1093/bioinformatics/btr709
25. Altmäe S, Koel M, Võsa U, Adler P, Suhorutšenko M, Laisk-Podar T, Kukushkina V, ET AL. Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers. Sci Rep 2017;7:10077. https://doi.org/10.1038/s41598-017-10098-3
26. Horcajadas J, Pellicer A, Simon C. Wide genomic analysis of human endometrial receptivity: new times, new opportunities. Hum Reprod Update 2007;13:77-86. https://doi.org/10.1093/humupd/dml046
27. Tseng L-H, Chen I, Chen M-Y, Yan H, Wang C-N, Lee C-L. Genome-based expression profiling as a single standardized microarray platform for the diagnosis of endometrial disorder: an array of 126-gene model. Fertil Steril 2010;94:114-119. https://doi.org/10.1016/j.fertnstert.2009.01.130

28. Rahmioglu N, Nyholt DR, Morris AP, Missmer SA, Montgomery GW, Zondervan KT. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. Hum Reprod Update 2014;20:702-716. https://doi.org/10.1093/humupd/dmu015
29. Bagheri M, Khansarinejad B, Mondanizadeh M, Azimi M, Alavi S. miRNAs related in signaling pathways of women's reproductive diseases: an overview. Mol Biol Rep 2024;51:414. https://doi.org/10.1007/s11033-024-09357-0
30. Zhou Y, Li Q, You S, Jiang H, Jiang L, He F, Hu L. Efficacy of mesenchymal stem cell-derived extracellular vesicles in the animal model of female reproductive diseases: A meta-analysis. Stem Cell Rev Rep 2023;19:2299-2310. https://doi.org/10.1007/s12015-023-10576-4
31. Robinson SW, Fernandes M, Husi H. Current advances in systems and integrative biology. Comput Struct Biotechnol J 2014;11:35-46. https://doi.org/10.1016/j.csbj.2014.08.007
32. Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, Sansone SA, ET AL. Data standards for omics data: the basis of data sharing and reuse. Methods Mol Biol 2011;719:31-69. https://doi.org/10.1007/978-1-61779-027-0_2
33. Altmäe S, Reimand J, Hovatta O, Zhang P, Kere J, Laisk T, Saare M, ET AL. Research resource: interactome of human embryo implantation: identification of gene expression pathways, regulation, and integrated regulatory networks. Mol Endocrinol 2012;26:203-217. https://doi.org/10.1210/me.2011-1196
34. Gracie S, Pennell C, Ekman-Ordeberg G, Lye S, McManaman J, Williams S, Palmer L, ET AL. An integrated systems biology approach to the study of preterm birth using"-omic" technology-a guideline for research. BMC Pregnancy Childbirth 2011;11:71. https://doi.org/10.1186/1471-2393-11-71
35. Mayhew T. Morphomics: An integral part of systems biology of the human placenta. Placenta 2015;36:329-340. https://doi.org/10.1016/j.placenta.2015.01.001
36. Jumeau F, Chalmel F, Fernandez-Gomez F-J, Carpentier C, Obriot H, Tardivel M, Caillet-Boudin ML, ET AL. Defining the human sperm microtubulome: an integrated genomics approach. Biol Reprod 2017;96:93-106. https://doi.org/10.1095/biolreprod.116.143479
37. Ghosh D, Sengupta J. A systems biology approach to elucidate the process of blastocyst implantation. Indian J Physiol Pharmacol 2010;54:41-50.
38. Diaz-Beltran L, Cano C, Wall DP, Esteban FJ. Systems biology as a comparative approach to understand complex gene expression in neurological diseases. Behav Sci 2013;3:253-272. https://doi.org/10.3390/bs3020253
39. Kyrgiou M, Pouliakis A, Panayiotides JG, Margari N, Bountris P, Valasoulis G, Paraskevaidi M, ET AL. Personalised management of women with cervical abnormalities using a clinical decision support scoring system. Gynecol Oncol 2016;141:29-35. https://doi.org/10.1016/j.ygyno.2015.12.032
40. Seli E, Robert C, Sirard M-A. OMICS in assisted reproduction: possibilities and pitfalls. Mol Hum Reprod 2010;16:513-530. https://doi.org/10.1093/molehr/gaq041
41. Haouzi D, Dechaud H, Assou S, Monzo C, De Vos J, Hamamah S. Transcriptome analysis reveals dialogues between human trophectoderm and endometrial cells during the implantation period. Hum Reprod 2011;26:1440-1449. https://doi.org/10.1093/humrep/der075
42. Aghajanova L, Shen S, Rojas AM, Fisher SJ, Irwin JC, Giudice LC. Comparative transcriptome analysis of human trophectoderm and embryonic stem cell-derived trophoblasts reveal key participants in early implantation. Biol Reprod 2012;86:1-21. https://doi.org/10.1095/biolreprod.111.092775
43. Leung A, Bader GD, Reimand J. HyperModules: identifying clinically and phenotypically significant network modules with disease mutations for biomarker discovery. Bioinformatics 2014;30:2230-2232. https://doi.org/10.1093/bioinformatics/btu172
44. Silva JV, Yoon S, Domingues S, Guimarães S, Goltsev AV, da Cruz E Silva EF, Mendes JF, ET AL. Amyloid precursor protein interaction network in human testis: sentinel proteins for male reproduction. BMC Bioinformatics 2015;16:12. https://doi.org/10.1186/s12859-014-0432-9
45. Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, Wong MP, ET AL. GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res 2012;40:D1047-D1054. https://doi.org/10.1093/nar/gkr1182

46. Beck T, Hastings RK, Gollapudi S, Free RC, Brookes AJ. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. European J Hum Genet 2014;22:949-952. https://doi.org/10.1038/ejhg.2013.274
47. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, Garvin T, ET AL. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS One 2013;8:e81148. https://doi.org/10.1371/journal.pone.0081148
48. Xiong Y, Wei Y, Gu Y, Zhang S, Lyu J, Zhang B, Chen C, ET AL. DiseaseMeth version 2.0: a major expansion and update of the human disease methylation database. Nucleic Acids Res 2017;45:D888-D895. https://doi.org/10.1093/nar/gkw1123
49. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2007;36:D154-D158. https://doi.org/10.1093/nar/gkm952
50. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA. org resource: targets and expression. Nucleic Acids Res 2008;36:D149-D153. https://doi.org/10.1093/nar/gkm995
51. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-210. https://doi.org/10.1093/nar/30.1.207
52. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, ET AL. ArrayExpress-a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 2007;35:D747-D750. https://doi.org/10.1093/nar/gkl995
53. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, ET AL. The PeptideAtlas project. Nucleic Acids Res 2006;34:D655-D658. https://doi.org/10.1093/nar/gkj040
54. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, ET AL. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 2015;44:D447-D456. https://doi.org/10.1093/nar/gkv1145
55. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, Ahmad Khan A, ET AL. Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res 2014;42:D959-D965. https://doi.org/10.1093/nar/gkt1251
56. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, ET AL. A draft map of the human proteome. Nature 2014;509:575-581. https://doi.org/10.1038/nature13302
57. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, ET AL. Mass-spectrometry-based draft of the human proteome. Nature 2014;509:582-587. https://doi.org/10.1038/nature13319
58. Yu W, Clyne M, Khoury MJ, Gwinn M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics 2010;26:145-146. https://doi.org/10.1093/bioinformatics/btp618
59. Tarín JJ, García-Pérez MA, Hamatani T, Cano A. Infertility etiologies are genetically and clinically linked with other diseases in single meta-diseases. Reprod Biol Endocrinol 2015;13:31. https://doi.org/10.1186/s12958-015-0029-9
60. Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A. OMICtools: an informative directory for multi-omic data analysis. Database (Oxford) 2014;2014:bau069. https://doi.org/10.1093/database/bau069
61. Hua J, Xu B, Yang Y, Ban R, Iqbal F, Cooke HJ, Zhang Y, Shi Q. Follicle Online: an integrated database of follicle assembly, development and ovulation. Database (Oxford) 2015;2015:bav036. https://doi.org/10.1093/database/bav036
62. Zhang Y, Zhong L, Xu B, Yang Y, Ban R, Zhu J, Cooke HJ, ET AL. SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. Nucleic Acids Res 2013;41:D1055-D1062. https://doi.org/10.1093/nar/gks1186
63. Luk AC-S, Gao H, Xiao S, Liao J, Wang D, Tu J, Rennert OM, ET AL. GermlncRNA: a unique catalogue of long non-coding RNAs and associated regulations in male germ cell development. Database (Oxford) 2015;2015:bav044. https://doi.org/10.1093/database/bav044
64. Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Le Bras Y, Jégou B, ET AL. The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. Nucleic Acids Res 2015;43:W109-W116. https://doi.org/10.1093/nar/gkv345
65. Bai W, Yang W, Wang W, Wang Y, Liu C, Jiang Q, Hua J, Liao M. GED: a manually curated comprehensive resource for epigenetic modification of gametogenesis. Brief Bioinform 2017;18:98-104. https://doi.org/10.1093/bib/bbw007

66. Leo CP, Vitt UA, Hsueh AJ. The Ovarian Kaleidoscope database: an online resource for the ovarian research community. Endocrinology 2000;141:3052-3054. https://doi.org/10.1210/endo.141.9.7679
67. Ben-Shlomo I, Vitt UA, Hsueh AJ. Perspective: the Ovarian Kaleidoscope database-II. Functional genomic analysis of an organ-specific database. Endocrinol 2002;143:2041-2044. https://doi.org/10.1210/endo.143.6.8851
68. Hsueh AJ, Rauch R. Ovarian Kaleidoscope database: ten years and beyond. Biol Reprod 2012;86:192. https://doi.org/10.1095/biolreprod.112.099127
69. Kim M, Cooper BA, Venkat R, Phillips JB, Eidem HR, Hirbo J, Nutakki S, ET AL. GEneSTATION 1.0: a synthetic resource of diverse evolutionary and functional genomic data for studying the evolution of pregnancy-associated tissues and phenotypes. Nucleic Acids Res 2016;44:D908-D916. https://doi.org/10.1093/nar/gkv1137
70. Uzun A, Triche EW, Schuster J, Dewan AT, Padbury JF. dbPEC: a comprehensive literature-based database for preeclampsia related genes and phenotypes. Database (Oxford) 2016;2016:baw006. https://doi.org/10.1093/database/baw006
71. Uzun A, Laliberte A, Parker J, Andrew C, Winterrowd E, Sharma S, Istrail S, Padbury JF. dbPTB: a database for preterm birth. Database (Oxford) 2012;2012:bar069. https://doi.org/10.1093/database/bar069
72. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, ET AL. TM4 microarray software suite. Methods Enzymol 2006;411:134-193. https://doi.org/10.1016/S0076-6879(06)11009-5
73. Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, Koski M, ET AL. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics 2011;12:507. https://doi.org/10.1186/1471-2164-12-507
74. Hilker R, Stadermann KB, Doppmeier D, Kalinowski J, Stoye J, Straube J, Winnebald J, Goesmann A. ReadXplorer-visualization and analysis of mapped sequences. Bioinformatics 2014;30:2247-2254. https://doi.org/10.1093/bioinformatics/btu205
75. Reimers M, Carey VJ. Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 2006;411:119-134. https://doi.org/10.1016/S0076-6879(06)11008-3
76. Zhang Y, Szustakowski J, Schinke M. Bioinformatics analysis of microarray data. Methods Mol Biol 2009;573:259-284. https://doi.org/10.1007/978-1-60761-247-6_15
77. Ekmekci B, McAnany CE, Mura C. An introduction to programming for bioscientists: a Python-based primer. PLoS Comput Biol 2016;12:e1004867. https://doi.org/10.1371/journal.pcbi.1004867
78. Rausch TK, Schillert A, Ziegler A, Lüking A, Zucht H-D, Schulz-Knappe P. Comparison of pre-processing methods for multiplex bead-based immunoassays. BMC Genomics 2016;17:601. https://doi.org/10.1186/s12864-016-2888-7
79. Esteban F, Cano C, De la Haza I, Cano-Ortiz A, Vélez de Mendizábal N, Goñi J, Horcajadas JA. Análisis bioinformático de datos: aplicación en microarrays. CMR 2008;14:87-96.
80. Sui Y, Zhao X, Speed TP, Wu Z. Background adjustment for DNA microarrays using a database of microarray experiments. J Comput Biol 2009;16:1501-1515. https://doi.org/10.1089/cmb.2009.0063
81. Mirroshandel SA, Ghasemian F, Monji-Azad S. Applying data mining techniques for increasing implantation rate by selecting best sperms for intra-cytoplasmic sperm injection treatment. Comput Methods Programs Biomed 2016;137:215-229. https://doi.org/10.1016/j.cmpb.2016.09.013
82. Goncalves A, Tikhonov A, Brazma A, Kapushesky M. A pipeline for RNA-seq data processing and quality assessment. Bioinformatics 2011;27:867-869. https://doi.org/10.1093/bioinformatics/btr012
83. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res 2016;5:ISCB Comm J-1408. https://doi.org/10.12688/f1000research.9005.2
84. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 2017;33:1179-1186. https://doi.org/10.1093/bioinformatics/btw777
85. Cazaly E, Thomson R, Marthick JR, Holloway AF, Charlesworth J, Dickinson JL. Comparison of pre-processing methodologies for Illumina 450k methylation array data in familial analyses. Clin Epigenetics 2016;8:75. https://doi.org/10.1186/s13148-016-0241-2
86. Cruz-Marcelo A, Guerra R, Vannucci M, Li Y, Lau CC, Man T-K. Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. Bioinformatics 2008;24:2129-2136. https://doi.org/10.1093/bioinformatics/btn398

87. Egea RR, Puchalt NG, Escrivá MM, Varghese AC. OMICS: current and future perspectives in reproductive medicine and technology. J Hum Reprod Sci 2014;7:73-92. https://doi.org/10.4103/0974-1208.138857
88. Silvestri E, Lombardi A, de Lange P, Glinni D, Senese R, Cioffi F, Lanni A, ET AL. Studies of complex biological systems with applications to molecular medicine: the need to integrate transcriptomic and proteomic approaches.
J Biomed Biotechnol 2011:810242. https://doi.org/10.1155/2011/810242
89. Li P, Piao Y, Shon HS, Ryu KH. Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. BMC Bioinformatics 2015;16:347. https://doi.org/10.1186/s12859-015-0778-7
90. Yang S, Mercante DE, Zhang K, Fang Z. An integrated approach for RNA-seq data normalization. Cancer Inform 2016;15:129-141. https://doi.org/10.4137/CIN.S39781
91. Borgaonkar SP, Hocker H, Shin H, Markey MK. Comparison of normalization methods for the identification of biomarkers using MALDI-TOF and SELDI-TOF mass spectra. OMICS 2010;14:115-126. https://doi.org/10.1089/omi.2009.0082
92. Chawade A, Alexandersson E, Levander F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. J Proteome Res 2014;13:3114-3120. https://doi.org/10.1021/pr401264n
93. Chadeau-Hyam M, Campanella G, Jombart T, Bottolo L, Portengen L, Vineis P, Liquet B, Vermeulen RC. Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. Environ Mol Mutagen 2013;54:542-557. https://doi.org/10.1002/em.21797
94. Wagner F. GO-PCA: An unsupervised method to explore gene expression data using prior knowledge. PLoS One 2015;10:e0143196. https://doi.org/10.1371/journal.pone.0143196
95. Jiang D, Tang C, Zhang A. Cluster analysis for gene expression data: a survey. IEEE Trans Knowl Data Eng 2004;16:1370-1386. https://doi.org/10.1109/TKDE.2004.68
96. Altmäe S, Tamm-Rosenstein K, Esteban FJ, Simm J, Kolberg L, Peterson H, Metsis M, ET AL. Endometrial transcriptome analysis indicates superiority of natural over artificial cycles in recurrent implantation failure patients undergoing frozen embryo transfer. Reprod Biomed Online 2016;32:597-613. https://doi.org/10.1016/j.rbmo.2016.03.004
97. Hatfield GW, Hung Sp, Baldi P. Differential analysis of DNA microarray gene expression data. Mol Microbiol 2003;47:871-877. https://doi.org/10.1046/j.1365-2958.2003.03298.x
98. Guo Y, Graber A, McBurney RN, Balasubramanian R. Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinformatics 2010;11:447. https://doi.org/10.1186/1471-2105-11-447
99. Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics 2015;8:33. https://doi.org/10.1186/s12920-015-0108-y
100. Franco D, Bonet F, Hernandez-Torres F, Lozano-Velasco E, Esteban FJ, Aranega AE. Analysis of microRNA microarrays in cardiogenesis. Methods Mol Biol 2016;1375:207-221. https://doi.org/10.1007/7651_2015_247
101. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, ET AL. Pathway enrichment analysis and visualization of omics data using g: Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 2019;14:482-517. https://doi.org/10.1038/s41596-018-0103-9
102. Paczkowska M, Barenboim J, Sintupisut N, Fox NS, Zhu H, Abd-Rabbo D, Mee MW, ET AL. Integrative pathway enrichment analysis of multivariate omics data. Nat Commun 2020;11:735. https://doi.org/10.1038/s41467-019-13983-9
103. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57. https://doi.org/10.1038/nprot.2008.211
104. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, ET AL. Gene ontology: tool for the unification of biology. Nat Genet 2000;25:25-29. https://doi.org/10.1038/75556
105. Reimand J, Arak T, Vilo J. g: Profiler-a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res 2011;39:W307-W315. https://doi.org/10.1093/nar/gkr378
106. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013;14:128. https://doi.org/10.1186/1471-2105-14-128

107. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 2011;6:e21800. https://doi.org/10.1371/journal.pone.0021800
108. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio-the analysis and navigation of molecular networks. Bioinformatics 2003;19:2155-2157. https://doi.org/10.1093/bioinformatics/btg290
109. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8(Suppl 4):S11. https://doi.org/10.1186/1752-0509-8-S4-S11
110. Santiago JA, Bottero V, Potashkin JA. Dissecting the molecular mechanisms of neurodegenerative diseases through network biology. Frontiers Aging Neurosci 2017;9:166. https://doi.org/10.3389/fnagi.2017.00166
111. Gehlenborg N, O'donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, ET AL. Visualization of omics data for systems biology. Nat Methods 2010;7(3 Suppl):S56-S68. https://doi.org/10.1038/nmeth.1436
112. Paquette J, Tokuyasu T. EGAN: exploratory gene association networks. Bioinformatics 2010;26:285-286. https://doi.org/10.1093/bioinformatics/btp656
113. Hayrabedyan S, Todorova K, Jabeen A, Metodieva G, Toshkov S, Metodiev MV, Mincheff M, Fernández N. Sertoli cells have a functional NALP3 inflammasome that can modulate autophagy and cytokine production. Sci Rep 2016;6:18896. https://doi.org/10.1038/srep18896
114. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, ET AL. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-D452. https://doi.org/10.1093/nar/gku1003
115. Horcajadas JA, Mínguez P, Dopazo J, Esteban FJ, Domínguez F, Giudice LC, Pellicer A, Simón C. Controlled ovarian stimulation induces a functional genomic delay of the endometrium with potential clinical implications. J Clin Endocrinol Metab 2008;93:4500-4510. https://doi.org/10.1210/jc.2008-0588
116. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-2504. https://doi.org/10.1101/gr.1239303
117. Liu H, Zeng L, Yang K, Zhang G. A network pharmacology approach to explore the pharmacological mechanism of xiaoyao powder on anovulatory infertility. Evid Based Complement Alternat Med 2016;2016:2960372. https://doi.org/10.1155/2016/2960372
118. Sabetian S, Shamsir MS. Systematic analysis of protein interaction network associated with azoospermia. Int J Mol Sci 2016;17:1857. https://doi.org/10.3390/ijms17111857
119. Biswas N, Chakrabarti S. Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. Front Oncol 2020;10:588221. https://doi.org/10.3389/fonc.2020.588221
120. Frank E, Hall MA, Witten IH. *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition". Morgan Kaufmann, The University of Waikato, New Zealand, 2016, p 128.
121. Hung F-H, Chiu H-W. Cancer subtype prediction from a pathway-level perspective by using a support vector machine based on integrated gene expression and protein network. Comput Methods Programs Biomed 2017;141:27-34. https://doi.org/10.1016/j.cmpb.2017.01.006
122. Liu Q, Gan M, Jiang R. A sequence-based method to predict the impact of regulatory variants using random forest. BMC Syst Biol 2017;11:7. https://doi.org/10.1186/s12918-017-0389-1
123. Way GP, Allaway RJ, Bouley SJ, Fadul CE, Sanchez Y, Greene CS. A machine learning classifier trained on cancer transcriptomes detects NF1 inactivation signal in glioblastoma. BMC Genomics 2017;18:127. https://doi.org/10.1186/s12864-017-3519-7