The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Multi-Omics and In Silico Drug Discovery Approaches to Identify Novel Biomarkers of Early Metastasis in Melanoma

Samaya Vaidya

The Cathedral and John Connon School

Dr. Begüm-Akman Tuncer

Abstract

Melanoma is one of the most aggressive skin cancers, with early metastasis being the primary cause of patient mortality. Despite advances in targeted and immunotherapies, reliable biomarkers that predict early metastatic potential remain unclear. This study sought to identify novel genes associated with melanoma metastasis through an integrative bioinformatics pipeline combining transcriptomic, network, pathway, and drug discovery analyses. Differential expression analysis of the GSE7553 dataset identified 25 significant DEGs ($p < 0.05$, $|\log2FC| > 2$), among which keratin and small proline-rich protein (SPRR) family members (KRT6A, KRT16, KRT17, SPRR1A, and SPRR3) were consistently upregulated in metastatic samples. Protein-protein interaction (PPI) networks and pathway enrichment analyses highlighted their central roles in

keratinisation, cornified envelope formation, and developmental biology pathways, suggesting dysregulation of epidermal differentiation as a driver of tumour progression. Cross-validation with independent datasets and OncoPrint analysis confirmed their clinical relevance, while Kaplan-Meier survival curves linked high KRT17 and SPRR3 expression with poor prognosis. Finally, all five candidate genes were evaluated using Drugnome AI to determine their raw scores and percentile rankings for druggability as small-molecule targets in an oncogenic setting. Based on this analysis, the two most promising druggable candidates were identified and prioritised for subsequent wet-lab validation.

Keywords: Melanoma, Early Metastasis, Multi-omics, Bioinformatics, in Silico Drug Discovery

Introduction

Melanoma, a malignant neoplasm arising from melanocytes, is one of the most aggressive forms of skin cancer, known primarily for its rapid progression and tendency to metastasise early, the primary cause of mortality in affected patients (Waseh & Lee, 2023). Skin cancer is currently the most commonly diagnosed group of cancer worldwide, with malignant melanoma accounting for almost 20% of these cancers - approximately 325,000 cases worldwide as of 2020 (Arnold et al., 2022). It is likely that most of these are due to increased exposure to UV radiation, a known carcinogen.

Metastatic melanoma, specifically stage III and stage IV diseases, carries an inferior prognosis (Waseh & Lee, 2023). Though there have been certain advancements in targeted therapy and immunotherapy, extremely reliable biomarkers that predict early metastasis remain

undiscovered. Current treatments include BRAF/MEK inhibitors and immune checkpoint blockade, extending survival in many patients (Sorino et al., 2024). However, resistance and lack of early predictive markers remain significant problems in treating melanoma. Importantly, there remains a pressing need for reliable early metastasis biomarkers to enable timely intervention and improve outcomes.

The clinical impact of early metastasis in melanoma is profound. Five-year survival drops drastically from over 99% for localised disease to just 35% for distant metastases (Sandru et al., 2014). For advanced-stage patients - particularly those with brain or visceral involvement - prognosis remains dismal, with median survival often less than one year and five-year survival rates as low as 5-19% (Sandru et al., 2014). These statistics underscore the urgent need for predictive biomarkers that can signal metastatic risk early, enabling timely and potentially life-saving interventions.

Previous studies have explored gene expression signatures and driver mutations such as BRAF, NRAS, and NF1 in melanoma progression (Rossi et al., 2019). However, most have focused solely on single-omic datasets or limited gene panels without integrating multiple parallel pipelines. The publicly available microarray dataset GSE7553, found on GEO2R, which profiles normal skin, primary melanoma, and metastatic samples, provides a valuable opportunity for such integrative analysis.

This study aims to identify novel genes associated with early metastasis in melanoma through an integrative analysis of multi-omics data, including gene expression, mutation, and copy number variation. By comparing the molecular profiles of primary and metastatic

melanoma tumours, the study aims to identify differentially expressed genes (DEGs) that may drive early tumour spread. The most significant candidate genes will then be evaluated for therapeutic relevance using protein-protein interaction analysis and molecular docking. This combined bioinformatics and in silico drug discovery approach aims to highlight new, targetable biomarkers that could inform future strategies for metastasis prevention and treatment in melanoma.

Methods

This section describes the hypothesis testing framework of this research in the subsequent sections:

Data Collection

Gene Expression Omnibus (GEO) was employed to download microarray gene expression data based on keywords associated with melanoma, including "Melanoma," "Primary Melanoma," and "Metastatic Melanoma." The dataset with GEO accession ID GSE7553 was chosen from the GEO database because it includes gene expression profiles of normal skin, primary melanoma, and metastatic melanoma samples. GEO2R, an online interactive web-based program offered by NCBI, was employed to do differential gene expression analysis using metastatic melanoma as the 'test' and primary melanoma as the 'control' group. The most significant DEGs were then downloaded for further consideration.

Genomic alteration information (mutations and copy number alterations) for shortlisted genes was accessed from The Cancer Genome Atlas - Skin Cutaneous Melanoma (TCGA-SKCM)

data through cBioPortal. Protein expression information was accessed through The Human Protein Atlas and cross-validated using MEL-PLOT.

For drug discovery, a library of 12,206 drug compounds was downloaded from ChEMBL by choosing under 'Small Molecules' the options Phase 1, Phase 2, Phase 3, Early Phase 1, and Approved. A second library of 21,931 compounds was downloaded from the Therapeutic Target Database (TTD) on PubChem. Databases were selected for their large coverage of bioactive molecules and their general popularity as sources for drug discovery research. The RCSB Protein Data Bank (PDB) was utilised to retrieve the 3D structure of the identified therapeutic target proteins.

Network Analysis

The DEGs of significant interest were submitted to STRING (Version 12.0) for the generation of protein–protein interaction (PPI) networks, with the addition of 5 and 10 interactors. Two network files were retrieved. STRING provides a database of known and predicted protein–protein interactions. The networks were visualised in Cytoscape (version 3.10.2) and examined with the AnalyzeNetwork app. Topological network parameters, including degree, betweenness centrality, and clustering coefficient, were employed to determine hub genes. The first 25 genes of each of the 2 networks were listed, and the first 10 genes shared across all networks that were also significantly differentially expressed were selected for further research.

Pathway Analysis

The list of genes from the most comprehensive and best-networked network (50 added interactors) was submitted to the Reactome Pathway Database (Version 89) to determine

significantly overrepresented pathways. Significant pathway results were also obtained without adding interactors for comparison. The occurrence of the shortlisted hub genes in these enriched pathways was validated, and overrepresented pathways with possible involvement in early melanoma metastasis were determined.

Druggability Assessment of Candidate Genes

All five candidate genes were analysed using Drugnome AI in order to assess their raw scores and percentile ranks when being tested as druggable targets for small molecules to bind to in an oncogenic context. Finally, from this, the top two druggable targets were shortlisted for wet-lab analysis.

Results and Discussion

This section presents the results obtained from the bioinformatics analysis of differentially expressed genes (DEGs), their pathways, and network interactions, providing a broader understanding of molecular mechanisms in melanoma and testing the hypothesis that key keratin-related genes are involved in disease progression.

Differential gene expression analysis

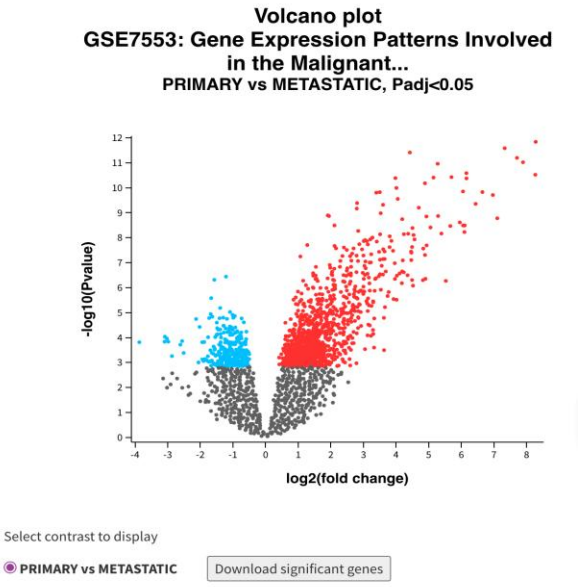
Using the RNASeq dataset GSE7553, differential expression analysis was performed using GEO2R. The expression of DEGs was visualised in the form of a volcano plot (Figure 1). In total, 25 DEGs were identified ($p < 0.05$, $|\log2FC| > 2$). A subset of five genes - SPRR1A, SPRR3, KRT6A, KRT16, and KRT17 - emerged as consistently upregulated and significant across

analyses. These were prioritised for further pathway and network investigation using STRING, Cytoscape

Figure 1 The volcano plot of differentially expressed genes in Melanoma. This plot shows the gene expression in blood from primary and metastatic samples with Thalassemia (GSE7553), identified using the GEO2R tool., and Reactome Pathway Analysis.

Figure 1

The Volcano Plot of Differentially Expressed Genes in Melanoma



This plot shows the gene expression in blood from primary and metastatic samples with Thalassemia (GSE7553), identified using the GEO2R tool.

Network and Pathway Analysis

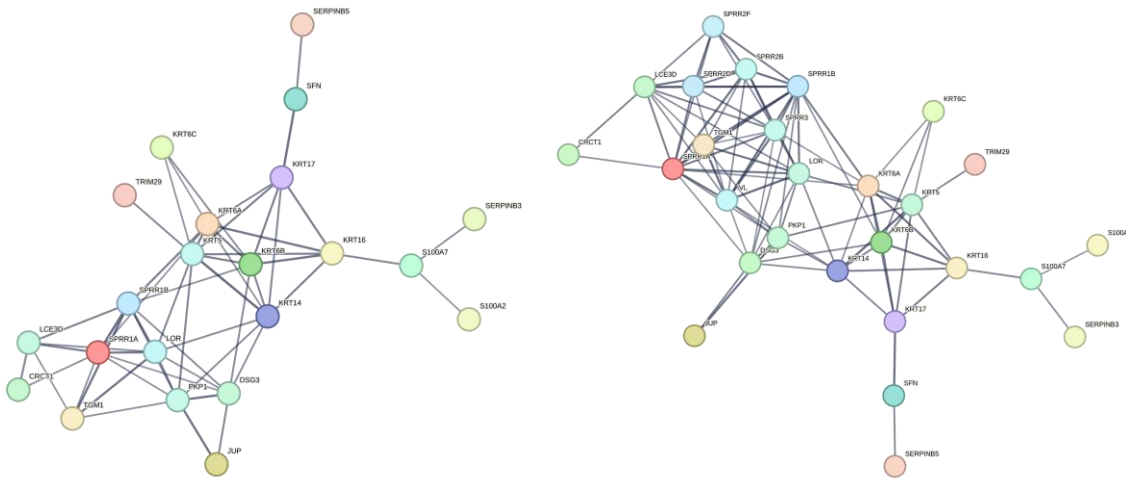
Protein-protein interaction (PPI) networks of significantly differentially expressed genes were generated using the STRING database and further analysed in Cytoscape. Networks were constructed with 5 and 10 added interactors (Figure 2 and Figure 3). The 10-interactor network exhibited a higher degree of connectivity compared to the 5-interactor network, suggesting that increasing the number of interactors allows better capture of disease-relevant associations in melanoma.

Topological analysis was performed in Cytoscape using measures such as degree, betweenness centrality, and clustering coefficient. After the genes were listed in order, they were assigned a numerical value, which was used to highlight the most central and well-connected genes, often considered hub genes within biological networks. Across both networks, SPRR1A, SPRR3, KRT6A, KRT16, and KRT17 consistently ranked highest in these measures and were significantly upregulated in melanoma datasets. Their central positions within the STRING-Cytoscape networks underscore their potential role as key regulators in melanoma progression and provide a strong basis for downstream pathway analysis.

Figure 2

The Interaction Network of Differentially Expressed Genes

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200



(A) **(B)**

(A) This shows an interaction network of differentially expressed genes with 5 added interactors in STRING. (B) This shows an interaction network of differentially expressed genes with 10 added interactors in STRING.

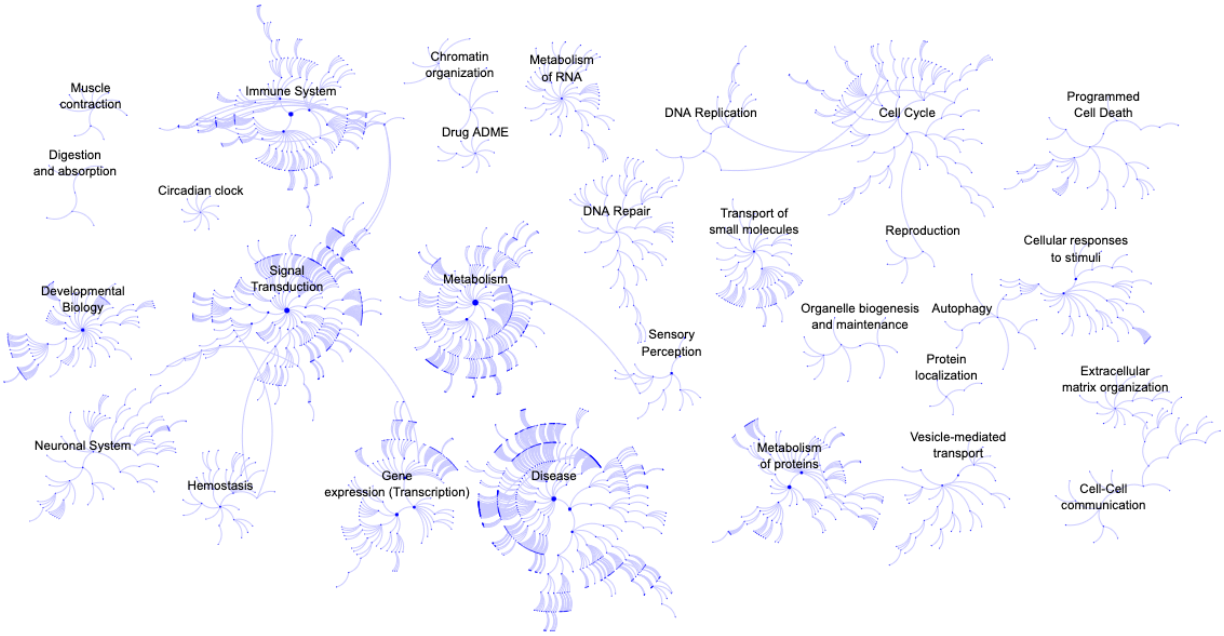
Pathway analysis revealed significant enrichment in the *Formation of the cornified envelope*, *Keratinisation*, and *Developmental Biology*. These pathways are largely driven by keratin and small proline-rich protein (SPRR) family members, highlighting widespread dysregulation of epidermal differentiation processes in melanoma. Specifically, keratin genes such as KRT6A, KRT16, and KRT17 consistently appeared in enriched pathways, indicating perturbations in cytoskeletal organisation and structural integrity of melanoma cells. Similarly, SPRR1A and SPRR3, normally involved in terminal differentiation and cornified envelope formation, were also enriched, suggesting aberrant remodelling of epithelial barrier functions during tumour progression.

The *Formation of the cornified envelope* pathway is particularly noteworthy, as it is typically associated with late epidermal differentiation and protective barrier formation. Its enrichment in melanoma suggests that tumour cells may co-opt or dysregulate this program to alter adhesion, invasion, or stress resistance. *Keratinisation* and *Developmental Biology* pathways further reflect a broader theme of disrupted cellular differentiation, where melanoma cells appear to adopt a more plastic phenotype, oscillating between epithelial-like and invasive states. Together, these findings underscore the central role of structural and developmental pathways in driving melanoma progression and highlight keratin/SPRR genes as potential biomarkers or therapeutic targets.

Figure 3

Pathway Overview Highlighting the Enrichment of Shortlisted Genes in Keratinisation and Formation of the Cornified Envelope, Key Processes Disrupted in Melanoma

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200



Transcriptomics

To corroborate the results of the differential expression and pathway analyses, gene expression and survival data were further examined using the UALCAN platform. All five shortlisted genes (KRT6A, KRT14, KRT17, SPRR1A, and SPRR3) were significantly upregulated in melanoma samples compared to normal skin, reinforcing the robustness of the DEG analysis.

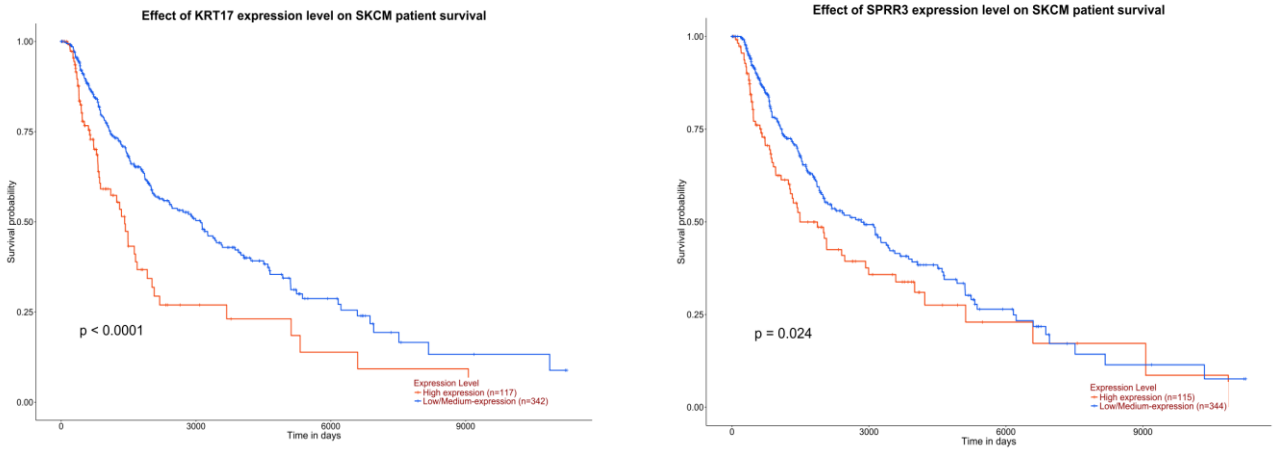
Expression boxplots consistently demonstrated markedly elevated transcript levels across melanoma cohorts. Kaplan-Meier survival analyses further revealed that higher expression of several candidates, most notably KRT17 (Figure 4) and SPRR3 (Figure 5), was associated with poorer overall survival. In particular, SPRR3 expression correlated with increased metastatic

potential, suggesting that its dysregulation may actively contribute to melanoma progression and dissemination.

Where available, phosphoproteomic data provided additional mechanistic insight, indicating that post-translational modifications could further regulate these keratin and SPRR-family proteins, thereby modulating cytoskeletal dynamics, barrier function, and signallingcascades implicated in tumour development and metastasis.

Figure 4

Kaplan-Meier Survival Analysis (A and B) Showing that High KRT17 and SPRR3 Expression Correlates with Significantly Poorer Survival, Indicating its Potential Role in Metastasis.



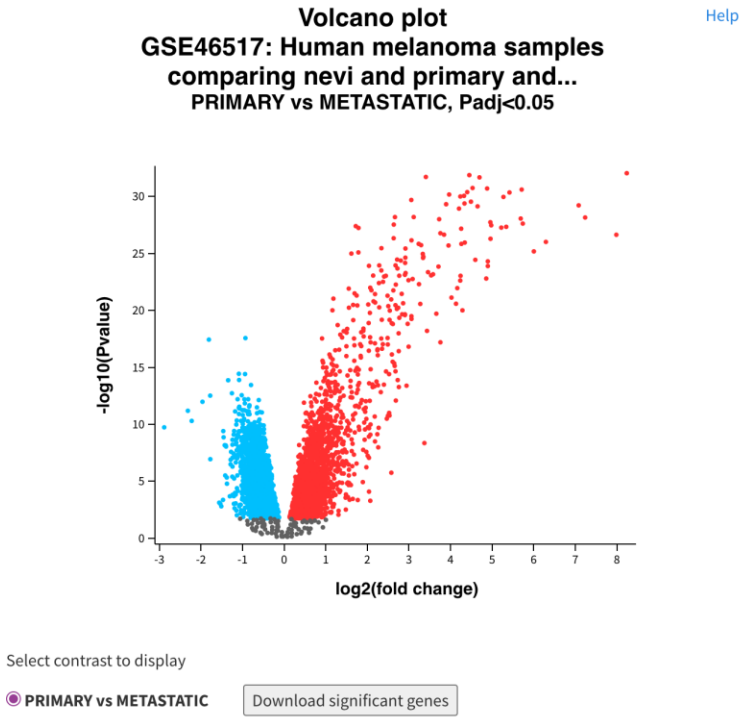
Cross-Validation

To further cross-validate the associations of survival, the GSE46517 dataset was used as an independent cohort. The five genes with the most significant differences in expression (KRT6A,

KRT14, KRT17, SPRR1A, and SPRR3) (Figure 6) were identified again in this dataset, confirming that their relevance was not merely limited to a single dataset, but can rather be generalised.

Figure 6

The Volcano Plot of Differentially Expressed Genes in Melanoma



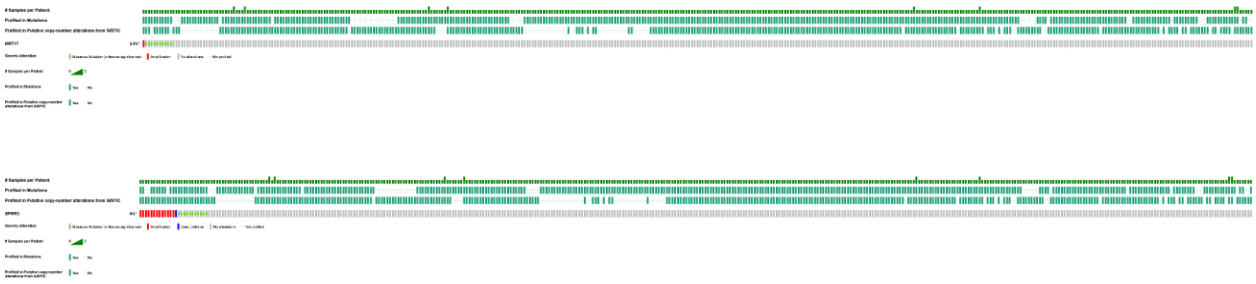
This plot shows the gene expression in blood from primary and metastatic samples with Thalassemia (GSE46517), identified using the GEO2R tool.

Next, further validation was conducted using their OncoPrint plots (Figures 7 and 8). This provided a clear overview of genomic events such as amplifications, deletions, and mutations in

melanoma samples. While most samples showed no real structural changes, the amplifications seen are consistent with the transcript upregulation in melanoma tissues. Though all five genes were studied for cross-validation, this paper depicts two out of the five for clarity. This is done to illustrate the research pipeline and provide a representative sample of how expression, validation, and survival associations were examined.

Figure 7

OncoPrint Visualisation of SPRR3 and KRT17 Alterations in Melanoma



The plot illustrates the spectrum and frequency of genomic alterations in SPRR3 and KRT17 across melanoma patient samples, including amplifications, deletions, and mutations.

Druggability Assessment of Candidate Genes

To delve deeper into the therapeutic relevance of the five melanoma-associated validated genes, DrugnomeAI was employed. The Cancer-SM score for each gene was recorded. This predicts the likelihood that proteins can be targeted by small-molecule drugs in oncology. By studying both the raw score and percentile ranks in relation to the entire set of protein-coding

genes, a deeper analysis was possible. Genes that rank above the 70th percentile have been previously seen to be enriched for known oncogenes and FDA-approved drug targets, hence serving as a valid threshold for drugability.

Table 1

Cancer-SM Druggability Predictions for Five Candidate Genes in Melanoma

Gene	Raw Score	Percentile Rank
KRT6A	0.0263	68.75
KRT16	0.0425	72.50
KRT17	0.0853	77.81
SPRR1A	0.0025	42.34
SPRR3	0.0058	54.04

Results showed that KRT16 (72.5th percentile) and KRT17 (77.8th percentile) exceeded the cutoff, while KRT6A (68.8th percentile) closely approached it, suggesting a possible degree of drugability. Contrastingly, SPRR1A and SPRR3 scored extremely low, indicating that it has low therapeutic potential. Hence, KRT16 and KRT17 were discovered to be potentially druggable oncogenic targets.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Conclusion

This study employed a multi-omics and in silico approach to identify novel biomarkers of early metastasis in melanoma. Differential gene expression analysis of the GSE7553 and GSE46517 datasets highlighted five consistently upregulated genes - KRT6A, KRT16, KRT17, SPRR1A, and SPRR3 - with possible roles in early metastasis of melanoma. Next, network and pathway analysis revealed the central role of these genes in keratinisation, cornified envelope formation, and developmental biology pathways, indicating that dysregulation contributes to metastatic potential. Survival analysis further linked high expression of all the genes to poor prognosis, supporting clinical relevance. Druggability assessment using DrugnomeAI identified KRT16 and KRT17 as the most probable targets, exceeding the 70th percentile threshold.

Overall, these findings establish a preliminary framework for prioritising keratin family members for wet-lab validation of oncogenic function. Future experimental work could potentially offer more insight into their causal roles in early melanoma and assess the feasibility of targeting them through drugs, thereby directly addressing the research question of how molecular markers can predict and potentially modulate primary metastasis in melanoma.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

References

Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., Cust, A. E., de Vries, E., Whiteman, D. C., & Bray, F. (2022). Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. *JAMA dermatology*, *158*(5), 495–503. https://doi.org/10.1001/jamadermatol.2022.0160

Rossi, A., Roberto, M., Panebianco, M., Botticelli, A., Mazzuca, F., & Marchetti, P. (2019). Drug resistance of BRAF-mutant melanoma: Review of up-to-date mechanisms of action and promising targeted agents. *European journal of pharmacology*, *862*, 172621. https://doi.org/10.1016/j.ejphar.2019.172621

Sandru, A., Voinea, S., Panaitescu, E., & Blidaru, A. (2014). Survival rates of patients with metastatic malignant melanoma. *Journal of medicine and life*, *7*(4), 572–576.

Sorino, C., Iezzi, S., Ciuffreda, L., & Falcone, I. (2024). Immunotherapy in melanoma: advances, pitfalls, and future perspectives. *Frontiers in molecular biosciences*, *11*, 1403021. https://doi.org/10.3389/fmolb.2024.1403021

Waseh, S., & Lee, J. B. (2023). Advances in melanoma: epidemiology, diagnosis, and prognosis. *Frontiers in medicine*, *10*, 1268479. https://doi.org/10.3389/fmed.2023.1268479