

School of Medicine

Department of Health Sciences

Master degree in Medical Biotechnologies

Laboratory of Biomedical Materials

Distance-Based Computational Analysis of Stem Cell Transcriptomic and Proteomic Responses to Zirconia doped with silver Coatings for Temporary Fixation Devices

Supervisor:
Dr.Mauro Nascimben

Academic Mentor:
Prof. Lia Rimondini, Prof. Anderia Cochis

Firmato digitalmente da Andrea Cochis
Data: 29.09.2025 12:17:53 CEST
Organizzazione: UNIVERSITA' DEGLI STUDI DEL PIEMONTE ORIENTALE/01943490027

MSc student: Fawzia Ibrahim

Matricula: 20058753

Academic year 2024/2025

Contents

Abbreviations 4
Abstract 6
1. Introduction 7
1.1. Transcriptomic data analysis and its challenges 9
1.2. Proteomic data analysis and its challenges 14
2. The aim of the study 18
3. Materials and Methods 19
3.1 Cell Lines and Experimental Conditions 19
3.2 Data Preprocessing 19
1.3. Distance-Based Analysis of Transcriptomic and Proteomic Data 20
3.1.1. Transcriptomic Analysis 21
3.1.1.1. Distance Matrix Calculation and Ratio 21
3.1.1.2. Selection of Top Gene Pairs 21
3.1.1.3. Functional Enrichment Analysis 21
3.1.2. Proteomic Analysis 22
3.1.2.1. Distance Matrix Calculation and Transformation 22
3.1.2.2. Distance Ratio Calculation and Clustering 22
3.1.2.3. Selection of Protein Pairs 22
3.1.2.4. Functional Enrichment Analysis 22
4. Results 23
1.4. Data Overview and Quality Assessment 23
4.1.1. Distribution of Expression Data (Boxplots and Statistical Evaluation) 23
4.1.2. Data Normalization and Outlier Evaluation 24
4.1.3. Correlation Assessment of Biological Replicates 24
1.5. Distance-Based Analysis 27
4.1.4. Transcriptomic Distance Matrix Analysis 27
4.1.5. Proteomic distance ratio analysis and functional enrichment 30
5. Discussion 34
6. Conclusion 37
7. Bibliography 38

Abbreviations

A

Artificial Intelligence
AI · 10

C

Cross normalization
CrossNorm · 13

D

deep learning
DL · 14
Deep Multiple Reaction Monitoring
DeepMRM · 17
deep neural networks
DNN · 14

G

Geen Ontology
GO · 21
Genotype-Tissue Expression
GTex · 10

I

Informative cross normalization
ICN · 13
interquartile range
IQR · 23

K

Kernel density weighted loess normalization
KDWL · 13

M

machine learning
ML · 11
mass spectrometry
MS · 8
Modified least-variant set normalization
LVSmiR · 13

N

next-generation sequencing
NGS · 8

O

one- or two-dimensional gel
1D or 2D · 15
one-class support vector machine
SVM · 19

R

Ras homolog family member A
RhoA · 35
reactive oxygen species
ROS · 8
RNA sequencing
RNA-seq · 9

S

silver
Ag · 8

T

Talin
TLN · 34
the Cancer Cell Line Encyclopedia

CCLE · 10
The Cancer Genome Atlas
TCGA · 10
the Sequence Read Archive
SRA · 10
titanium
Ti · 18

Z

zirconia doped with silver
ZrO_2–Ag · 8

Abstract

Temporary fixation devices must balance two crucial needs: preventing infection and supporting healthy bone healing. These devices should encourage the right amount of cell adhesion-enough to let bone repair itself, but not so much that the device becomes difficult to remove later. To explore how to achieve this balance, this thesis looked at how human mesenchymal stem cells (hMSCs) interact with titanium surfaces coated with zirconia doped with silver (ZrO_2–Ag). Rather than just measuring whether certain genes or proteins went up or down, a new distance-based computational method was used to see how relationships between them changed. This method analyzed over five million gene pairs and more than 700 protein pairs, tracking shifts in how closely related they were and what biological functions were affected.

The findings showed that adding silver did consistently reshape how genes and proteins connect, especially in pathways related to cell adhesion and the cytoskeleton. Interestingly, both stronger and weaker molecular interactions still centered around the same adhesion-related functions, and some genes and proteins acted as central hubs whose roles shifted depending on the context. While the protein data set was smaller and had more variability, it backed up the results seen in gene expression and confirmed a clear remodeling of adhesion and cytoskeletal networks.

These results build on earlier research by revealing that silver influences how cells interact with materials, not just by turning adhesion molecules on or off, but by dynamically reorganizing the networks they belong to. In summary, this thesis shows that zirconia-silver coatings offer both infection protection and adaptable cellular responses, making them promising for temporary fixation devices. The distance-matrix method proved especially useful for detecting these subtle but important changes, highlighting its potential as a tool for future biomaterials research.

1.Introduction

Implants consist of biological or artificial materials that are inserted or grafted into the body [1]. Their use can be traced back to ancient Egyptian civilizations and cultures in South-Central America. Thanks to advancements in material and biological sciences, remarkable progress has been achieved in this field. Improvements in the quality and availability of implant materials have made this treatment option extremely promising, innovative, and widely adopted in contemporary practice [2].

As implants became more sophisticated, it became clear that their success depends not only on the bulk material but also on the properties of their surfaces. In particular, the way cells and bacteria interact with implant surfaces plays a crucial role in determining outcomes such as tissue integration and infection risk [3]. Consequently, a significant amount of research has focused on developing pro-regenerative and antibacterial surface modifications, especially for permanent implants [4], [5]. However, comparatively little attention has been given to temporary implants. These devices, which are used briefly to stabilize fractures in polytrauma patients before being removed for definitive treatment, pose unique requirements [6].

Consequently, materials intended for temporary fixation devices must not encourage cell adhesion or proliferation (unlike permanent implants), as removing them before implanting the prosthesis might harm the bone if tissue ingrowth has taken place [7]; however, these materials must still be cytocompatible to avoid the release of harmful compounds at the healing site. The primary concerns related to temporary fixation devices are bacterial contamination and excessive tissue growth. To address these challenges, certain studies emphasize the importance of surface roughness in tissue adhesion, proposing the creation of notably smooth surfaces to hinder tissue overgrowth on temporary fixation devices [8], [9]. Additionally, it can be inferred that surface roughness significantly impacts bacterial adhesion as well. In addition to controlling roughness, other research explores the creation of innovative coatings through functionalization, such as using vitamin E [10]. Currently, there is only a temporary fixation device that features a biodegradable coating designed to release the antibiotic Gentamicin, along with two devices that have coatings that release silver ions. Building on previous findings, thin inorganic coatings of potential relevance for temporary fixation devices are being

developed using zirconia doped with silver (ZrO_2–Ag), characterized by a very smooth surface topography that can help diminish bacterial adhesion [11].

Metal ions have been historically utilized as antimicrobial agents and disinfectants. Specifically, silver (Ag) demonstrates significant effectiveness in eliminating multidrug-resistant bacteria and in enhancing the success rates of treatments for infections caused by resistant strains [12], [13].

There has been a significant rise in the use of silver in the healthcare sector because of its exceptional antibacterial characteristics [14], [15].

Silver is extensively used in antibacterial coatings due to its recognized capability to eliminate bacteria by compromising membrane integrity, inhibiting protein synthesis, disrupting DNA replication, or inducing oxidative stress through the generation of reactive oxygen species (ROS) in the cytoplasm [16].

While the antibacterial potential of silver and surface modifications has been widely recognized, the precise biological responses they trigger at the molecular level remain complex and not fully understood. In the era of high-throughput biology, advanced technologies such as next-generation sequencing (NGS) and mass spectrometry (MS) have revolutionized our ability to investigate these interactions. Two major fields that have grown out of these advances are transcriptomics and proteomics. This thesis explores how cells respond at the molecular level to zirconia doped with silver coatings on temporary fixation devices by applying transcriptomic and proteomic analyses.

To fully understand how surface modifications influence cellular behavior, it is essential to move beyond macroscopic observations and explore molecular-level responses. High-throughput techniques such as transcriptomics and proteomics offer powerful tools to capture these changes in detail. By examining gene and protein expression patterns, we can uncover how cells react to different implant surfaces, providing insights that are critical for designing materials with improved biological performance.

1.1. Transcriptomic data analysis and its challenges

The transcriptome acts as a vital intermediary, rich with information that provides significant insights into the biology of diseases. It includes the complete set of RNA transcripts produced in a cell, reflecting both fundamental genetic and epigenetic factors, as well as environmental influences. This creates a comprehensive view of the functional states of cells at any given moment [17].

Recent advancements in technology now enable us to examine the transcriptome at the individual cell level [17].

Since 1950, the primary focus of RNA studies has been its role in protein synthesis. Subsequent discoveries revealed that only a minor portion of RNA is responsible for coding proteins, while the remainder may serve various functional purposes. With the rise of NGS, particularly RNA sequencing (RNA-seq) technology, the costs associated with sequencing production significantly decreased. Among the uses of NGS, transcriptome analysis examining transcripts within a cell and quantifying them for specific developmental stages or treatment conditions has become increasingly prevalent in laboratories. Consequently, the last decade has brought forth new insights into transcriptome complexity and the roles of RNA molecules in cellular functions. In terms of computational progress, bioinformatics research has introduced new methodologies for analyzing RNA-seq data. Comparing transcriptome profiles from multiple samples often poses challenges for those without programming expertise [18].

Currently, the use of NGS technology for transcriptome analysis is regarded as the gold standard, completely supplanting microarray-based expression profiling, which has dominated the field for over a decade. Over ten years ago, with the rise of massively parallel sequencing, RNA-seq emerged as a crucial technological tool in molecular biology [19], [20], [21] to enhance our understanding of gene expression processes by assessing expressed mRNAs [22] through the sequencing of a complete transcriptome in various cell and tissue types and conditions. The significance of this sequencing technology lies in its ability to explore numerous aspects of molecular biology, such as studying mRNA splicing [23], the regulation of gene expression mediated by noncoding RNAs [24], and the role of enhancer RNAs [25]. More than 100,000 alternative isoforms have been identified through RNA-seq experiments [23], yet pinpointing functional transcripts remains a challenge. In an NGS experiment, a

typical laboratory workflow initiates with RNA extraction and concludes with library preparation for sequencing. The sequencing process generates several million short reads, usually comprising hundreds of bases in length. The main computational application of RNA-seq involves determining the quantitative variations in expression levels between experimental groups (for example, at the gene or transcript levels). Other computational approaches are also explored based on specific biological questions, including alternative splicing events, alternative polyadenylation sites, and fusion events. However, as the volume of experimental data continues to escalate, bioinformatics has effectively entered the big data era, making the interpretation of results from comparisons between different RNA-seq samples increasingly intricate. In conjunction with technological advancements in NGS, extensive primary repositories of raw sequencing data (such as the Sequence Read Archive SRA) [26], The Cancer Genome Atlas TCGA, Genotype-Tissue Expression GTex [27], and the Cancer Cell Line Encyclopedia CCLE) have been rapidly populated with a vast amount of sequencing data. These archives house an impressive collection of RNA-seq samples that can facilitate new analyses [28], [29], [30] and enable the reuse of data for in silico comparisons and validations. Over the past decade, numerous pipeline tools have been developed specifically for RNA-seq data analysis [31].

Advancements in sequencing technologies, particularly NGS and the groundbreaking third-generation sequencing, have transformed the study of genetic sequences in a cost-effective way [32]. These techniques have produced enormous volumes of data related to DNA, RNA, and protein sequences [32]. Specifically, RNA sequence data is being leveraged to reveal hidden insights, such as the unique functions of RNAs in living organisms (for instance, in protein synthesis and gene regulation) and their connections to various diseases, including cancers, genetic illnesses, and neurodegenerative disorders [33]. To obtain a deeper understanding of RNA sequence information, researchers are employing wet-lab experimental methods for a range of RNA sequence analysis tasks [34]. However, RNA sequence analysis based on wet-lab experiments tends to be time-intensive and costly. Due to the inherent drawbacks of traditional wet-lab methods, researchers are exploring the capabilities of Artificial Intelligence (AI) techniques to create applications for AI-driven RNA sequence analysis [34]. The majority of AI-driven RNA sequence analysis applications are categorized under regression, clustering, and classification approaches. The objective of the clustering approach is to group RNA

sequences that exhibit similar traits [32], [34]. The regression approach is centered around predicting continuous numerical values based on RNA-seq data [32], [34]. For example, researchers might employ regression to forecast how the expression level of a particular gene could fluctuate under different environmental conditions [32], [34]. The classification approach focuses on assigning RNA sequences to predefined groups [32], [34]. A comprehensive workflow for these three distinct paradigm-based AI-driven RNA sequence analysis predictive pipelines is depicted in Fig. 1. A closer examination of Fig. 1 demonstrates that the functioning paradigm of AI-driven RNA sequence analysis predictive pipelines can be divided into four separate stages. The first stage concentrates on gathering and curating high-quality benchmark datasets. This stage either utilizes datasets produced by prior studies or generates new datasets. The generation of new datasets involves obtaining RNA sequences and their related information from public databases or obtaining data through wet-lab experiments. The second stage, known as representation learning, utilizes various techniques to capture the informative distribution of nucleotides in RNA sequences and transform this information into statistical vectors. This conversion is crucial because AI techniques fundamentally depend on statistical vectors. The third stage employs the statistical vectors of RNA sequences along with machine learning (ML) or deep learning (DL) algorithms to make predictions. The aim of the fourth stage is to assess the effectiveness of predictive pipelines that deploy representation learning and machine/deep learning methods. Among all four stages, the representation learning stage is the most pivotal, as high-quality statistical vectors enable even basic ML algorithms to perform effectively, whereas inadequate vectors can impede the performance of advanced algorithms [35].

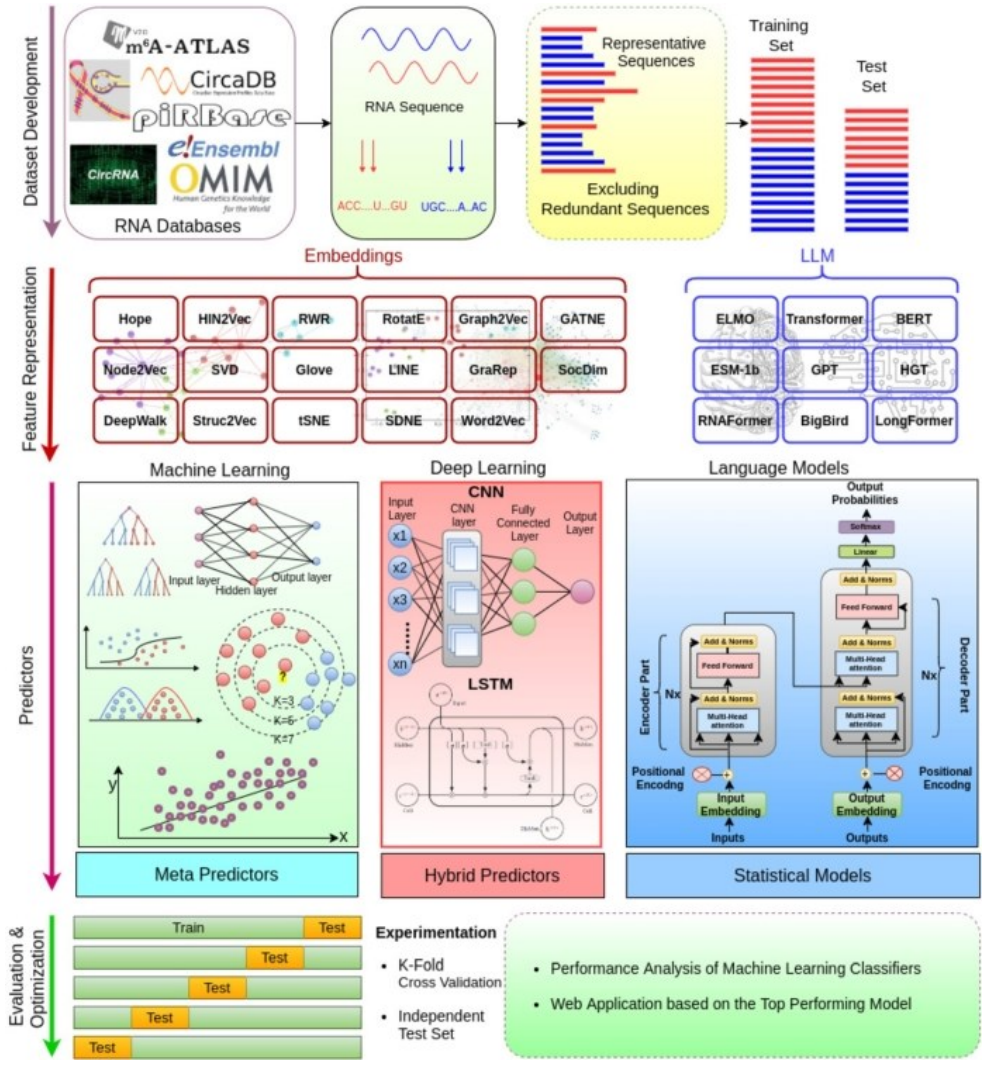


Figure 1 Predictive pipeline of RNA Sequence Analysis Tasks [35].

Challenges in transcriptomic data analysis

Although transcriptomics offers promising possibilities for understanding disease mechanisms, it also presents considerable hurdles, especially regarding the analysis and interpretation of the large amounts of data produced using an expanding variety of analytical tools [36].

RNA-seq has numerous applications, but a universal analysis pipeline is not applicable for all scenarios [37].

Large transcriptomic datasets, especially those featuring various experimental conditions or time-series analyses, require significant computational resources and effective algorithms.

Achieving scalability and preserving performance continue to be ongoing challenges for large-scale transcriptomic initiatives [38].

The normalization of transcriptomic data continues to be a major challenge, particularly in scenarios where the datasets are unbalanced, featuring an uneven number of up-regulated versus down-regulated genes. Many current techniques operate under the assumption of balanced expression, and breaching this assumption can result in significant biases. There is no one-size-fits-all normalization approach that suits every experimental design, and choosing the wrong method can skew biological interpretations [39].

Recognizing this issue has significantly advanced the creation of normalization techniques for shifted expression data. Over the past twenty years, more than 20 methods have been introduced that either avoid or rely less heavily on fundamental assumptions:

Cross normalization (CrossNorm) [40], Informative cross normalization (ICN) [41], Modified least-variant set normalization (LVSmiR) [42], Kernel density weighted loess normalization (KDWL) [43]

While bulk transcriptomic investigations have greatly advanced our understanding of complex diseases, their ability to explore is limited by the diverse compositions of biological samples. Specifically, by averaging the expression levels across various cell types, RNA-Seq analysis fails to account for changes at the cellular level, which complicates the identification of the underlying components within tissues. There are ongoing challenges that need to be overcome in order to accurately assess closely related cell types from RNA sequencing data, highlighting the complementary role of single-cell RNA-Seq for this purpose [44].

1.2. Proteomic data analysis and its challenges

The term "proteome" refers to the full array of proteins present in a cell. It is recognized as the extensive analysis of all proteins within entire cells, tissues, and organisms. This field focuses on the investigation of protein functions, their locations, and interactions among proteins [45]. The primary goal of proteomics is to understand where proteins are located by creating a three-dimensional representation of a cell. The inherent dynamism of proteomics aids in identifying both the external and internal signals that influence the biochemical systems associated with cells [45]. It also provides insight into post-translational modifications, protein synthesis, and degradation. This process can be viewed as capturing a momentary snapshot of the protein environment at a specific time [46].

Proteins are one of the most vital and informative molecules involved in biological functions [47]. They play both structural and functional roles in cellular processes. In biomedical research, it is vital to analyze our proteome to identify diseases and understand their behavior. Traditional techniques often fall short and lack the ability to simultaneously analyze many proteins from the vast array of complex biological data samples [48]. Consequently, MS is one of the modern methods employed to carry out the aforementioned task in a cohesive manner [49]. To understand the connection between a disease and its phenotypic behavior, as well as the progression and regulatory mechanisms of the disease in the body, must be examined [49]. Through mass spectrometry, we can compile a list of proteins that are readily identifiable using the available information [49].

Additionally, various bioinformatics tools are employed to abstract data from the large datasets generated, utilizing deep neural networks (DNN) constructed through DL algorithms.

Typically, DL neural networks possess two key characteristics: first, they consist of multiple layers formed by multi-layered structures of nonlinear processing units, and second, diverse types of supervised and unsupervised algorithms are applied at each layer to extract features from the specific dataset. The applications of DL are extensive and adaptable within the scope of bioinformatics and biomedical data analysis, including genome sequencing, among other areas [50].

Proteomics experiments can be defined by the volume of experimental data produced during the studies and the accompanying informatics models. Small-scale experiments yield relatively limited data, such as identifying a specific group of proteins in a one- or two-dimensional gel

(1D or 2D) [51]. These experiments are specifically designed around well-defined hypotheses, which are rigorously tested through the experimental findings. The analysis requires significant intuitive iteration and re-evaluation by qualified observers, making it hard to automate profitably. Consequently, the significance of the information's structure diminishes: each observer tends to have distinct techniques for deriving insights and knowledge from the dataset. Medium-scale proteomics employs many of the same techniques as small-scale experiments, but the dataset's size becomes so large that thorough evaluation by trained observers is unfeasible [52]. The objective of these experiments remains to extract the most knowledge from the data, but the sheer volume of information presents a genuine obstacle to the iterative, intuitive methodology utilized in small-scale studies. Observers are engaged to sift through the information to assess whether it has been accurately constructed from the data. The informatics model for this type of project is necessarily more rigid, curtailing the reporting flexibility that small-scale experiments typically allow. Data and information storage transform into a form-filling task, where the aim is to fill a relational database that permits extensive examination of the information. Large-scale proteomics initiatives rely on fully automated systems to gather and process data, transforming it into information [53]. Observer involvement is minimal, except for quality control purposes. The informatics model must be entirely fixed, adhering to a set of external guidelines dictated by the experimental design [54].

A growing volume of proteomics data is being stored in databases. Specifically, pathway databases (Table 1) enable researchers to investigate molecular mechanisms and interactions among proteins linked to a specific condition identified in a proteomics study. The biological information structured in pathway databases can be utilized to formulate robust hypothesis through an inductive approach and to assign an expectation value to it [55].

Over-representation analysis can be utilized to emphasize significant pathways in order to functionally interpret variations in protein abundance identified through differential quantitative proteomics. The hypergeometric test, also known as Fisher's test, is typically employed to determine a p-value for each pathway that is either over or under-represented. Fisher's test is a conditional analysis, where the p-value for an over-represented pathway is derived from the count of differentially expressed proteins within the experimental data set and a reference database, regardless of their pathway assignment. A variety of software tools exist to carry out this type of analysis (Table 2) [55].

Table 1
Main pathway databases.

Database	Description
Reactome	It collects the so called "reactions", where molecules (*e.g.*, proteins and metabolites) form a network of biological interactions and are grouped into pathways (https://reactome.org/) [10].
KEGG	It integrates molecular building blocks with the knowledge on molecular wiring diagrams of interaction, reaction and relation networks (https://www.genome.jp/kegg/) [11].
GO Consortium	It is based on the Gene Ontology (GO) classification, which defines concepts used to describe gene function (http://geneontology.org/) [12].
WIKIPATHWAY	It is an open, collaborative platform for capturing and disseminating models of biological pathways for data visualization and analysis (https://www.wikipathways.org/index.php/WikiPathways) [13].
Pathway Commons	It provides a web-based interface that enables biologists to browse and search a collection of pathways from multiple sources (http://www.pathwaycommons.org/pc/) [14].

Table 2
Main pathway analysis software tools.

Tool	Description
DAVID	It provides a set of functional annotation tools for investigators to understand biological meaning, such as identification of enriched GO terms, clustering of redundant annotation terms and visualization of genes using BioCarta & KEGG pathway maps (https://david.ncifcrf.gov/) [15].
Enrichr	It contains a large collection of gene set libraries available for analysis and download. It is a source of curated gene sets and a search engine that accumulates biological knowledge for further biological discoveries (http://amp.pharm.mssm.edu/Enrichr/) [16].
KOBAS	A web server for functional gene set enrichment that can accept either gene lists or gene expression data as input. It generates enriched gene sets with the corresponding name, the *p*-value or a probability of enrichment and an enrichment score based on results of multiple methods (http://kobas.cbi.pku.edu.cn/) [17].
PANTHER	It classifies proteins according to family (evolutionarily related proteins) and subfamily (proteins with the same function), molecular function, biological process and pathway (http://www.pantherdb.org/) [18].
PathVisio	It provides a basic set of features for pathway drawing, analysis and visualization (https://www.pathvisio.org/) [19].
WebGestalt	A functional enrichment analysis web tool. It supports several organisms gene identifiers and three methods for enrichment analysis, including over-representation analysis, Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA) [20].
Analyse data (Reactome)	It is implemented for over-representation, pathway-topology and expression analyses (https://reactome.org/) [10].

[55]

Challenges in proteomic data analysis

The comparison, sharing, and exchange of data or analytical results obtained from various platforms or different laboratories remain significant challenges, primarily due to the lack of standardized data formats, processing parameters, and quality assessment criteria. As a result, the of a unified pipeline for the processing and analyzing complex proteomics datasets has become increasingly essential [56].

In peptide-based MS quantification, the amount of a protein is estimated from a limited selection of peptides, sometimes only a single peptide. To improve the reliability of these estimates, researchers generally apply thresholds for the minimum number of peptides considered. For example, Peaks Studio Quantification typically relies on the top three peptides for protein quantification, although ideally, all reliably detected peptides should be incorporated [57].

Similarly, in top-down proteomics, many data analysis tools either emphasize protein identification, have a complicated user interface, or are exclusive to instrument manufacturers, thereby limiting accessibility for researchers [58].

Another major challenge arises from the intrinsic characteristics of proteomics data, which is often high dimensionality and very limited sample sizes, creating substantial difficulties when applying ML methods to extract the most pertinent information. While many statistical techniques and various analysis pipelines are utilized in analyzing proteomics data, it remains uncertain which methods yield the most effective, reproducible, and clinically relevant outcomes [59].

Machine learning in proteomic data analysis

Just as proteomics serves as a broad term that encompasses numerous workflows and concepts, the term ML also carries a similarly expansive interpretation [60]. ML methodologies have enabled systematic, data-driven exploration of comprehensive proteomic platforms, thereby uncovering novel biomarkers [61]. This advancement has created opportunities to evaluate both disease-specific and shared molecular features across different conditions. In the broader context of biomedical science, ML is driving discoveries in areas where traditional approaches, based primarily on human-derived knowledge, have achieved limited success [62]. However, progress in testing and applying these technologies has been hindered by the lack of prospective studies that link diverse diseases with electronic health records, representing a significant barrier to their wider clinical translation [63].

One notable example of a successful ML application in proteomics is Deep Multiple Reaction Monitoring (DeepMRM), a targeted proteomic data interpretation tool that employs DL algorithms for object detection, specifically created to significantly minimize the manual inspection workload, even for complex and noisy datasets. By recontextualizing the task of identifying peaks for targeted peptides as an object detection problem, DeepMRM locates instances of peak groups within a series of 1D chromatograms, similar to how conventional object detection frameworks recognize object instances of a certain class in a 2D image. DeepMRM takes transition chromatograms of different lengths and quantities as input, producing chromatographic peaks of targeted peptides along with their corresponding abundances and confidence scores. The detection outcomes and transition chromatograms are displayed in a graphical user interface, allowing users to quickly review and modify the results as required. DeepMRM enhances the efficiency of high-throughput targeted proteomics analysis through automated quantification and quality control, eliminating the need for decoy transitions [64].

2.The aim of the study

The objective of this research is to computationally assess pre-existing transcriptomic and proteomic datasets derived from human mesenchymal stem cells grown on bare titanium (Ti control) and titanium surfaces coated with low concentrations of zirconia doped with silver (Ag low). By employing distance matrix–based methods, the study measures alterations in gene and protein expression relationships between different conditions and identifies the gene and protein pairs that are most significantly affected. Following this, additional analyses such as clustering, classification, and functional enrichment are conducted to uncover important molecular characteristics, protein–protein interactions, and biological processes linked to silver-doped coatings.

3.Materials and Methods

3.1 Cell lines and experimental conditions

Human mesenchymal stem cells (hMSCs) derived from three independent cell lines were used in this study. Cells were cultured on titanium alloy (Ti6Al4V) substrates to investigate molecular responses to surface modification by silver doping. Two experimental conditions were compared: bare titanium (Ti used as control) and zirconia-coated titanium doped with a low concentration of silver (Ag low). The Ag low coating was designed to release silver ions in a controlled manner, thereby providing bacteriostatic effects while maintaining substrate stability.

Cells from each of the three lines were seeded on both Ti and Ag low substrates under identical culture conditions. After defined exposure times, proteins and total RNA were extracted separately from each line and prepared for proteomic and transcriptomic profiling. The proteomic profile was analyzed through MS using a Q Exactive Plus Mass Spectrometer. The cell culture experiments and initial omics data generation were previously carried out and published by *Ferraris et al. 2025* [65].

In this thesis, these existing datasets were re-analyzed using a distance-based computational approach aimed at identifying the gene and protein pairs most affected by the silver-doped coating.

3.2 Data preprocessing

The proteomic and transcriptomic datasets were first imported into the Orange tool (version 3.38.1) for preprocessing. To ensure that all features were directly comparable across genes and conditions, interval normalization (0-1) was applied. Data distribution was explored using box plots, which allowed visualization of variation across the three cell lines under each condition. To evaluate whether there were global differences between the two experimental groups, a T-test was performed, comparing mean expression and protein values between the Ag low and Ti conditions.

Outlier detection was then performed to refine the dataset. A one-class support vector machine (SVM) implemented in Orange with a 0.01 Karnel coefficient was used to identify and exclude outliers, with only inlier data points retained for further analyses.

The one-class SVM is a variation of SVM designed to work with unlabeled data [66], and as a well-established method for detecting outliers, it has found extensive use across various applications [67]. The primary distinction between the two classifiers lies in the fact that the one-class SVM focuses solely on a single class. This approach was originally developed for unsupervised detection of outliers and anomalies in datasets containing only one class [68], [69], [70], [71].

The one-class SVM is quite similar to traditional SVMs, which were initially presented by Vapnik [71], [72], but it differs in its application. Specifically, the one-class SVM identifies a decision boundary that wraps around the smallest area possible around a chosen set of examples from the target class. Once this boundary is created, it can be used to determine whether new test examples are outliers (if they fall outside the boundary) or non-outliers (if they are within the boundary) [73].

Genes or proteins that were absent in one of the two conditions were also removed to maintain consistency and avoid misleading comparisons.

To assess reproducibility between biological replicates, Pearson correlation coefficients were calculated for the three cell lines under each condition.

The Pearson correlation coefficient is commonly utilized for data that follows a joint normal distribution, and the correlation quantifies the degree of a monotonic relationship between two variables. In data that exhibits correlation, a change in the scale of one variable corresponds to a change in the scale of another variable, either in the same direction or in the opposite direction. In simpler terms, greater values of one variable are generally linked to either greater (positive correlation) or lesser (negative correlation) values of the other variable, and vice versa [74].

1.3. Distance-based analysis of transcriptomic and proteomic data

Studying gene expression and other types of high-dimensional genomic data can be challenging due to the sheer number of variables involved. Over the years, a variety of methods have been developed to simplify and make sense of this complexity [75]. One common strategy is data reduction, for example, using cluster analysis which helps to identify patterns and group similar samples in a meaningful way [76].

A particularly powerful approach is the construction of a similarity or distance matrix. This matrix provides a straightforward way to quantify how similar or different each pair of samples is based on their gene or protein expression profiles. The idea of using distance matrices dates back to the earliest gene expression studies and has since become a standard practice. Today, it underpins many modern tools for visualizing and analyzing omics data and forms the foundation for most clustering and network-based analyses [77], [78], and it forms the foundation for most clustering techniques [78].

In this study, we applied this framework to both transcriptomic and proteomic datasets, enabling us to identify genes and proteins whose expression relationships were most affected by the silver-doped titanium surfaces.

3.1.1. Transcriptomic analysis

3.1.1.1. Distance matrix calculation and ratio

For the transcriptomic data, pairwise Euclidean distances were calculated between the mean expression values of the three cell lines under each condition (Ag low and Ti), generating distance matrices. To facilitate downstream analysis, each distance matrix was transformed into a table with three columns: GeneA, GeneB, and the corresponding distance. Immediately after this transformation, self-comparisons (e.g., GeneA vs. GeneA) and duplicate pairs (e.g., GeneA–GeneB vs. GeneB–GeneA) were removed to retain unique gene pairs for downstream analyses. To directly compare the two conditions, the distance ratio (Ag low / Ti) was then calculated for each gene pair. This allowed us to quantify how relationships between genes were altered by the silver-doped coating.

3.1.1.2. Selection of top gene pairs

Using these distance ratios, the top and bottom 5% of gene pairs showing the largest differences between Ag low and Ti were selected.

3.1.1.3. Functional enrichment analysis

The selected gene pairs were analyzed for functional enrichment using the ToppGene portal, with Bonferroni correction applied. Genes with $p < 0.01$ were considered significant. Adhesion-related processes were cross-checked against GO terms in Ensembl BioMart (GO:0007155, GO:0022610,

GO:0031589, and GO:0098609) to ensure accurate annotation. Exceptionally small *p*-values reported as zero were replaced with 1×10^{-300} to allow calculation of $-\log_{10}$ (*p*-values).

For visualization, the 20 most significantly enriched biological processes were plotted for both the "Ag low increased distance" and "Ag low shortened distance" groups. Each group was assigned a distinct color to facilitate direct comparison. Relevant processes, including "regulation of programmed cell death" and "response to wounding," were highlighted to illustrate key functional changes induced by the silver-doped titanium surfaces.

3.1.2. Proteomic analysis

3.1.2.1. Distance matrix calculation and transformation

Proteomic data was processed similarly. Pairwise Euclidean distances between proteins were calculated for each condition, generating distance matrices. Each matrix was then transformed into a table with three columns: Protein A, Protein B, and the corresponding distance. After this transformation, self-comparisons (e.g., Protein A vs. Protein A) and duplicate pairs (e.g., Protein A_ Protein B vs. Protein B_ Protein A) were removed to retain only unique protein pairs for downstream analyses.

3.1.2.2. Distance ratio calculation and clustering

For each protein pair, the ratio of distances between Ag low and Ti conditions was calculated, followed by a $\log_2$ transformation. These transformed ratios were clustered using the k-means algorithm in Orange (version 3.38.1) with k = 5, random initialization, 10 re-runs, and a maximum of 300 iterations. Clustering quality was evaluated using the Silhouette coefficient, which measures how well clusters are separated while remaining internally cohesive.

3.1.2.3. Selection of protein pairs

To focus on biologically meaningful differences, only protein pairs with at least a two-fold change in distance ratios were retained for further analysis.

3.1.2.4. Functional enrichment analysis

These selected protein pairs were submitted separately to STRING software for protein-protein interaction analysis, using a minimum interaction score of 0.7 to ensure reliability.

4.Results

1.4. Data overview and quality assessment

To explore the molecular responses of hMSC to zirconia coatings, the transcriptomic and proteomic datasets were first examined through descriptive and comparative analyses.

The initial transcriptomic dataset contained 21,398 genes, evenly distributed between conditions (Ag low = 10,699; Ti = 10,699). The proteomic dataset included 236 proteins, also evenly divided (Ag low = 118; Ti = 118). These datasets were subjected to preliminary quality assessments to ensure reliability for downstream computational analyses.

4.1.1. Distribution of expression data (Boxplots and statistical evaluation)

Boxplots were generated from the raw data to examine the initial distributions of expression values across conditions. The plots were created using the Box Plot widget in Orange (v3.38.1). Each box shows the interquartile range (IQR), with the median indicated by a central line, whiskers extending to 1.5 × IQR, and outliers plotted individually. This visualization allowed the detection of extreme values and provided an overview of the overall data spread.

Transcriptomic dataset: Distributions were broadly comparable (Figure 2), a significant difference in mean values was detected ($t = 34.877$, $p < 0.001$, $N = 21{,}398$), suggesting that transcript levels were more sensitive to the silver coating.

Proteomic dataset: Distributions across Ti and Ag low were highly similar (Figure 3), and statistical analysis confirmed no significant difference in mean values ($t = 0.714$, $p = 0.476$, $N = 236$).

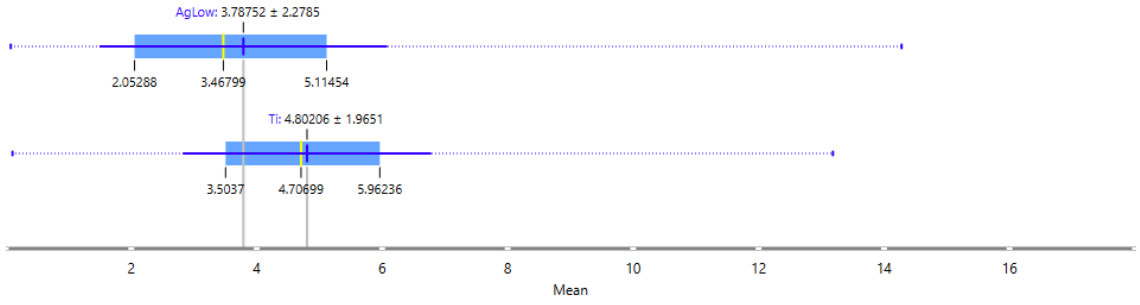


Figure 2 Boxplot showing the distribution of transcriptomic expression values for the two experimental conditions (Ti and Ag low). The plot represents the mean expression values across the three hMSC lines for each condition. The central line indicates

the mean, the shaded box represents the standard deviation, and the whiskers extend to the minimum and maximum values within 1.5 × the standard deviation. Points lying outside this range are displayed as outliers. A significant difference in mean expression was observed between conditions (Ti: 4.80 ± 1.97; Ag low: 3.79 ± 2.28, t = 34.877, p < 0.001, N = 21,398).

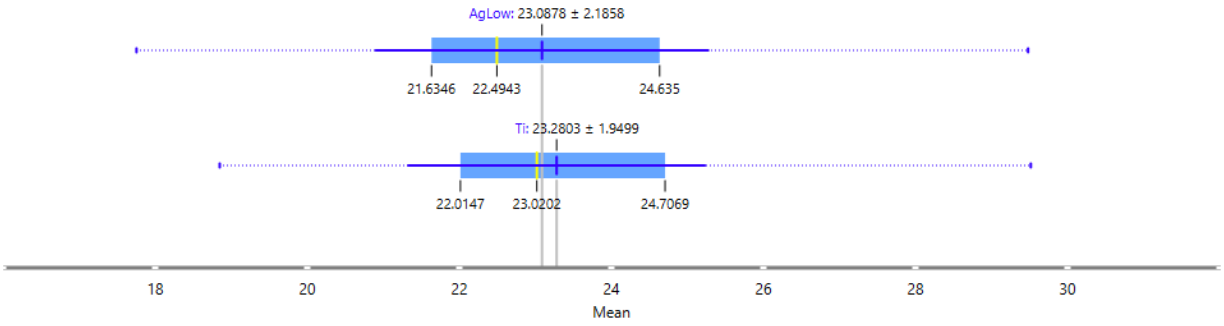


Figure 3 Boxplot showing the distribution of proteomic expression values for the two experimental conditions (Ti and Ag low). The plot represents the mean expression values across the three hMSC lines for each condition. The central line indicates the mean, the shaded box represents the standard deviation, and the whiskers extend to the minimum and maximum values within 1.5 × the standard deviation. Points lying outside this range are displayed as outliers. No significant difference in mean expression was observed between conditions (Ti: 23.28 ± 1.95; Ag low: 23.08 ± 2.19, t = 0.714, p = 0.476, N = 236), indicating comparable overall protein expression levels.

4.1.2. Data normalization and outlier evaluation

Following initial visualization, both transcriptomic and proteomic datasets were normalized using interval scaling (0-1) in Orange (v3.38.1) to place all features on a comparable scale across genes/proteins and conditions. After normalization, outlier detection was performed using a one-class SVM, which flagged genes or proteins with expression values deviating strongly from the main distribution. Only inlier data points were retained for downstream analyses. After this refinement, the proteomic dataset contained 119 proteins (Ag low = 57; Ti = 62), and the transcriptomic dataset contained 10,699 genes (Ag low = 5,051; Ti = 5,648). This combined preprocessing ensured that distance matrix construction, ratio calculations, clustering, and functional enrichment analyses were not biased by extreme or inconsistent values.

4.1.3. Correlation assessment of biological replicates

The reproducibility of the datasets was assessed by calculating Pearson correlation coefficients between the three hMSC lines for each condition.

For the **transcriptomic dataset (Figure 4)**, reproducibility was markedly higher. In the Ag low condition, correlation values were 0.96 (cell line 1 vs. cell line 2), 0.97 (cell line 1 vs. cell line 3),

and 0.97 (cell line 2 vs. cell line 3). Similarly, in the Ti condition, correlations were 0.96 (cell line 1 vs. cell line 2), 0.93 (cell line 1 vs. cell line 3), and 0.93 (cell line 2 vs. cell line 3). These consistently high values demonstrate excellent agreement between transcriptomic replicates in both conditions. The corresponding scatter plots are shown in Figure 4.

For the **proteomic dataset (Figure.5)**, the Ag low condition showed correlations of 0.68 (cell line 1 vs. cell line 2), 0.65 (cell line 1 vs. cell line 3), and 0.83 (cell line 2 vs. cell line 3). In contrast, the Ti condition exhibited lower correlation values of 0.45 (cell line 1 vs. cell line 2), 0.65 (cell line 1 vs. cell line 3), and 0.39 (cell line 2 vs. cell line 3). These results suggest that while reproducibility was moderate in the Ti group, the Ag low group displayed higher consistency among replicates. The corresponding scatter plots are shown in Figure 5.

Taken together, the correlation analysis confirmed that the transcriptomic dataset was highly reproducible, while the proteomic dataset showed moderate reproducibility, particularly under Ti. Nevertheless, proteomic studies frequently report greater variability compared to transcriptomics, and averaging biological replicates is a well-established strategy to minimize noise and improve robustness in downstream analyses [79], [80]. Therefore, averaged values across the three replicates for each condition were used in subsequent analyses, ensuring comparability between transcriptomic and proteomic datasets.

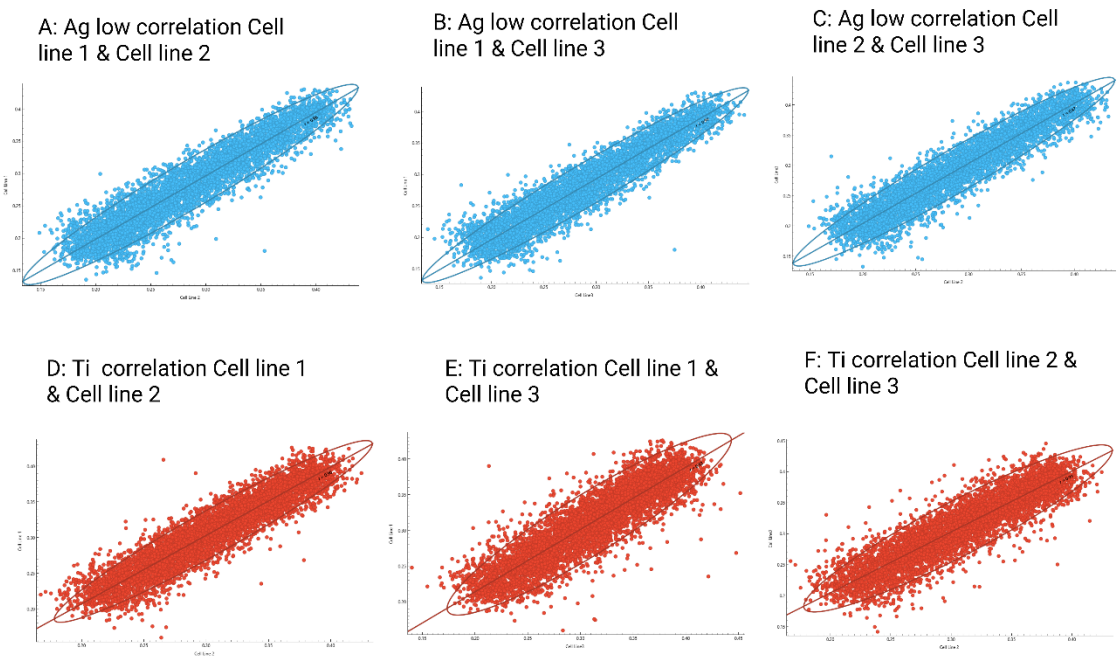


Figure 4 Transcriptomic expression correlations between hMSC cell lines under Ag low (top row: A–C) and Ti (bottom row: D–F) conditions. Data points are tightly clustered along the diagonal in all comparisons, indicating excellent reproducibility. Confidence ellipses and regression lines further illustrate the strong agreement in gene expression profiles, with Pearson correlation values ranging from 0.93 to 0.97.

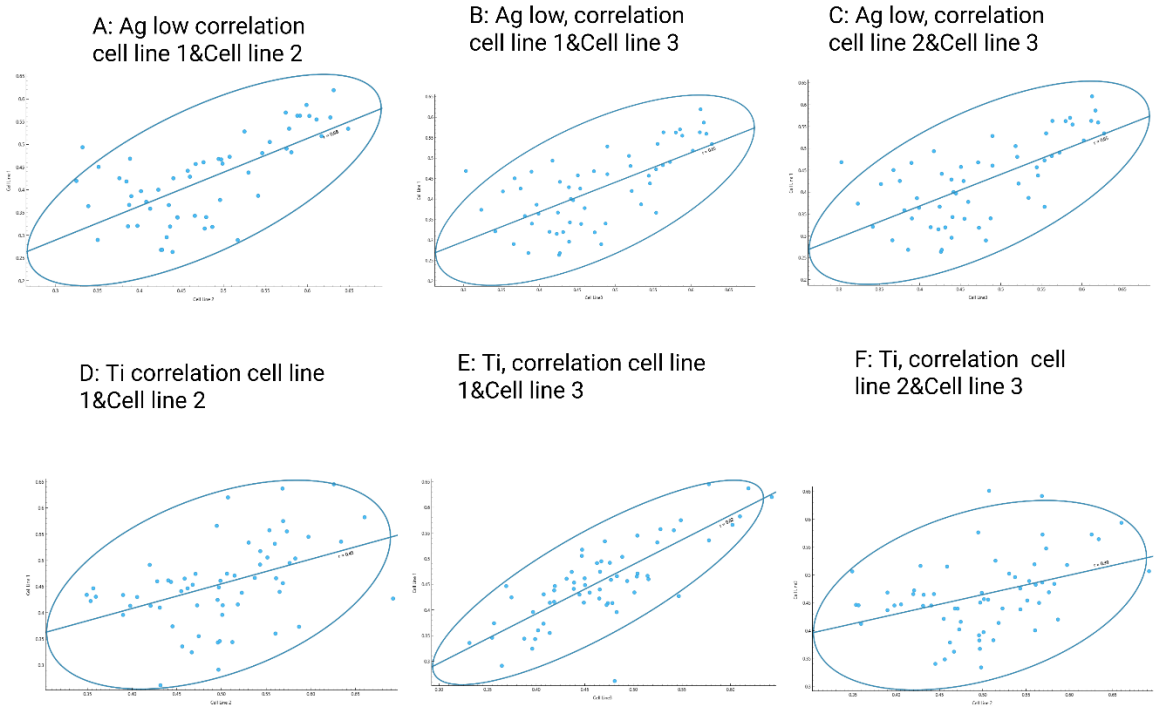


Figure 5 Proteomic expression correlations between hMSC cell lines under Ag low (top row: A–C) and Ti (bottom row: D–F) conditions. Each scatter plot shows how closely protein levels align between two cell lines, with a fitted regression line and confidence ellipse highlighting the degree of correlation. The Ag low condition shows stronger consistency across replicates, while Ti displays more variability.

1.5. Distance-based analysis

Following inlier determination, the datasets were refined to ensure comparability between conditions. Since the number of proteins and genes was not evenly distributed, a one-step removal of unmatched features was applied to both datasets. This filtering resulted in 76 proteins (Ag low: 38; Ti: 38) and 6,564 genes (Ag low: 3,282; Ti: 3,282). For each condition, the average expression values across the three cell lines were then calculated, reducing replicate variability and producing consolidated datasets suitable for subsequent analyses.

Pairwise distance matrices were generated separately for each condition, quantifying the similarity or dissimilarity between every pair of proteins or genes. To avoid redundancy, each matrix was transformed into a three-column long format (feature 1, feature 2, distance), and duplicate entries (e.g., Gene A_Gene B vs. Gene B_Gene A) as well as self-comparisons (Gene A–Gene A) were removed. This filtering resulted in a total of 5,380,840 transcriptomic gene pairs per condition and 703 proteomic protein pairs per condition. To highlight condition-specific relationships, distance ratios (Ag low/Ti) were calculated for each pair and used as the basis for downstream functional enrichment analyses.

4.1.4. Transcriptomic distance matrix analysis

For transcriptomics, the top and bottom 5% of gene pairs from the distance ratio distribution were selected, corresponding to 269,042 gene pairs in each tail (out of 5,380,840 total). These represented the most condition-specific increases or decreases in co-expression distances between genes under Ag low relative to Ti.

The genes represented in these extreme pairs were subjected to functional enrichment analysis using ToppGene, with Bonferroni correction applied and a significance threshold of $p < 0.01$. Adhesion-related terms were further validated in Ensembl BioMart against the following Gene Ontology categories: cell adhesion (GO:0007155), biological adhesion (GO:0022610), cell-substrate adhesion (GO:0031589), and cell-cell adhesion (GO:0098609). To ensure proper visualization of extreme results, enrichment outputs reported as $p = 0$ were substituted with 1×10^{-300} prior to -log10 transformation.

The 20 most significantly enriched biological processes were visualized separately for gene pairs with Ag low increased distances and those with Ag low shortened distances (Figure. 6). Both profiles were highly similar, dominated by terms related to adhesion and structural organization. Processes such as cell adhesion, cell-cell adhesion, cell-substrate adhesion, cell junction organization, and junction assembly appeared in both groups at the top of the rankings. Additional categories, including cell migration, morphogenesis, and response to wounding, reinforced the focus on pathways governing cellular interactions and tissue architecture.

This similarity reflects the fact that many of the same genes were present in both distance tails. Importantly, these overlapping genes likely act as network hubs whose transcriptional relationships shift in multiple directions: becoming more tightly co-expressed with some partners while loosening their co-expression with others.

Thus, the enrichment results point not to two distinct biological programs, but to a reorganization of a common set of adhesion and structure related genes. Ag low appears to perturb this core network bidirectionally, producing both expanded and contracted co-expression distances around the same functional modules.

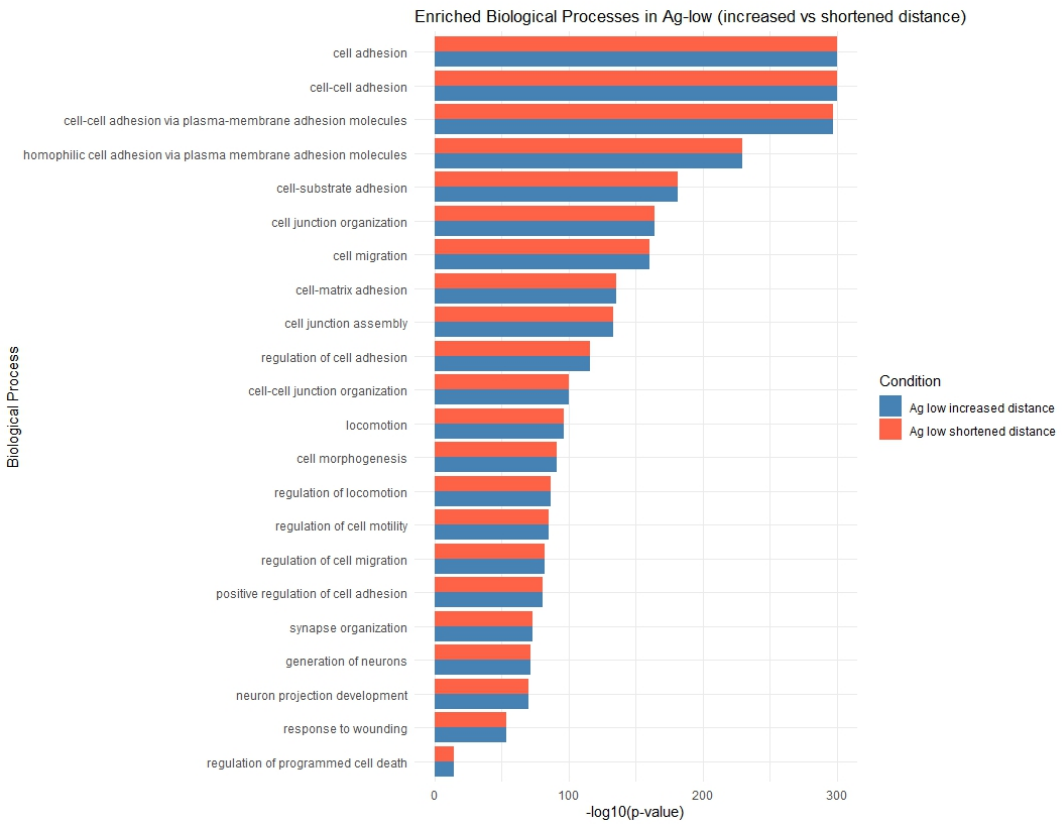


Figure 6 Functional enrichment of transcriptomic gene pairs with altered distances in Ag low relative to Ti.
The bar plot shows the 20 most significantly enriched biological processes (ToppGene, Bonferroni-corrected $p < 0.01$) among the top and bottom 5% of distance ratio gene pairs (Ag low/Ti; 269,042 pairs per group). Blue bars represent pairs with Ag low–increased distances, and red bars represent pairs with Ag low–shortened distances. Both categories showed strong enrichment for adhesion-related processes, including cell adhesion, cell–cell adhesion, and cell-substrate adhesion, suggesting that silver-doped zirconia coatings influence pathways central to hMSC adhesion.

4.1.5. Proteomic distance ratio analysis and functional enrichment

To assess protein level differences between Ag low and Ti conditions, pairwise distances were computed among all proteins, and the ratio of distances (Ag low/Ti) was calculated for each protein pair. A scatter plot of distances under Ag low versus Ti (Figure.7) indicates condition-specific reorganization of proteomic relationships.

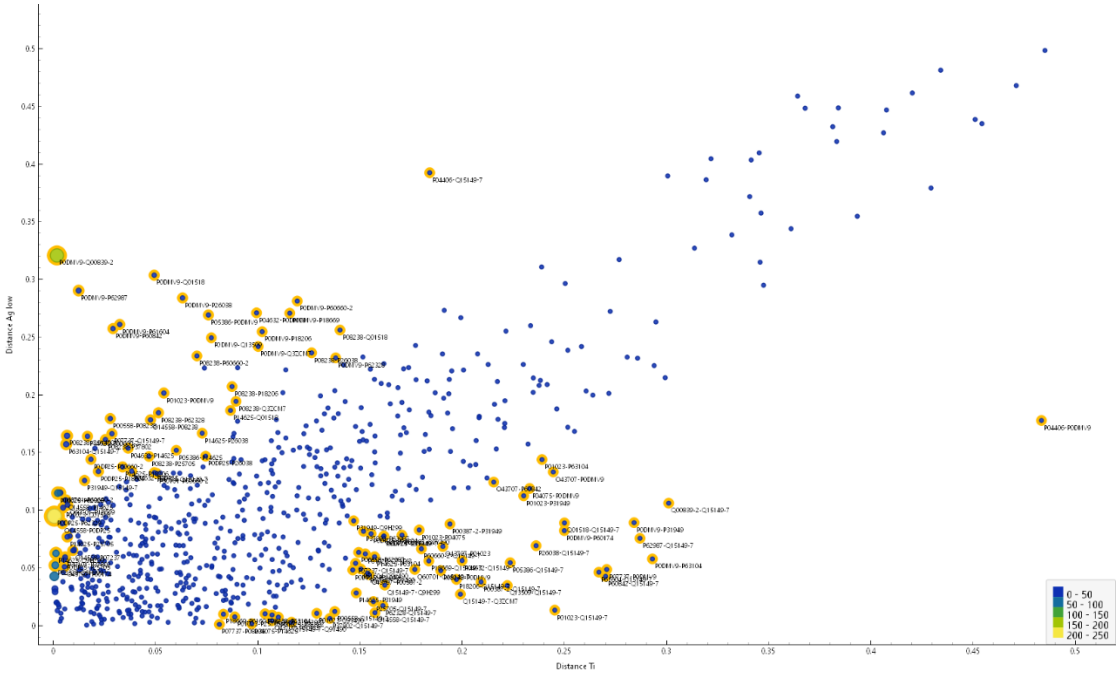


Figure 7 Scatter plot of proteomic distances under Ti (x-axis) and Ag low (y-axis) conditions. Each point represents a protein pair. Pairs deviating from the diagonal highlight condition-specific distance changes.

For systematic evaluation, $\log_2$-transformed distance ratios were subjected to k-means clustering using Orange (v.3.38.1) with k = 5 clusters, random initialization, 10 re-runs, and 300 maximum iterations. The resulting clusters (Figure.8) displayed Silhouette values ranging from 0.49 to 0.71 (average ~0.6), indicating reasonably well-separated groups and confirming that most protein pairs were consistently assigned.

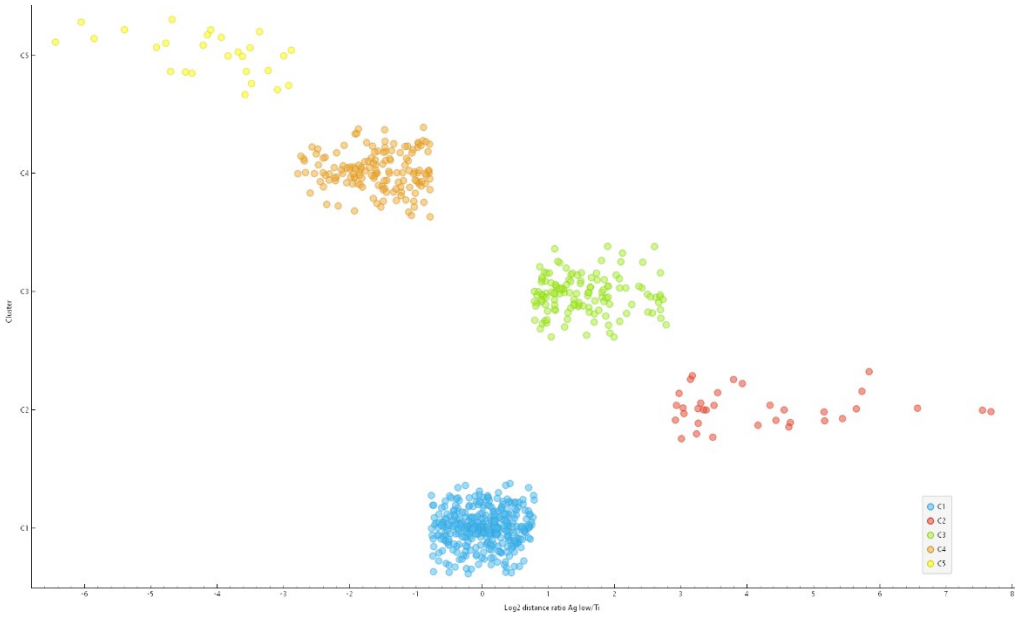


Figure 8 K-means clustering of log₂-transformed proteomic distance ratios (Ag low/Ti). Five clusters (C1–C5) were identified, with an average Silhouette coefficient of ~0.6, indicating moderately well-defined cluster separation.

While clustering captured global patterns, functional enrichment using STRING software was performed separately for increased and decreased distance protein sets to ensure biologically interpretable results. Downstream analyses were restricted to protein pairs with at least a two-fold change in distance ratios. This filtering step yielded 136 pairs with increased distances (19.3%) and 152 pairs with decreased distances (21.6%). Proteins from these two sets were then subjected to functional enrichment analysis in STRING software (minimum confidence score = 0.7).

STRING analysis revealed that both increased- and decreased-distance sets converged on highly overlapping functional categories (Figure. 9A and 9B).

Prominent molecular functions included cell adhesion molecule binding, structural molecule activity, cytoskeletal protein binding, cadherin binding, and actin binding (Figure. 10A and 10B), reflecting the central role of adhesion and cytoskeletal dynamics in the proteomic reorganization associated with Ag low implants. These results paralleled the transcriptomic findings, where adhesion and migration emerged as dominant biological themes.

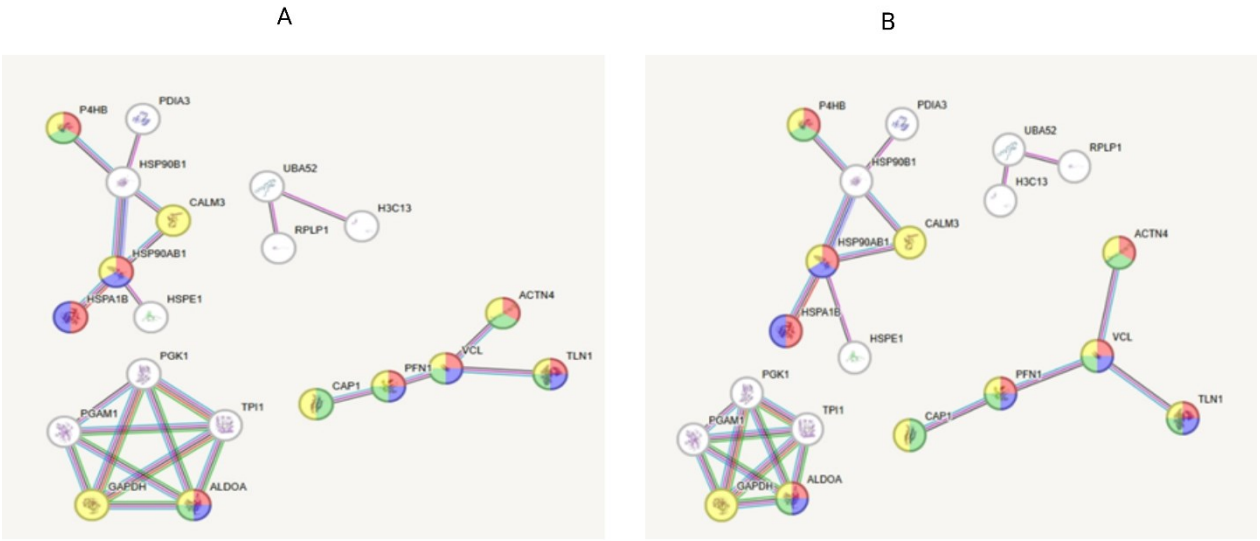


Figure 9 STRING protein–protein interaction (PPI) network of proteins with (A)increased distance lengths (≥ two-fold Ag low/Ti ratio), annotated with enriched molecular functions, (B) decreased distance lengths (≤ 0.5-fold Ag low/Ti ratio), annotated with enriched molecular functions.. Edges represent predicted interactions, and nodes are colored according to functional annotation: red = cell adhesion molecule binding, blue = cadherin binding, green = actin binding, yellow = cytoskeletal protein binding.

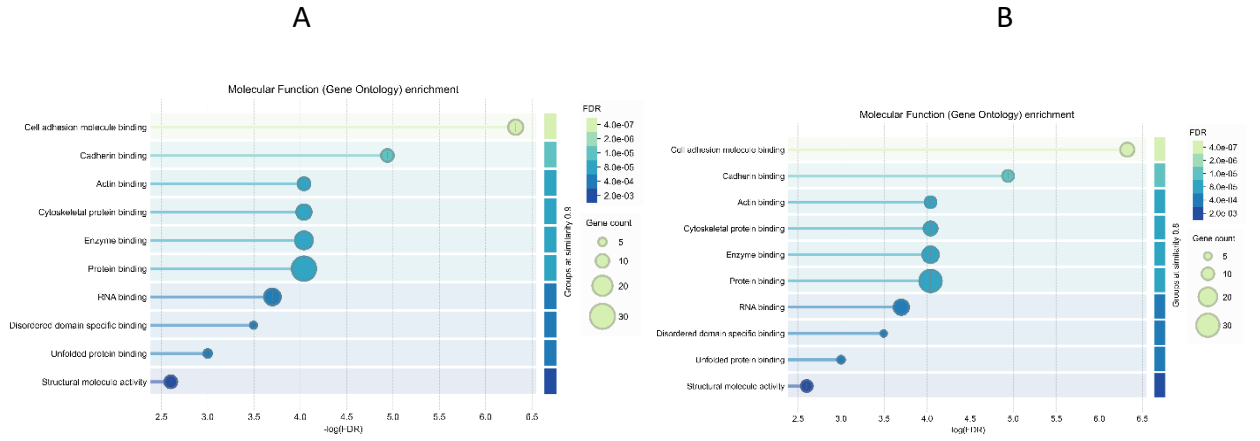


Figure 10 Bubble plots generated in STRING showing molecular functions enriched among protein pairs with increased (10A) and decreased (10B) proteomic distances under Ag low conditions. Each bubble represents a functional category, with size indicating gene count and color intensity reflecting the false discovery rate (FDR). Both sets converged on highly overlapping biological themes, with cell adhesion molecule binding emerging as the top enriched function, followed by cadherin binding, cytoskeletal protein binding, and actin binding. These results underscore the central role of adhesion and cytoskeletal dynamics in the cellular response to silver-doped zirconia coatings

In summary, by using a distance matrix-based strategy, we were able to look beyond single expression changes and instead ask how genes and proteins reorganize their relationships under

different material conditions. This approach highlighted that the silver-doped surface did not simply switch pathways on or off but rather reshaped the existing interaction network in two directions at once: some gene and protein pairs became more tightly connected, while others drifted apart. Strikingly, the same biological themes, cell adhesion, cytoskeletal regulation, and structural organization, kept resurfacing in both transcriptomic and proteomic analyses, regardless of whether distances increased or decreased. This consistency suggests that the central response of hMSCs to Ag low is not the activation of entirely new programs, but the reorganization of their structural and adhesive machinery. In this way, the distance-based method provided a lens to see the subtle rewiring of cellular networks, offering a more nuanced view of how cells sense and adapt to zirconia doped with silver coatings.

5.Discussion

This study looked at how hMSCs respond to silver-doped zirconia coatings on titanium surfaces. Our goal was to understand how these coatings influence the organization of genes and proteins inside the cells. Unlike conventional analyses that focus only on changes in overall expression levels, we used a distance-based computational approach to see how relationships between genes and proteins reorganize under different surface conditions. Interestingly, while the enriched pathways remained largely similar between silver-coated and control surfaces, the internal molecular networks were significantly reshaped. Adhesion and cytoskeletal pathways consistently emerged as the most affected processes, suggesting that the cells' main response to silver is not turning on new programs but reorganizing existing structural machinery.

At the transcriptomic level, we analyzed more than 5 million gene pairs. By focusing on the top and bottom 5% of distance changes, we uncovered large-scale shifts in how genes interact. Functional analysis showed that both increasing and decreasing distance groups involved similar processes such as cell adhesion, cell–cell adhesion, and cytoskeletal organization. This suggests that silver affects the same set of adhesion-related genes in two directions: some gene pairs become more tightly coordinated, while others drift apart. Many genes appeared in both groups, acting as hubs that strengthen connections with some partners while weakening others. This bidirectional network remodeling gives a more nuanced picture of how cells adjust to surface changes, fine-tuning their structural programs rather than simply turning them up or down.

Proteomic data, although smaller, supported these findings. Protein pairs with at least a twofold change in distance were enriched for cytoskeletal and adhesion-related functions, such as cadherin and actin binding. STRING analysis confirmed that these proteins formed overlapping functional modules, echoing the transcriptomic results. Reproducibility was higher in transcriptomics than proteomics, consistent with the higher variability often seen in proteomic datasets. Yet, the shared themes across both layers strengthen confidence that silver coatings reorganize cellular adhesion machinery rather than simply switching it on or off.

Our findings align with previous work by Ferraris et al. [65] who showed that zirconia doped with silver coatings reduce adhesion in hMSCs without being toxic. They reported down-regulation of cadherin- and integrin-related genes and decreased protein levels of Talin (TLN)

and Ras homolog family member A (RhoA). Our study extends this by showing that silver does not simply decrease expression; it reorganizes adhesion and cytoskeletal networks, with genes sometimes playing dual roles depending on their interaction partners. This highlights how distance-based network analysis can reveal subtleties that traditional fold-change methods may miss.

Silver is well known for its antimicrobial properties; it can damage bacterial membranes, block protein synthesis, disrupt DNA replication, and generate oxidative stress [16]. However, its effects on mammalian cells are less studied. Our results suggest that silver also reshapes how cells interact with surfaces, particularly by reorganizing adhesion and cytoskeletal pathways. This dual effect may be particularly useful for temporary implants: it can prevent bacterial colonization while controlling cell adhesion to avoid complications during device removal.

From a methodological standpoint, distance matrices offered new insights. Instead of looking at gene or protein changes individually, we measured shifts in their pairwise relationships, capturing network-level reorganizations. Similar approaches have been applied elsewhere [81] used distance matrices to reveal relationships in microbiome data, and [82]. applied them to adjust batch effects in RNA-seq. While those studies focused on different systems, they show that distance-based analyses can uncover hidden structure that standard methods might miss. Our work adapts this idea to biomaterials, showing how gene and protein networks reorganize in response to surface modifications.

Biologically, adhesion and cytoskeletal remodeling appear central to how hMSCs sense and respond to silver-doped surfaces. Genes that appear in both increased- and decreased-distance groups likely act as network hubs, adjusting their roles depending on context. This dynamic rewiring allows cells to fine-tune adhesion, balancing attachment and flexibility crucial for temporary implants, where too little adhesion could destabilize the device, and too much could make removal difficult. Silver coatings seem to guide cells toward this adaptive balance.

There are some limitations to note. The transcriptomic dataset is very large, and analyzing millions of gene pairs increases the risk of spurious associations. Enrichment analyses did not weight the magnitude of distance changes, so large shifts may be treated the same as minor ones. The overlap of genes between increased- and decreased-distance groups makes interpretation more complex, and the proteomic dataset, though supportive, was smaller and more variable. Future studies could

improve methods to account for direction and magnitude simultaneously and expand proteomic coverage.

Despite these limitations, the convergence of transcriptomic and proteomic findings, along with previous experimental results, suggests that our conclusions are robust. Adhesion and cytoskeletal pathways consistently emerge as key responses of hMSCs to silver-coated surfaces. Future work could include experimental validation through adhesion and migration assays, as well as adding single-cell transcriptomics or other omics layers like metabolomics to explore cellular heterogeneity.

In summary, silver-doped zirconia coatings reshape molecular networks in hMSCs, particularly within adhesion and cytoskeletal pathways. Distance-matrix analysis allowed us to detect subtle but meaningful reorganizations beyond simple expression changes. These findings confirm and extend previous studies, showing that silver modulates cell–material interactions in a dynamic and adaptive way. This understanding could guide the design of temporary implants that balance antimicrobial activity with controlled cellular responses.

6.Conclusion

This thesis investigated how hMSCs respond to zirconia doped with silver coatings on titanium, with a focus on their application in temporary fixation devices. A distance-based computational analysis of transcriptomic and proteomic data showed that silver reorganizes interactions within molecular networks. Adhesion and cytoskeletal processes were consistently the most affected pathways, indicating that silver directly impacts how cells attach and respond to modified surfaces. The study makes two key contributions: first, it confirms that silver acts as both a well-known antimicrobial agent and a modulator of mammalian cell behavior by reshaping adhesion-related pathways. Second, it highlights the value of distance-matrix analysis as a helpful method alongside traditional fold-change techniques, revealing subtle yet important changes in cellular networks. Although the proteomic dataset was relatively small and enrichment tools did not fully account for the significance of distance changes, the consistent results across both omics layers, along with alignment with previous research, support the reliability of these findings. Moving forward, future research should work to increase proteomic coverage, improve computational methods to include the size and direction of distance changes, and validate predictions through functional tests like adhesion and migration assays. Applying this framework to single-cell datasets or adding more omics layers would deepen our understanding of how biomaterials affect cellular responses. In conclusion, this thesis confidently shows that silver-doped zirconia coatings reorganize stem cell molecular networks in ways that support their use in temporary fixation devices. By combining biological insights with a new computational method, it lays a strong foundation for future studies focused on designing implant surfaces that balance antimicrobial protection with positive host responses.

7.Bibliography

[1] A. Lodder, M. Kamath, A. R. Upton, and D. Armstrong, "Evaluation of the Efficacy and Performance of Medical Implants: A Review," *J Long Term Eff Med Implants*, vol. 20, no. 3, pp. 173–185, 2010, doi: 10.1615/JLongTermEffMedImplants.v20.i3.20.

[2] M. Saini, "Implant biomaterials: A comprehensive review," *WJCC*, vol. 3, no. 1, p. 52, 2015, doi: 10.12998/wjcc.v3.i1.52.

[3] B. Kasemo, "Biological surface science," *Surface Science*, vol. 500, no. 1–3, pp. 656–677, Mar. 2002, doi: 10.1016/S0039-6028(01)01809-X.

[4] J. Raphel, M. Holodniy, S. B. Goodman, and S. C. Heilshorn, "Multifunctional coatings to simultaneously promote osseointegration and prevent infection of orthopaedic implants," *Biomaterials*, vol. 84, pp. 301–314, Apr. 2016, doi: 10.1016/j.biomaterials.2016.01.016.

[5] S. Spriano, S. Yamaguchi, F. Baino, and S. Ferraris, "A critical review of multifunctional titanium surfaces: New frontiers for improving osseointegration and host response, avoiding bacteria contamination," *Acta Biomaterialia*, vol. 79, pp. 1–22, Oct. 2018, doi: 10.1016/j.actbio.2018.08.013.

[6] D. Bose and D. Piper, "Temporary external fixation in the management of orthopaedic trauma," *Orthopaedics and Trauma*, vol. 35, no. 2, pp. 80–83, Apr. 2021, doi: 10.1016/j.mporth.2021.01.003.

[7] A. Fatehi Hassanabad *et al.*, "Prevention of Post-Operative Adhesions: A Comprehensive Review of Present and Emerging Strategies," *Biomolecules*, vol. 11, no. 7, p. 1027, July 2021, doi: 10.3390/biom11071027.

[8] AO Research Institute Davos, AO Foundation, Clavadelerstrasse, 8, CH-7270, Davos Platz, Switzerland *et al.*, "Surface polishing positively influences ease of plate and screw removal," *eCM*, vol. 19, pp. 117–126, Feb. 2010, doi: 10.22203/eCM.v019a12.

[9] J. S. Hayes and R. G. Richards, "Surfaces to control tissue adhesion for osteosynthesis with metal implants: *in vitro* and *in vivo* studies to bring solutions to the patient," *Expert Review of Medical Devices*, vol. 7, no. 1, pp. 131–142, Jan. 2010, doi: 10.1586/erd.09.55.

[10] F. Gamna *et al.*, "The use of vitamin E as an anti-adhesive coating for cells and bacteria for temporary bone implants," *Surface and Coatings Technology*, vol. 444, p. 128694, Aug. 2022, doi: 10.1016/j.surfcoat.2022.128694.

[11] W. H. Goldmann, "Biosensitive and antibacterial coatings on metallic material for medical applications," *Cell Biology International*, vol. 45, no. 8, pp. 1624–1632, Aug. 2021, doi: 10.1002/cbin.11604.

[12] J. A. Lemire, J. J. Harrison, and R. J. Turner, "Antimicrobial activity of metals: mechanisms, molecular targets and applications," *Nat Rev Microbiol*, vol. 11, no. 6, pp. 371–384, June 2013, doi: 10.1038/nrmicro3028.

[13] H. Li, R. Wang, and H. Sun, "Systems Approaches for Unveiling the Mechanism of Action of Bismuth Drugs: New Medicinal Applications beyond *Helicobacter Pylori* Infection," *Acc. Chem. Res.*, vol. 52, no. 1, pp. 216–227, Jan. 2019, doi: 10.1021/acs.accounts.8b00439.

[14] S. Eckhardt, P. S. Brunetto, J. Gagnon, M. Priebe, B. Giese, and K. M. Fromm, "Nanobio Silver: Its Interactions with Peptides and Bacteria, and Its Uses in Medicine," *Chem. Rev.*, vol. 113, no. 7, pp. 4708–4754, July 2013, doi: 10.1021/cr300288v.

[15] J.-Y. Maillard and P. Hartemann, "Silver as an antimicrobial: facts and gaps in knowledge," *Critical Reviews in Microbiology*, vol. 39, no. 4, pp. 373–383, Nov. 2013, doi: 10.3109/1040841X.2012.713323.

[16] T. C. Dakal, A. Kumar, R. S. Majumdar, and V. Yadav, “Mechanistic Basis of Antimicrobial Actions of Silver Nanoparticles,” *Front. Microbiol.*, vol. 7, Nov. 2016, doi: 10.3389/fmicb.2016.01831.
[17] E. De Jong and A. Bosco, “Unlocking immune-mediated disease mechanisms with transcriptomics,” *Biochemical Society Transactions*, vol. 49, no. 2, pp. 705–714, Apr. 2021, doi: 10.1042/BST20200652.
[18] M. D’Antonio, P. Libro, E. Picardi, G. Pesole, and T. Castrignanò, “RAP: A Web Tool for RNA-Seq Data Analysis,” in *RNA Bioinformatics*, vol. 2284, E. Picardi, Ed., in Methods in Molecular Biology, vol. 2284. , New York, NY: Springer US, 2021, pp. 393–415. doi: 10.1007/978-1-0716-1307-8_21.
[19] S. J. Emrich, W. B. Barbazuk, L. Li, and P. S. Schnable, “Gene discovery and annotation using LCM-454 transcriptome sequencing,” *Genome Res.*, vol. 17, no. 1, pp. 69–73, Jan. 2007, doi: 10.1101/gr.5145806.
[20] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, “Mapping and quantifying mammalian transcriptomes by RNA-Seq,” *Nat Methods*, vol. 5, no. 7, pp. 621–628, July 2008, doi: 10.1038/nmeth.1226.
[21] U. Nagalakshmi *et al.*, “The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing,” *Science*, vol. 320, no. 5881, pp. 1344–1349, June 2008, doi: 10.1126/science.1158441.
[22] Z. Wang, M. Gerstein, and M. Snyder, “RNA-Seq: a revolutionary tool for transcriptomics,” *Nat Rev Genet*, vol. 10, no. 1, pp. 57–63, Jan. 2009, doi: 10.1038/nrg2484.
[23] E. T. Wang *et al.*, “Alternative isoform regulation in human tissue transcriptomes,” *Nature*, vol. 456, no. 7221, pp. 470–476, Nov. 2008, doi: 10.1038/nature07509.
[24] K. V. Morris and J. S. Mattick, “The rise of regulatory RNA,” *Nat Rev Genet*, vol. 15, no. 6, pp. 423–437, June 2014, doi: 10.1038/nrg3722.
[25] W. Li, D. Notani, and M. G. Rosenfeld, “Enhancers as non-coding RNA transcription units: recent insights and future perspectives,” *Nat Rev Genet*, vol. 17, no. 4, pp. 207–223, Apr. 2016, doi: 10.1038/nrg.2016.4.
[26] R. Leinonen, H. Sugawara, M. Shumway, and on behalf of the International Nucleotide Sequence Database Collaboration, “The Sequence Read Archive,” *Nucleic Acids Research*, vol. 39, no. Database, pp. D19–D21, Jan. 2011, doi: 10.1093/nar/gkq1019.
[27] J. Lonsdale *et al.*, “The Genotype-Tissue Expression (GTEx) project,” *Nat Genet*, vol. 45, no. 6, pp. 580–585, June 2013, doi: 10.1038/ng.2653.
[28] E. Picardi, C. Manzari, F. Mastropasqua, I. Aiello, A. M. D’Erchia, and G. Pesole, “Profiling RNA editing in human tissues: towards the inosinome Atlas,” *Sci Rep*, vol. 5, no. 1, p. 14941, Oct. 2015, doi: 10.1038/srep14941.
[29] M. Cirilli *et al.*, “PeachVar-DB: A Curated Collection of Genetic Variations for the Interactive Analysis of Peach Genome Data,” *Plant and Cell Physiology*, vol. 59, no. 1, pp. e2–e2, Jan. 2018, doi: 10.1093/pcp/pcx183.
[30] K. Licht *et al.*, “A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing,” *Genome Res.*, vol. 29, no. 9, pp. 1453–1463, Sept. 2019, doi: 10.1101/gr.242636.118.
[31] S. Gioiosa *et al.*, “Massive NGS data analysis reveals hundreds of potential novel gene fusions in human cell lines,” *GigaScience*, vol. 7, no. 10, p. giy062, Oct. 2018, doi: 10.1093/gigascience/giy062.
[32] J. Shendure and H. Ji, “Next-generation DNA sequencing,” *Nat Biotechnol*, vol. 26, no. 10, pp. 1135–1145, Oct. 2008, doi: 10.1038/nbt1486.

[33] K. R. Kukurba and S. B. Montgomery, “RNA Sequencing and Analysis,” *Cold Spring Harb Protoc*, vol. 2015, no. 11, p. pdb.top084970, Nov. 2015, doi: 10.1101/pdb.top084970.
[34] D. Pandey and P. Onkara Perumal, “A scoping review on deep learning for next-generation RNA-Seq. data analysis,” *Funct Integr Genomics*, vol. 23, no. 2, p. 134, June 2023, doi: 10.1007/s10142-023-01064-6.
[35] M. N. Asim, M. A. Ibrahim, T. Asif, and A. Dengel, “RNA sequence analysis landscape: A comprehensive review of task types, databases, datasets, word embedding methods, and language models,” *Heliyon*, vol. 11, no. 2, p. e41488, Jan. 2025, doi: 10.1016/j.heliyon.2024.e41488.
[36] E. De Jong and A. Bosco, “Unlocking immune-mediated disease mechanisms with transcriptomics,” *Biochemical Society Transactions*, vol. 49, no. 2, pp. 705–714, Apr. 2021, doi: 10.1042/BST20200652.
[37] A. Conesa *et al.*, “A survey of best practices for RNA-seq data analysis,” *Genome Biol*, vol. 17, no. 1, p. 13, Dec. 2016, doi: 10.1186/s13059-016-0881-8.
[38] W. V. Li and J. J. Li, “Modeling and analysis of RNA-seq data: a review from a statistical perspective,” *Quant. Biol.*, vol. 6, no. 3, pp. 195–209, Sept. 2018, doi: 10.1007/s40484-018-0144-7.
[39] X. Liu *et al.*, “Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review,” *Front. Bioeng. Biotechnol.*, vol. 7, p. 358, Nov. 2019, doi: 10.3389/fbioe.2019.00358.
[40] L. Cheng, L.-Y. Lo, N. L. S. Tang, D. Wang, and K.-S. Leung, “CrossNorm: a novel normalization strategy for microarray data in cancers,” *Sci Rep*, vol. 6, no. 1, p. 18898, Jan. 2016, doi: 10.1038/srep18898.
[41] L. Cheng *et al.*, “ICN: a normalization method for gene expression data considering the over-expression of informative genes,” *Mol. BioSyst.*, vol. 12, no. 10, pp. 3057–3066, 2016, doi: 10.1039/C6MB00386A.
[42] C. Suo, A. Salim, K.-S. Chia, Y. Pawitan, and S. Calza, “Modified least-variant set normalization for miRNA microarray,” *RNA*, vol. 16, no. 12, pp. 2293–2303, Dec. 2010, doi: 10.1261/rna.2345710.
[43] W.-P. Hsieh, T.-M. Chu, Y.-M. Lin, and R. D. Wolfinger, “Kernel density weighted loess normalization improves the performance of detection within asymmetrical data,” *BMC Bioinformatics*, vol. 12, no. 1, p. 222, Dec. 2011, doi: 10.1186/1471-2105-12-222.
[44] B. Chassagnol, G. Nuel, and E. Becht, “An updated State-of-the-Art Overview of transcriptomic Deconvolution Methods,” 2023, *arXiv*. doi: 10.48550/ARXIV.2310.14722.
[45] B. Aslam, M. Basit, M. A. Nisar, M. Khurshid, and M. H. Rasool, “Proteomics: Technologies and Their Applications,” *J Chromatogr Sci*, vol. 55, no. 2, pp. 182–196, Feb. 2017, doi: 10.1093/chromsci/bmw167.
[46] V. Dhingra, M. Gupta, T. Andacht, and Z. F. Fu, “New frontiers in proteomics research: A perspective,” *International Journal of Pharmaceutics*, vol. 299, no. 1–2, pp. 1–18, Aug. 2005, doi: 10.1016/j.ijpharm.2005.04.010.
[47] S.-E. Ong and M. Mann, “A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC),” *Nat Protoc*, vol. 1, no. 6, pp. 2650–2660, Dec. 2006, doi: 10.1038/nprot.2006.427.
[48] M. Padula, I. Berry, M. O′Rourke, B. Raymond, J. Santos, and S. P. Djordjevic, “A Comprehensive Guide for Performing Sample Preparation and Top-Down Protein Analysis,” *Proteomes*, vol. 5, no. 2, p. 11, Apr. 2017, doi: 10.3390/proteomes5020011.

[49] S. Lin, T. A. Shaler, and C. H. Becker, “Quantification of Intermediate-Abundance Proteins in Serum by Multiple Reaction Monitoring Mass Spectrometry in a Single-Quadrupole Ion Trap,” *Anal. Chem.*, vol. 78, no. 16, pp. 5762–5767, Aug. 2006, doi: 10.1021/ac060613f.
[50] N. H. Tran *et al.*, “Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry,” *Nat Methods*, vol. 16, no. 1, pp. 63–66, Jan. 2019, doi: 10.1038/s41592-018-0260-3.
[51] J. S. Andersen *et al.*, “Directed Proteomic Analysis of the Human Nucleolus,” *Current Biology*, vol. 12, no. 1, pp. 1–11, Jan. 2002, doi: 10.1016/S0960-9822(01)00650-9.
[52] A.-C. Gavin *et al.*, “Functional organization of the yeast proteome by systematic analysis of protein complexes,” *Nature*, vol. 415, no. 6868, pp. 141–147, Jan. 2002, doi: 10.1038/415141a.
[53] Y. Ho *et al.*, “Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry,” *Nature*, vol. 415, no. 6868, pp. 180–183, Jan. 2002, doi: 10.1038/415180a.
[54] G. D. Bader and C. W. V. Hogue, “BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways,” *Bioinformatics*, vol. 16, no. 5, pp. 465–477, May 2000, doi: 10.1093/bioinformatics/16.5.465.
[55] C. Monti, M. Zilocchi, I. Colugnat, and T. Alberio, “Proteomics turns functional,” *Journal of Proteomics*, vol. 198, pp. 36–44, Apr. 2019, doi: 10.1016/j.jprot.2018.12.012.
[56] B. Domon and R. Aebersold, “Challenges and Opportunities in Proteomics Data Analysis,” *Molecular & Cellular Proteomics*, vol. 5, no. 10, pp. 1921–1926, Oct. 2006, doi: 10.1074/mcp.R600012-MCP200.
[57] C. Chen, J. Hou, J. J. Tanner, and J. Cheng, “Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis,” *IJMS*, vol. 21, no. 8, p. 2873, Apr. 2020, doi: 10.3390/ijms21082873.
[58] J. A. Melby *et al.*, “Novel Strategies to Address the Challenges in Top-Down Proteomics,” *J. Am. Soc. Mass Spectrom.*, vol. 32, no. 6, pp. 1278–1294, June 2021, doi: 10.1021/jasms.1c00099.
[59] K. H. Lee, S. Assassi, C. Mohan, and C. Pedroza, “Addressing statistical challenges in the analysis of proteomics data with extremely small sample size: a simulation study,” *BMC Genomics*, vol. 25, no. 1, p. 1086, Nov. 2024, doi: 10.1186/s12864-024-11018-2.
[60] H. Desaire, E. P. Go, and D. Hua, “Advances, obstacles, and opportunities for machine learning in proteomics,” *Cell Reports Physical Science*, vol. 3, no. 10, p. 101069, Oct. 2022, doi: 10.1016/j.xcrp.2022.101069.
[61] S. A. Williams *et al.*, “Plasma protein patterns as comprehensive indicators of health,” *Nat Med*, vol. 25, no. 12, pp. 1851–1857, Dec. 2019, doi: 10.1038/s41591-019-0665-2.
[62] J. Fang *et al.*, “Endophenotype-based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer’s disease,” *Nat Aging*, vol. 1, no. 12, pp. 1175–1188, Dec. 2021, doi: 10.1038/s43587-021-00138-z.
[63] J. Carrasco-Zanini *et al.*, “Proteomic prediction of diverse incident diseases: a machine learning-guided biomarker discovery study using data from a prospective cohort study,” *The Lancet Digital Health*, vol. 6, no. 7, pp. e470–e479, July 2024, doi: 10.1016/S2589-7500(24)00087-6.
[64] J. Park *et al.*, “Targeted proteomics data interpretation with DeepMRM,” *Cell Reports Methods*, vol. 3, no. 7, p. 100521, July 2023, doi: 10.1016/j.crmeth.2023.100521.

[65] S. Ferraris *et al.*, “Bacteriostatic, silver-doped, zirconia-based thin coatings for temporary fixation devices tuning stem cells’ expression of adhesion-relevant genes and proteins,” *Biomaterials Advances*, vol. 176, p. 214360, Nov. 2025, doi: 10.1016/j.bioadv.2025.214360.
[66] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution,” *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, July 2001, doi: 10.1162/089976601750264965.
[67] Z. Que and C.-J. Lin, “One-Class SVM Probabilistic Outputs,” *IEEE Trans. Neural Netw. Learning Syst.*, vol. 36, no. 4, pp. 6244–6256, Apr. 2025, doi: 10.1109/TNNLS.2024.3395148.
[68] Kun-Lun Li, Hou-Kuan Huang, Sheng-Feng Tian, and Wei Xu, “Improving one-class SVM for anomaly detection,” in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, Xi’an, China: IEEE, 2003, pp. 3077–3081. doi: 10.1109/ICMLC.2003.1260106.
[69] Yanxin Wang, Johnny Wong, and A. Miner, “Anomaly intrusion detection using one class SVM,” in *Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004.*, West Point, NY, USA: IEEE, 2004, pp. 358–364. doi: 10.1109/IAW.2004.1437839.
[70] M. Amer, M. Goldstein, and S. Abdennadher, “Enhancing one-class support vector machines for unsupervised anomaly detection,” in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, Chicago Illinois: ACM, Aug. 2013, pp. 8–15. doi: 10.1145/2500853.2500857.
[71] A. Thielmann, C. Weisser, A. Krenz, and B. Säfken, “Unsupervised document classification integrating web scraping, one-class SVM and LDA topic modelling,” *Journal of Applied Statistics*, vol. 50, no. 3, pp. 574–591, Feb. 2023, doi: 10.1080/02664763.2021.1919063.
[72] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, Pittsburgh Pennsylvania USA: ACM, July 1992, pp. 144–152. doi: 10.1145/130385.130401.
[73] J. Mourão-Miranda *et al.*, “Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine,” *NeuroImage*, vol. 58, no. 3, pp. 793–804, Oct. 2011, doi: 10.1016/j.neuroimage.2011.06.042.
[74] P. Schober, C. Boer, and L. A. Schwarte, “Correlation Coefficients: Appropriate Use and Interpretation,” *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.
[75] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, “Microarray data analysis: from disarray to consolidation and consensus,” *Nat Rev Genet*, vol. 7, no. 1, pp. 55–65, Jan. 2006, doi: 10.1038/nrg1749.
[76] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, “Cluster analysis and display of genome-wide expression patterns,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 25, pp. 14863–14868, Dec. 1998, doi: 10.1073/pnas.95.25.14863.
[77] P. D’haeseleer, “How does gene expression clustering work?,” *Nat Biotechnol*, vol. 23, no. 12, pp. 1499–1501, Dec. 2005, doi: 10.1038/nbt1205-1499.
[78] M. A. Zapala and N. J. Schork, “Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 51, pp. 19430–19435, Dec. 2006, doi: 10.1073/pnas.0609333103.
[79] S. R. Kim *et al.*, “Comparative proteomics: assessment of biological variability and dataset comparability,” *BMC Bioinformatics*, vol. 16, p. 121, Apr. 2015, doi: 10.1186/s12859-015-0561-9.

[80] T.-H. Tsai, M. Wang, and H. W. Ressom, “Preprocessing and Analysis of LC-MS-Based Proteomic Data,” in *Statistical Analysis in Proteomics*, vol. 1362, K. Jung, Ed., in Methods in Molecular Biology, vol. 1362. , New York, NY: Springer New York, 2016, pp. 63–76. doi: 10.1007/978-1-4939-3106-4_3.
[81] G. A. Satten, R. E. Tyx, A. J. Rivera, and S. Stanfill, “Restoring the Duality between Principal Components of a Distance Matrix and Linear Combinations of Predictors, with Application to Studies of the Microbiome,” *PLoS ONE*, vol. 12, no. 1, p. e0168131, Jan. 2017, doi: 10.1371/journal.pone.0168131.
[82] T. Fei and T. Yu, “scBatch: batch-effect correction of RNA-seq data through sample distance matrix adjustment,” *Bioinformatics*, vol. 36, no. 10, pp. 3115–3123, May 2020, doi: 10.1093/bioinformatics/btaa097.