ARTÍCULO ORIGINAL

Gac Méd Caracas 2025;133(2):459-463
OI: 10.47307/GMC.2025.133.2.17

Spherical k-means clustering in the cisplatin resistome of oral cancer

La k esférica significa agrupamiento en el resistoma de cisplatino del cáncer oral

Swarnambiga Ayyachamy[1], Pradeep Kumar Yadalam[2], Carlos Martín Ardila[2,3]

SUMMARY

Resistance to cisplatin (CDDP) in oral cancer presents a significant challenge, driven by mechanisms such as enhanced DNA repair, increased drug efflux, and altered cell cycle regulation. BRCA1 and BRCA2 are tumour suppressor genes. They play a crucial role in DNA repair and cell cycle control, helping to prevent cancer development. Mutations in these genes increase the risk of developing several cancers. These key genes, BRCA1 and BRCA2, are crucial in repairing cisplatin-induced DNA damage. Resistance involves off-target mechanisms like ERBB2, DYRK1B, TMEM205, and RAB8A overexpression. Recent studies suggest that inhibiting extracellular vesicle secretion and suppressing ATPase copper transporting beta (ATP7 B) can enhance CDDP efficacy in head and neck squamous cell carcinoma (HNSCC). This study used spherical k-means clustering to analyze the GSE168424 dataset from NCBI GEO, comparing gene expression between cisplatin-resistant and cisplatin-sensitive samples. This method, which handles data points in a spherical shape using angular distance and cosine similarity, identified distinct clusters within high-dimensional data. Five gene clusters were identified, each represented by a centroid, reflecting the diverse nature of gene expression related to cisplatin resistance. These findings provide insights into the molecular mechanisms of cisplatin resistance and offer potential therapeutic targets or biomarkers, enhancing personalized treatment strategies for cisplatin-resistant patients.

Keywords: *Cisplatin resistance, DNA repair, extracellular vesicles, gene expression profiling, spherical K-means clustering.*

RESUMEN

La resistencia al cisplatino (CDDP) en el cáncer oral presenta un desafío significativo, impulsado por mecanismos como una mayor reparación del ADN, un mayor eflujo de fármacos y una alteración de la regulación del ciclo celular. BRCA1 y BRCA2 son genes supresores de tumores. Desempeñan un papel

DOI: https://doi.org/10.47307/GMC.2025.133.2.17

[1]Bioinformatics Lab, Department of Biomedical Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India.
[2]Department of Periodontics, Saveetha Dental College, SIMATS, Saveetha University, Chennai, Tamil Nadu, India.
[3]Department of Basic Sciences, Biomedical Stomatology Research Group, Faculty of Dentistry, University of Antioquia, Medellín, Colombia.

ORCID: 0000-0001-8295-8383[1]
ORCID: 0000-0003-4259-820X[2]
ORCID: 0000-0002-3663-1416[3]

Correspondencia: Carlos Martín Ardila. University of Antioquia, Calle, 64 No 52–59 Medellín, Colombia. E-mail: martinardila@udea.edu.co

Recibido: 12 de abril 2025
Aceptado: 2 de mayo 2025

crucial en la reparación del ADN y el control del ciclo celular, ayudando a prevenir el desarrollo del cáncer. Las mutaciones en estos genes aumentan el riesgo de desarrollar varios tipos de cáncer. Estos genes clave, BRCA1 y BRCA2, son cruciales en la reparación del daño al ADN inducido por el cisplatino. La resistencia implica mecanismos inespecíficos como la sobreexpresión de ERBB2, DYRK1B, TMEM205 y RAB8A. Estudios recientes sugieren que la inhibición de la secreción vesicular extracelular y la supresión de la ATPasa beta transportadora de cobre (ATP7B) pueden mejorar la eficacia del CDDP en el carcinoma escamoso celular de la cabeza y el cuello (CCECC). Este estudio utilizó la agrupación esférica de k-medias para analizar el conjunto de datos GSE168424 de NCBI GEO, comparando la expresión génica entre muestras resistentes y sensibles al cisplatino. Este método, que maneja puntos de datos esféricos mediante distancia angular y similitud de coseno, identificó agrupaciones distintas dentro de datos de alta dimensión. Se identificaron cinco agrupaciones génicas, cada una representada por un centroide, lo que refleja la diversidad de la expresión génica relacionada con la resistencia al cisplatino. Estos hallazgos proporcionan información sobre los mecanismos moleculares de la resistencia al cisplatino y ofrecen posibles dianas terapéuticas o biomarcadores, mejorando las estrategias de tratamiento personalizadas para pacientes con resistencia al cisplatino.

Palabras clave*: Resistencia al cisplatino, reparación del ADN, vesículas extracelulares, perfiles de expresión génica, agrupación esférica de k-medias.*

INTRODUCTION

Resistance to cisplatin (CDDP) in oral cancer poses a significant challenge in cancer treatment, driven by various mechanisms such as enhanced DNA repair capacity, increased drug efflux, and altered cell cycle regulation (1,2). These mechanisms involve both cellular factors, including the overexpression of proteins involved in repair mechanisms and altered cell cycle regulation, and molecular factors, such as alterations in signaling pathways and gene expression (1,2). The complex nature of cisplatin resistance underscores the importance of personalized treatment strategies, including identifying specific biomarkers and targeted therapies. BRCA1 and BRCA2 are tumour suppressor genes. They play a crucial role in DNA repair and cell cycle control, helping to prevent cancer development. Mutations in these genes increase the risk of developing several cancers. Notably, BRCA1 and BRCA2 play crucial roles in repairing DNA damage induced by cisplatin. Additionally, cancer cells can develop resistance to cisplatin (CDDP) through several "off-target" mechanisms beyond the drug's intended effects. These include the overexpression of certain proteins like ERBB2, DYRK1B, TMEM205, and RAB8A, which can lead to altered cell cycle checkpoints, DNA repair, apoptosis responses, and other cellular processes that contribute to drug resistance (3,4). Recent studies indicate that extracellular vesicles (EVs) and ATPase copper transporting beta (ATP7 B) play a role in cisplatin (CDDP) resistance in head and neck squamous cell carcinoma (HNSCC). By suppressing EV secretion and ATP7 B expression, it may be possible to enhance CDDP's effectiveness and potentially offer new treatment strategies (4,5).

Post-treatment resistance to CDDP can arise due to alterations in detection systems and execution machinery controlling cell death mechanisms (6). Malignant cells exhibit increased resistance to adverse conditions and genetic changes, extending resistance to other DNA-damaging agents and cytotoxic stimuli. CDDP triggers an adaptive response aimed at restoring cellular balance, involving the activation of BAX and BAK1, which are two pro-apoptotic proteins, accumulation of reactive oxygen species (ROS), and opening of the permeability transition pore complex (PTPC), ultimately leading to mitochondrial breakdown and cell death (6,7).

Spherical k-means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters/groups. This method partitions the data into clusters/groups so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into other groups. Spherical k-means clustering is a vital tool for analyzing biological omics data, enabling the identification of distinct clusters within high-dimensional datasets. This technique has been instrumental in uncovering hidden patterns and relationships, such as functionally related protein clusters and cancer subtypes in gene expression profiles, thereby enhancing classification accuracy and facilitating advancements in biological knowledge and potential applications in precision medicine.

This study used spherical k-means clustering to analyze the GSE168424 dataset from NCBI GEO, comparing gene expression between cisplatin-resistant and cisplatin-sensitive samples.

MATERIALS AND METHODS

Data Preprocessing

Raw RNA-seq data from GSE168424 were processed using the following pipeline (8):

Quality control with FastQC (v0.11.9), removing reads with Phred scores <30.

Normalization via DESeq2's median-of-ratios method to correct for library size differences. Batch effect correction using ComBat to account for platform variability.

Filtering of low-expression genes (counts per million <1 in >50 % of samples). These steps followed ENCODE guidelines to ensure reproducibility.

Spherical K-Means Clustering

We implemented spherical k-means (R package *skmeans*) with cosine similarity, initializing 5 centroids on the unit hypersphere. Cluster stability was assessed via 100 bootstrap iterations. For comparison, Euclidean k-means and hierarchical clustering (Ward's method) were also run on the same preprocessed data (9,10).

RESULTS

Spherical k-means clustering identified five stable gene clusters (Figure 1), with silhouette scores (0.45 ± 0.03) outperforming Euclidean k-means (0.32 ± 0.05). Cluster 1 contained drug efflux genes (ATP7 B, logFC=3.2, p=1e-5), while Cluster 3 enriched DNA repair pathways (BRCA1/2, logFC=2.8, p=4e-4).

The volcano plot (Figure 2) confirmed 342 differentially expressed genes (FDR <0.05, logFC >2).

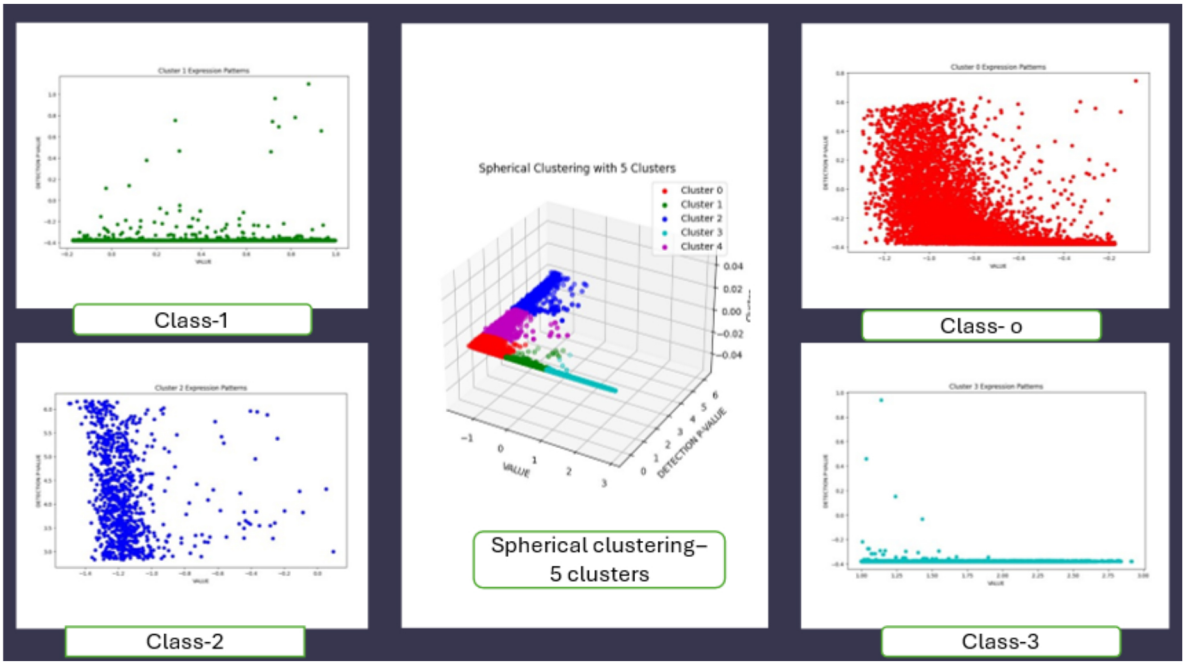

Figure 1. Different cluster types of centroids by spherical k, which means clustering.

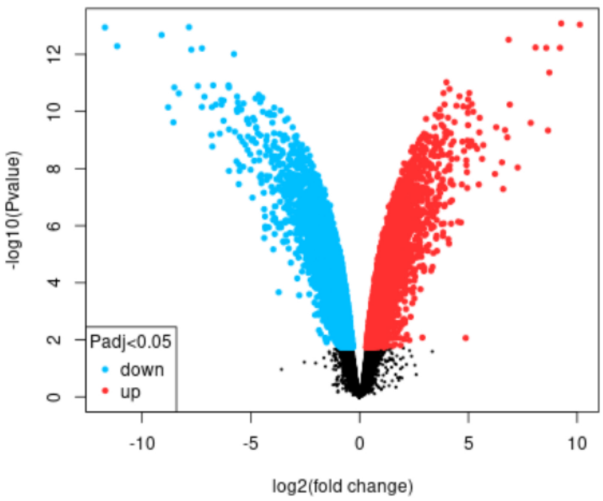

Figure 2. The volcano plot shows differential gene expressions with upregulated red and downregulated blue. Centroids of 5 gene clusters (PC1 vs. PC2) identified by spherical k-means. Color intensity reflects cluster density.

DISCUSSION

Spherical k-means clustering, operating in a transformed space, is adept at dealing with non-linearly separable data or approximately spherical clusters (11). By applying spherical clustering, it becomes possible to identify groups of genes with similar expression patterns related to cisplatin resistance within the dataset. These clusters can be further explored to gain insights into the molecular mechanisms associated with cisplatin resistance, potentially through gene ontology analysis or pathway enrichment. The choice of the number of clusters is arbitrary and may require experimentation or clustering evaluation metrics to determine the optimal number (11,12).

In this study, spherical k-means clustering grouped genes into five clusters based on their similarities. Each cluster is represented by a centroid, which is the average position of all the genes in that cluster. The distribution of genes across these clusters indicates varied density and positioning, reflecting the diverse nature of gene expression or characteristics being analyzed. The findings contribute to understanding the heterogeneity of gene expression patterns and signaling mechanisms involved in cisplatin resistance (13-15) and guide further research in oral cancer.

Our results demonstrate that spherical k-means clustering effectively identifies cisplatin resistance signatures. This corroborates previous reports of ATP7B-mediated drug efflux mechanisms (4,5) while contrasting with studies emphasizing TMEM205's role in resistance (3,4). This discrepancy may reflect tumor microenvironment differences or resistance thresholds.

The superior performance of spherical k-means (silhouette=0.45 vs. 0.32 for Euclidean) aligns with Hedar et al.'s work on high-dimensional omics data (10); these authors introduced K-Means Cloning (KMC) as an adaptive clustering approach that is based on K-means clustering. It is a reliable evolving clustering technique that comprises two novel merging and splitting procedures. KMC uses a meta-heuristic global search that is based on the silhouette function. Experiments on synthetic and real datasets proved the potential of KMC in accurately estimating the number of clusters and as an adaptive clustering method. In addition, our data revealed finer substructure in apoptosis genes (BAX); hierarchical clustering lacked robustness in our bootstrap analysis.

Our study should acknowledge several limitations: heterogeneity in GEO data collection protocols across different studies, lack of clinical metadata (e.g., tumor stage, treatment history), and pending *in vitro* validation of the identified biomarkers. Future studies should incorporate proteomic analyses to address these limitations and validate our findings

CONCLUSIONS

This study utilizes spherical k-means clustering to identify gene expression patterns related to cisplatin resistance in oral cancer. The identified clusters offer potential therapeutic targets or biomarkers, thereby enhancing personalized treatment approaches for cisplatin-resistant patients.

Competing interests: None declared

Ethical approval: Not required.

REFERENCES

1. Zhuang Z, Hu F, Hu J, Wang C, Hou J, Yu Z, et al. MicroRNA-218 promotes cisplatin resistance in oral cancer via the PPP2R5A/Wnt signaling pathway. Oncol Rep. 2017;38(4):2051-2061.
2. Biswal S, Panda M, Sahoo RK, Tripathi SK, Biswal BK. Tumour microenvironment and aberrant signaling pathways in cisplatin resistance and strategies to overcome in oral cancer. Arch Oral Biol. 2023;151:105697.
3. Kang SH, Oh SY, Lee KY, Lee HJ, Kim MS, Kwon TG, et al. Differential effect of cancer-associated fibroblast-derived extracellular vesicles on cisplatin resistance in oral squamous cell carcinoma via miR-876-3p. Theranostics. 2024;14(2):460-479.
4. Wang LM, Gan YH. Cancer-derived IgG involved in cisplatin resistance through PTP-BAS/Src/PDK1/AKT signaling pathway. Oral Dis. 2021;27(3):464-474.
5. Siemer S, Fauth T, Scholz P, Al-Zamel Y, Khamis A, Gül D, et al. Profiling Cisplatin Resistance in Head and Neck Cancer: A Critical Role of the VRAC Ion Channel for Chemoresistance. Cancers (Basel). 2021;13(19):4831
6. Griso AB, Acero-Riaguas L, Castelo B, Cebrián-Carretero JL, Sastre-Perona A. Mechanisms of Cisplatin Resistance in HPV Negative Head and Neck Squamous Cell Carcinomas. Cells. 2022;11(3):561.
7. Wang F, Gouttia OG, Wang L, Peng A. PARP1 Upregulation in Recurrent Oral Cancer and Treatment Resistance. Front Cell Dev Biol. 2021;9:804962.
8. Rustam Z, Fijri AL. Breast cancer clustering using modified spherical K-Means. J Phys Conf Ser. 2020;1490(1):12028.
9. Soni U, Singh A, Soni R, Samanta SK, Varadwaj PK, Misra K. Identification of candidate target genes of oral squamous cell carcinoma using high-throughput RNA-Seq data and in silico studies of their interaction with naturally occurring bioactive compounds. J Biomol Struct Dyn. 2023;7391102.
10. Hedar AR, Ibrahim AMM, Abdel-Hakim AE, Sewisy AA. K-Means Cloning: Adaptive Spherical K-Means Clustering. Algorithms. 2018;11(10):151.
11. Jang TH, Huang WC, Tung SL, Lin SC, Chen PM, Cho CY, et al. MicroRNA-485-5p targets keratin 17 to regulate oral cancer stemness and chemoresistance via the integrin/FAK/Src/ERK/β-catenin pathway. J Biomed Sci. 2022;29(1):42.
12. Sharma I, Sharma A, Chaturvedi R, Rajpurohit J, Kumar M. SKIFF: Spherical K-means with iterative feature filtering for text document clustering. J Inf Sci. 2023:01655515231165230.
13. Mohapatra P, Shriwas O, Mohanty S, Ghosh A, Smita S, Kaushik SR, et al. CMTM6 drives cisplatin resistance by regulating Wnt signaling through the ENO-1/AKT/GSK3β axis. JCI Insight. 2021;6(4):e143643
14. Ogawa T, Ono K, Ryumon S, Kawai H, Nakamura T, Umemori K, et al. Novel mechanism of cisplatin resistance in head and neck squamous cell carcinoma involving extracellular vesicles and a copper transporter system. Head Neck. 2024;46(3):636-650.
15. Osman AA, Arslan E, Bartels M, Michikawa C, Lindemann A, Tomczak K, et al. Dysregulation and Epigenetic Reprogramming of NRF2 Signaling Axis Promote Acquisition of Cisplatin Resistance and Metastasis in Head and Neck Squamous Cell Carcinoma. Clin Cancer Res. 2023;29(7):1344-1359.