ACCSCIENCE
PUBLISHING

Tumor Discovery

ISSN: 3060-8597 (Print)
ISSN: 2810-9775 (Online)
Volume 2 • Issue 1
March 2023

Tumor Discovery

Print ISSN: 3060–8597
Online ISSN: 2810–9775

Tumor Discovery is a peer-reviewed and open-access journal that aims to present new cancer research with strong emphasis on fundamental and translational studies. *Tumor Discovery* covers topics such as etiology and pathogenesis of cancer, mechanisms and molecular pathways underlying cancer initiation and progression, tumor metastasis, etc.

Scan to access website:

Scan to submit papers:

About the Publisher

AccScience Publishing is a publishing company based in Singapore. We publish a range of high-quality, open-access, peer-reviewed journals and books from a broad spectrum of disciplines.

Contact Us

Managing Editor
td.office@accscience.sg

AccScience Publishing
8 Burn Road, #15-03 Trivex, Singapore 369977.

Volume 2 • Issue 1 • March 2023
ISSN 3060-8597 (print) ISSN 2810-9775 (online)

TUMOR DISCOVERY

ACCSCIENCE PUBLISHING

Access Science Without Barriers

ACCSCIENCE
PUBLISHING

Full issue copyright © 2023 AccScience Publishing
All rights reserved. Without permission in writing from the publisher, this full issue publication in its entirety may not be reproduced or transmitted for commercial purposes in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system. Permissions may be sought from td.office@accscience.sg.

Article copyright © Respective Author(s)
See articles for copyright year. All articles in this full issue publication are open-access. There are no restrictions in the distribution and reproduction of individual articles, provided the original work is properly cited. However, permission to reuse copyrighted materials of an article for commercial purposes is applicable if the article is licensed under Creative Commons Attribution-NonCommercial License. Check the specific license before reusing.

TUMOR DISCOVERY

ISSN: 3060-8597 (print)
ISSN: 2810-9775 (online)

Editorial and Production Credits

Publisher: AccScience Publishing
Managing Editor: Zoe Zhang
Production Editor: Sharmila Velapasamy
Journal Development Editor: Felicia Wang
Special Issue Commissioning Editor: Hannah Zhang
Article Layout and Typeset: Sinjore Technologies (India)
Cover Design: ProPub (China)

For all advertising queries, contact td.office@accscience.sg.

Supplementary file

Supplementary files of articles can be obtained at https://accscience.com/journal/TD/2/1.

Disclaimer

AccScience Publishing is not liable to the statements, perspectives, and opinions contained in the publications. The appearance of advertisements in the journal shall not be construed as a warranty, endorsement, or approval of the products or services advertised and/or the safety thereof. AccScience Publishing disclaims responsibility for any injury to persons or property resulting from any ideas or products referred to in the publications or advertisements. AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tumor Discovery

Editorial Board

Editors-in-Chief

Yan Hou, *China*
Helmut H. Popper, *Austria*
Mingzhu Yin, *China*

Associate Editors

Jan B. Vermorken, *Belgium*
Zhimin Bian, *China*
Shuangqi Cai, *China*
Paolo Caliceti, *Italy*
Amancio Carnero Moya, *Spain*
Silvia Deaglio, *Italy*
Jinhai Deng, *UK*
Emilio Hirsch, *Italy*
Sung-hoon Kim, *Korea*
Jesang Ko, *South Korea*
Massimo Libra, *Italy*
Yong-beom Lim, *South Korea*
Tian-Jie Lyu, *China*
Wenping Ma, *China*
Fabio Malavasi, *Italy*
Kishor Pant, *USA*
Athanasios Papavassiliou, *Greece*
Silvia R Rogatto, *Denmark*
Alfred Sze Lok Cheng, *China*
João T. Barata, *Portugal*
Youtao Yu, *China*
Xin Zhao, *China*

*Editorial Board Members**

Ahmed Abu-Zaid, *USA*
Zohreh Amoozgar, *USA*
Hugo Arias-Pulido, *USA*
Nicolae Bacalbasa, *Romania*
Meriem Bahri, *UK*
Armand Bensussan, *France*
Prashanth K.B. Nagesh, *USA*
Paolo Boffano, *Italy*
Roberta Bortolozzi, *Italy*
Steven Brower, *USA*
Jian Cao, *USA*
Darren R Carpizo, *USA*
Min Soon Cho, *USA*
Lili Cui, *China*
Jennifer A. Doll, *USA*
Bertani Emilio, *Italy*
Luca Ermini, *Luxembourg*
Marco Falasca, *Australia*
Ana Faustino, *Portugal*
Pierfrancesco Franco, *Italy*
Niccola Funel, *Italy*
Jean Gabert, *France*
Francesca Giordano, *Italy*
Zhaohui Gong, *China*
Qinghua Guo, *China*
Ken H Young, *USA*
Jens Claus Hahne, *UK*
W. Hohenforst-Schmidt, *Germany*
Peter Huppert, *Germany*
Kiss István, *Hungary*
Weilin Jin, *China*
Kalevi Kairemo, *USA*
M.A. Kamal, *Saudi Arabia*
Dionyssios Katsaros, *Italy*
Ilya Klabukov, *Russia*
Koji Komori, *Japan*
Omer Kucuk, *USA*
Jong-Young Kwak, *Korea*
Seok-geun Lee, *Korea*
Sukmook Lee, *South Korea*
Robert Leonard, *UK*
Zhipin Liang, *USA*
Yifei Liu, *China*
Jose Manuel Lopes, *Portugal*
Domenica Mangieri, *Italy*
Francesco Marampon, *Italy*
Cioce Mario, *Italy*
Conti Matteo, *Italy*
Ammendola Michele, *Italy*
Maria Beatrice Morelli, *Italy*
Moe Muhith, *UK*
Atsushi Otsuka, *Japan*
Gabriela Raso, *USA*
Erle Robertson, *USA*
Giovanni Rosti, *Italy*
Ravi P. Sahu, *USA*
Ahmad Sayasneh, *UK*
A. Schonthal, *USA*Dian Wang, *USA*
Gautam Sethi, *Singapore*
Vishal Shelat, *Singapore*
Jingdong Shi, *China*
Xiaoyu Shi, *China*
Alexander Shtil, *Russia*
Hifzur R Siddique, *India*
Cynthia Simbulan-Rosenthal, *USA*
Zheng Song, *China*
Maria Patrizia Stoppelli, *Italy*
S. Subramanian, *Ethiopia*
Myron Szewczuk, *Canada*
Maria Teresa Vietri, *Italy*
Qiujun Wang, *China*
Yanjun Wei, *Texas*
Guifang Xu, *China*
Yan Xu, *China*
Jun Xu, *China*
Qin Yan, *USA*
Huike Yang, *China*
Bin Yi, *USA*
Chunyang Zhang, *China*
Meiling Zhang, *USA*
Xinyuan Zhao, *China*
Shaoquan Zheng, *China*
Xingang Zhou, *China*
Massimo Zollo, *Italy*

Youth Editorial Board

Tariq A. Bhat, *USA*
Yiyang Chen, *China*
Xinpei Deng, *China*
Angelo Corso Faini, *Italy*
Alessandra Ferraresi, *Italy*
Jindong Xie, *China*

ACCSCIENCE
PUBLISHING

Volume 2 | Issue 1 | March 2023 |

Tumor Discovery

https://accscience.com/journal/TD

CONTENTS

REVIEW ARTICLES

1 **Receptors of advanced glycation end products in oral squamous cell carcinoma: A systematic review**
Sinduja Palati, Pratibha Ramani, Saravanan Sekaran

2 **An overview of the risk, underlying factors, and mechanism of cancer progression in polycystic ovary syndrome**
Rubiat Afrin Ayon, Md. Azmain Faike, Sumiya Zaman Ononna, Aminul Hassan, Shoumik Kundu, Farhana Akhter, Mohammad Mahfuz Ali Khan Shawan, Md. Mozammel Hossain, Md. Ibrahim Khalil, Md. Ashraful Hasan

ORIGINAL RESEARCH ARTICLES

3 **An approach for classification of lung nodules**
Naveen HM, Naveena C, Manjunath Aradhya VN2

4 **Malignant versus normal breast tissue: Optical differentiation exploiting hyperspectral imaging system**
Mohamed Hisham Aref, Ibrahim H. Aboughaleb, Abdallah Abdelkader Hussein, Ayman Mohammed Farag, Sara Abd El-Ghaffar, Yasser H. El-Sharkawy

5 **A novel gene prognostic signature lymphocyte cytosolic protein 2 regulates melanoma progression by activating tumor-infiltrating CD8$^+$ T-cells through the interferon regulatory factor 5 signaling pathway**
Hongyin Sun, Kui Deng, Xingchen Zhou, Dongsheng Cao, Yan Cheng, Xiang Chen, Mingzhu Yin

ACCSCIENCE PUBLISHING

Tumor Discovery

REVIEW ARTICLE

Receptors of advanced glycation end products in oral squamous cell carcinoma: A systematic review

Sinduja Palati*, Pratibha Ramani, and Saravanan Sekaran

Department of Oral and Maxillofacial pathology, Saveetha Dental College and Hospitals, Saveetha Institute for Medical and Technical Sciences, Chennai, Tamil Nadu, India

Abstract

Oxidative stress markers have been shown to be elevated in oral squamous cell carcinomas; plays a crucial role in the build-up of advanced glycation end-receptors of advanced glycation end (AGE-RAGE) products; and has been shown to exacerbate cellular dysfunction, vascular change, apoptosis, and activate inflammatory pathways. The purpose of this study is to assess comprehensively the involvement of RAGE in oral squamous cell malignancies. The findings imply that these receptors and their associated ligands play a significant role in the growth and spread of the tumor, hence impacting the prognosis and life expectancy of the affected individual. This comprehensive review uncovers promising evidence for the clinical use of these molecules, such as RAGEs, in prognostic considerations or as molecular targets for therapy. The available literature shows a role for RAGE in invasion, migration, and angiogenesis in oral cancers. These preliminary findings are encouraging for the therapeutic use of these molecules for prognostic considerations or molecularly targeted therapy.

Keywords: Receptors of advanced glycation end products; Advanced glycation end products; Oral squamous cell carcinoma; Oral cancers

***Corresponding author:**
Sinduja Palati
(sindujap.sdc@saveetha.com)

Citation: Palati S, Ramani P, Sekaran S, 2023, Receptors of advanced glycation end products in oral squamous cell carcinoma: A systematic review. *Tumor Discov*, 2(1): 244. https://doi.org/10.36922/td.244

Received: November 2, 2022
Accepted: January 27, 2023

1. Introduction

Oral squamous cell carcinoma (OSCC) is the sixth most common type of cancer in the world. OSCC accounts for at least 40% of all cancers; it is especially prevalent in India and Sri Lanka. OSCC is a complex cancer in which genetic mutations, environmental factors, and other risks contribute considerably to its malignancy[1]. Local growth and lymph node metastasis are intrinsically linked to the malignant potential of OSCC[2]. It has been established that oxidative stress is enhanced in oral cancers and plays a significant role in the build-up of several toxic substances. Oxidative stress is characterized by an imbalance between the production of free radicals and antioxidants. In addition, it is illustrated by a phenomenon implicated in the pathophysiology of numerous inflammatory disorders and cancers.

Glycation is a non-enzymatic, spontaneous interaction between free reducing sugars and free amino groups of proteins, DNA, and lipids that result in the creation of Amadori products. Multiple irreversible dehydration and rearrangement result in the formation of advanced glycation end products (AGEs). It has been demonstrated that the presence of AGEs increases cellular dysfunction, vascular alteration, apoptosis, and activation of inflammatory pathways. Receptors of advanced glycation end products (RAGEs) are a receptor for several ligands, including amphoterin, advanced glycation products, b-amyloids, and S100 proteins[3,4].

The gene for RAGE is located on chromosome 6p21.3 at the class II/III junction of the major histocompatibility complex and has a 1.7 kb 5' flanking region, 11 exons, 10 introns, and a 3' untranslated region[5,6]. RAGE products are raised in a variety of clinical conditions, including several cancers. At each stage of the lesion, RAGE is expressed in the tissue in a distinct manner. The RAGE, commonly referred to as a "pattern recognition receptor," is a member of the immunoglobulin superfamily of cell surface molecules with a diverse array of ligand specificities[7].

By interacting with its varied ligand families, RAGE orchestrates several intracellular signaling pathways to govern numerous cellular functions, including inflammation, apoptosis, proliferation, and autophagy. In animal models, it has also been demonstrated that inhibiting RAGE signaling inhibits cancer growth and metastasis through regulating tumor proliferation, invasion, and matrix metalloproteinase expression[8]. The purpose of this literature review is to establish the relationship between RAGEs and OSCC by clarifying the importance of RAGE as a diagnostic and predictive biomarker for oral malignancies as well as a potential treatment strategy.

2. Materials and methods

We conducted a comprehensive search in PubMed, Google, and Cochrane for publications published up to December 2017 that explored the connection between RAGE and oral cancer. Several crucial terminologies were used separately and in conjunction: RAGE, receptor for advanced glycation end products, or AGE, oral cancer, carcinoma, oral neoplasm, and squamous cell carcinoma.

Exclusion criteria included non-English publications, conference abstracts, and studies that did not involve human subjects or samples, reviews, and articles relating to other head and neck tumors, and studies assessing the effect of drug therapy. Articles on oral cancer, cross-sectional research, and RAGE values are included as inclusion criteria. Following a review of article titles and abstracts, seven full-text studies were retrieved for inclusion in the study.

Due to the heterogeneity of the analyzed studies, a meta-analysis was not possible. The collected papers were subjected to a systematic review, and the results were tabulated and analyzed in Table 1.

3. Results

About 28% of the articles utilizing immunohistochemistry (IHC) demonstrated that RAGE products are more prevalent in the invasive front of tumor tissues, and 43% of the articles demonstrated that high expression of these receptors increased the motility of tumor cells. Due to their compatibility with several ligands, these receptors are crucial for tumor cell angiogenesis, invasiveness, and metastasis, making RAGEs an essential biomarker for determining the prognosis of oral carcinomas.

4. Discussion

RAGE is a cell surface receptor belonging to the immunoglobulin superfamily. They display molecule types produced by the non-enzymatic glycation of proteins through Millard's reaction. RAGE is expressed at low levels in normal tissue and vasculature during development. It appears uncontrolled wherever its ligands congregate[9] and is elevated in numerous clinical and non-pathological situations. The pro-inflammatory RAGE ligand high motility group box 1 (HMGB-1) is frequently released by necrotic cells. HMGB-1 and RAGE have also been reported to interact in chronic inflammation and the immune system. The average RAGE concentration in cancer samples was 57 ± 1.9 pg/mL. In a study, Sasahira *et al.* hypothesized that the concentration of RAGE declines with the progression of cancer; this is due to the consumption of such molecules during the process of cancer initiation, and the emergence of numerous other components in the later stages of cancer[10]. Consequently, the concentration of these receptors may be depleted during the later stages of cancer[11].

In seven studies encompassing 820 cases and 626 controls, the overall RAGE levels were shown to be higher in well-differentiated OSCC and lower as differentiation decreased in this systematic review.

Due to its potential to alter the HMGB-1, it is evident that RAGE has a strong relationship with cell motility; this has been demonstrated in 43% of the publications, hence boosting the invasiveness of the tumor. Using the Boyden Chamber experiment, Choi *et al.* demonstrated that RAGE, in conjunction with HMGB-1, considerably increases the motility of the cells, while Ujjal *et al.* demonstrated that antisense RAGE inhibits this motility. In their study, Shun *et al.* demonstrated that RAGE antibodies inhibit cell motility, underscoring the importance of these receptors in the enhanced cell motility observed in cases of OSCC. The antibody against RAGE may inhibit the motility and migration of cells, thereby facilitating targeted therapy.

Table 1. Details of the studies selected for the review

Study	Number of samples	Cell line	Case	Control	Marker	Method	Observation	*P*-value	Inference	Limitations
Choi *et al.*[10]	10		10	-	RAGE HMGB-1	Reverse transcriptase PCR	RAGE mRNA expression (534 bp) was detected		The mobility of cells was found to be increased when both RAGE and HMGB-1 is present.	Various tissue sources were utilized, and the number of cases was reduced.
						Boyden Chamber Assay	At 2000 ng/ml concentration of HMGB-1 the cell migration was increased.			
Bhawal *et al.*[21]	30	10	10-Primary 10-Metastatic	_	RAGE	Reverse transcriptase PCR	RAGE mRNA was identified at various levels in all oral carcinoma cell lines and very weakly in normal mucosa		Metastatic cases expressed RAGE mRNA at a high level compared to primary tumours	Not mentioned.
						Antisense Phosphorothioate (S)-Oligodeoxynucleotide Assay	RAGE antisense (S) oligodeoxynucleotide-treated LMF4 cells migrating was significantly lower (37 8 8.88 cells per well.	$P = 0.0017$		
						Western Blot	After treatment with antisense S-oligodeoxynucleotide, RAGE protein expression was significantly reduced in metastatic cells.			
Shun-Yao Ko[15]					AGE RAGE	Trypan Blue and WST-1	AGEs or BSA were applied to SAS cells. AGEs lowered the number of cells considerably.	$P = 0.01$	AGEs enhance cell migration and also reduce cell Viability. RAGE antibodies suppress cell migration.	Not mentioned.
						Western Blot	Treatment with AGEs significantly increased ERK phosphorylation.	$P = 0.009$		
Su *et al.*[19]	1210		618	592	RAGE pleomorphism	Real time PCR	When combined effect of Heterozygous genotype (TC) for rs1800625 with the homozygotes (CC) for the variant allele of rs1800625, its association with the risk of oral cancer is increased.	-	SNP rs1800652 was found to be associated with risk and progression of oral cancers	Note on the environmental risks not given.

(Cont'd...)

Table 1. *(Continued)*

Study	Number of samples	Cell line	Case	Control	Marker	Method	Observation	*P*-value	Inference	Limitations
Sasahira *et al.*[16]	74	_	74	_	RAGE	IHC	The disease-free survival rate of RAGE-H patients was considerably lower than that of individuals with a low grade.	$P < 0.00001$	RAGE-H suggests low disease-free survival rate.	Not mentioned
Landesberg *et al.*[22]	50	_	38	14	RAGE	IHC	10 of 10 (100%) well-differentiated OSCCs were RAGE-positive, while 0 of 8 (0%) were poorly differentiated.	$P < 0.05$	The RAGE expression decreases in OSCC as they become less differentiated	Tumor thickness cannot be determined on biopsy derived material
						Western Blot	Reduced band intensity at the molecular weight 55 kDa, seen more in OSCC than normal control.			
Sasahira *et al.*[23]	40		20	20	RAGE VEGF	Elisa	HSC-3 and HSC-4 cells treated with RAGE antisense S-ODN secreted considerably less VEGF	$P = 0.0123$ and 0.0344, respectively	RAGE plays a function in the expression of VEGF in tumors, which is linked to tumor angiogenesis	No correlations between RAGE and histological differentiation

RAGE: Receptors of advanced glycation end products, VEGF: Vascular endothelial growth factor, AGE: Advanced glycation end products, IHC: Immunohistochemistry, PCR: Polymerase chain reaction

IHC is used to study the distribution of these receptors in tumor tissue sections, which revealed that 28% of the articles studied the expression of RAGE in tumor tissues and the surrounding normal mucosa using IHC and concluded that RAGE positivity is more prevalent at the invasive front and is expressed more in well-differentiated than poorly differentiated OSCCs. This demonstrates that the presence and activity of these receptors and their ligands influence the invasiveness of tumor tissues.

Vascular endothelial growth factor (VEGF) is the primary facilitator of tumor angiogenesis, encouraging the creation of new blood vessels from adjacent capillaries, and granting tumors access to the oxygen and nutrients they require to flourish[12]. VEGF also plays a crucial role in tumor development by protecting tumor neovasculature against apoptosis through the activation of anti-apoptotic proteins Bcl-2 and survivin[13]. The elevated amounts of VEGF in tumors also result in architecturally distinct blood vessels compared to normal blood vessels. In contrast to the architecture of normal vasculature, tumor vasculature is irregularly formed, dilated, convoluted, and characterized by a large number of blind ends. Due to the chaotic design of tumor vasculature, tumor blood flow is often poor, with areas of stasis caused by dead-end arteries and disorganized blood flow caused by aberrant vascular connections[13]. This, in turn, predisposes regions to hypoxia, which further stimulates VEGF release and generates disorganized vasculature. *In situ* study of tumor tissues undergoing neovascularization indicates the close proximity of clusters of capillaries and VEGF-producing cells to necrosis[14].

In one of their articles, Sasahira *et al.* evaluated the relationship between RAGE and VEGF, thereby anticipating its role in angiogenesis[15,16]. Long recognized as the most potent angiogenic factor, VEGF-A is frequently related with increased MDV and unfavorable clinicopathological characteristics and results[17]. The expression of VEGF and VEGF-C was substantially correlated with MVD and LVD, suggesting that RAGE also plays a role in angiogenesis. There was no association between VEGF and MVD numbers, either individually or in terms of grade[18].

Higher levels of AGE receptors were detected near the invasive front of tumors, suggesting that these molecules play a role in the invasiveness and spread of cancers. These results imply that these receptors and their associated ligands play a significant role in the growth and spread of the tumor, hence affecting the prognosis and life expectancy of the affected individual.

Su *et al.* examined the genetic predisposition for oral squamous cell carcinoma; this investigation revealed that the presence of at least one polymorphic allele of rs1800625 increases the risk of OSCC[19]. Allelic discriminations of the RAGE rs1800625, rs1800624, rs2070600, and rs184003 allelic polymorphisms were investigated, and it was observed that the RAGE gene polymorphism rs1800625 not only raised the risk of oral cancer but was also related with late-stage and large-size tumors. The study also sheds light on environmental factors that may alter RAGE expression and oral cancer susceptibility. Even in the absence of tobacco chewing or smoking, polymorphism of the rs184003 allele was found to predispose an individual to a higher risk of oral squamous cell cancer. RAGE-ligand antagonists may be able to effectively target the "Achilles' heel" of cancers, namely, poor prognosis, silent metastasis, drug resistance, and cancer recurrence, given that inhibition of RAGE, HMGB-1, or S100 proteins alone has demonstrated a significant reduction in tumor size, invasion, and angiogenesis in a number of cancers[20].

In conclusion, there is promise for the creation of targeted therapeutics using RAGE products. Its potential significance in angiogenesis and tumor progression (cell invasion and motility) makes it an attractive adjuvant therapy target. *In vitro* inhibition of RAGE signaling with RAGE antibodies decreased cell differentiation and migration, thereby establishing RAGE as a unique and specific target for the treatment and management of OSCCs.

5. Conclusion

Despite insufficient understanding of the mechanism of interaction between various ligands and families, the available evidence supports the function of RAGE in invasion, migration, and angiogenesis in oral cancers. There has been no investigation on the diagnostic use of these markers for oral malignancies. Multiple clinicopathological factors are associated with these oral cancer indicators. This evidence is promising for the therapeutic application of these compounds in prognostic considerations and molecular target recognition treatment. To design a more targeted or possibly individualized treatment for patients, further study must be undertaken to acquire a better knowledge of the molecular events.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Palati Sinduja

Formal analysis: Palati Sinduja, Pratibha Ramani

Writing – original draft: Palati Sinduja

Writing – review & editing: Pratibha Ramani, Saravanan Sekaran

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Scully C, Bagan J, 2009, Oral squamous cell carcinoma overview. *Oral Oncol*, 45(4–5): 301–308.

   https://doi.org/10.1016/j.oraloncology.2009.01.004

2. López-Graniel CM, de León DT, Meneses-García A, *et al.*, 2001, Tumor angiogenesis as a prognostic factor in oral cavity carcinomas. *J Exp Clin Cancer Res*, 20(4): 463–468.

3. Hofmann MA, Drury S, Fu C, *et al.*, 1999, RAGE mediates a novel proinflammatory axis: A central cell surface receptor for S100/calgranulin polypeptides. *Cell*, 97(7): 889–901.

   https://doi.org/10.1016/s0092-8674(00)80801-6

4. Hori O, Brett J, Slattery T, *et al.*, 1995, The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin mediation of neurite outgrowth and co-expression of rage and amphoterin in the developing nervous system. *J Biol Chem*, 270(43): 25752–25761.

   https://doi.org/10.1074/jbc.270.43.25752

5. Sugaya K, Fukagawa T, Matsumoto K, *et al.*, 1994, Three genes in the human MHC class III region near the junction with the class II: Gene for receptor of advanced glycosylation end products, PBX2 homeobox gene and a notch homolog, human counterpart of mouse mammary tumor gene int-3. *Genomics*, 23(2): 408–419.

   https://doi.org/10.1006/geno.1994.1517

6. Vissing H, Aagaard L, Tommerup N, *et al.*, 1994, Localization of the human gene for advanced glycosylation end product-specific receptor (AGER) to chromosome 6p21.3. *Genomics*, 24(3): 606–608.

   https://doi.org/10.1006/geno.1994.1676

7. Bierhaus A, Humpert PM, Morcos M, *et al.*, 2005, Understanding RAGE, the receptor for advanced glycation end products. *J Mol Med (Berl)*, 83(11): 876–886.

   https://doi.org/10.1007/s00109-005-0688-7

8. Taguchi A, Blood DC, del Toro G, *et al.*, 2000, Blockade of RAGE-amphoterin signalling suppresses tumour growth and metastases. *Nature*, 405(6784): 354–360.

   https://doi.org/10.1038/35012626

9. Schmidt AM, Stern DM, 2001, Receptor for age (RAGE) is a gene within the major histocompatibility class III region: Implications for host response mechanisms in homeostasis and chronic disease. *Front Biosci J Virtual Libr*, 6(3): 1151–1160.

   https://doi.org/10.2741/schmidt

10. Choi J, Lee MK, Oh KH, *et al.*, 2011, Interaction effect between the receptor for advanced glycation end products (RAGE) and high-mobility group box-1 (HMGB-1) for the migration of a squamous cell carcinoma cell line. *Tumori J*, 97(2): 196–202.

    https://doi.org/10.1177/030089161109700211

11. Kierdorf K, Fritz G, 2013, RAGE regulation and signaling in inflammation and beyond. *J Leukoc Biol*, 94(1): 55–68.

    https://doi.org/10.1189/jlb.1012519

12. Bergers G, Benjamin LE, 2003, Tumorigenesis and the angiogenic switch. *Nat Rev Cancer*, 3(6): 401–410.

    https://doi.org/10.1038/nrc1093

13. Carmeliet P, 2005, VEGF as a key mediator of angiogenesis in cancer. *Oncology*, 69 Suppl 3: 4–10.

    https://doi.org/10.1159/000088478

14. Shweiki D, Itin A, Soffer D, *et al.*, 1992, Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature*, 359(6398): 843–845.

    https://doi.org/10.1038/359843a0

15. Ko SY, Ko HA, Shieh TM, *et al.*, 2014, Cell migration is regulated by AGE-RAGE interaction in human oral cancer cells *in vitro*. *PLoS One*, 9(10): e110542.

    https://doi.org/10.1371/journal.pone.0110542

16. Sasahira T, Kirita T, Bhawal UK, *et al.*, 2007, Receptor for advanced glycation end products (RAGE) is important in the prediction of recurrence in human oral squamous cell carcinoma. *Histopathology*, 51(2): 166–172.

    https://doi.org/10.1111/j.1365-2559.2007.02739.x

17. Hutajulu SH, Paramita DK, Santoso J, *et al.*, 2018, Correlation between vascular endothelial growth factor-A expression and tumor location and invasion in patients with colorectal cancer. *J Gastrointest Oncol*, 9(6): 1099–1108.

    https://doi.org/10.21037/jgo.2018.07.01

18. Kukreja I, Kapoor P, Deshmukh R, *et al.*, 2013, VEGF and CD 34: A correlation between tumor angiogenesis and microvessel density-an immunohistochemical study. *J Oral Maxillofac Pathol*, 17(3): 367–373.

https://doi.org/10.4103/0973-029X.125200

19. Su S, Chien M, Lin C, *et al.*, 2015, RAGE gene polymorphism and environmental factor in the risk of oral cancer. *J Dent Res*, 94(3): 403–411.

    https://doi.org/10.1177/0022034514566215

20. Palanissami G, Paul SF, 2018, RAGE and its ligands: Molecular interplay between glycation, inflammation, and hallmarks of cancer-a review. *Horm Cancer*, 9(5): 295–325.

    https://doi.org/10.1007/s12672-018-0342-9

21. Bhawal UK, Ozaki Y, Nishimura M, *et al.*, 2005, Association of expression of receptor for advanced glycation end products and invasive activity of oral squamous cell carcinoma. *Oncology*, 69(3): 246–255.

    https://doi.org/10.1159/000087910

22. Landesberg R, Woo V, Huang L, *et al.*, 2008, The expression of the receptor for glycation endproducts (RAGE) in oral squamous cell carcinomas. *Oral Surg Oral Med Oral Pathol Oral Radiol Endodontol*, 105(5): 617–624.

    https://doi.org/10.1016/j.tripleo.2007.08.006

23. Sasahira T, Kirita T, Bhawal UK, *et al.*, 2007, The expression of receptor for advanced glycation end products is associated with angiogenesis in human oral squamous cell carcinoma. *Virchows Arch*, 450(3): 287–295.

    https://doi.org/10.1007/s00428-006-0359-2

ACCSCIENCE PUBLISHING

Tumor Discovery

REVIEW ARTICLE

An overview of the risk, underlying factors, and mechanism of cancer progression in polycystic ovary syndrome

Rubiat Afrin Ayon[1], Md. Azmain Faike[1], Sumiya Zaman Ononna[1], Aminul Hassan[1], Shoumik Kundu[1], Farhana Akhter[2], Mohammad Mahfuz Ali Khan Shawan[1], Md. Mozammel Hossain[1], Md. Ibrahim Khalil[1]*, and Md. Ashraful Hasan[1]*

[1]Laboratory of Preventive and Integrative Biomedicine, Department of Biochemistry and Molecular Biology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

[2]Government Unani and Ayurvedic Medical College Hospital, Mirpur-13, Dhaka 1221, Bangladesh

Abstract

Polycystic ovary syndrome (PCOS) is the most common endocrine and metabolic disorder among women of reproductive age. PCOS is characterized by ovulatory dysfunction, clinical or biochemical features of hyperandrogenism, and polycystic ovaries. The risk of cancer among PCOS patients has been a topic of discussion for decades due to the overlapping metabolic and endocrine abnormalities. This review article focuses on the association of PCOS with various types of reproductive (such as endometrial cancer, ovarian cancer, and breast cancer) and non-reproductive cancers, considering different aspects, such as the risk of cancer progression in PCOS patients, the underlying factors, and the mechanism through which PCOS might progress to cancer. The information provided in this article would help create awareness among PCOS patients about the need to take risk-reducing measures. This article might also aid in the effort of identifying novel therapeutic targets to counteract the progression of cancer in PCOS.

Keywords: Polycystic ovary syndrome; Cancer; Risk; Mechanism

***Corresponding authors:**
Md. Ibrahim Khalil
(drmikhalil@gmail.com)
Md. Ashraful Hasan
(ashrafulhasan@juniv.edu)

Citation: Ayon RA, Faike MA, Ononna SZ, *et al.*, 2023, An overview of the risk, underlying factors, and mechanism of cancer progression in polycystic ovary syndrome. *Tumor Discov,* 2(1): 328. https://doi.org/10.36922/td.328

Received: January 16, 2023
Accepted: March 3, 2023
Published Online: March 17, 2023

1. Introduction

Polycystic ovary syndrome (PCOS) is the most prevalent endocrine disorder among females of reproductive age in developed countries. According to reports, 6–20% of women of reproductive age suffer from PCOS[1]. This syndrome, which is heterogeneous, may be defined by a combination of signs and symptoms related to androgen excess and ovarian dysfunction in the absence of any specific diagnoses[1]. Women who are diagnosed with PCOS may complain of heavy or irregular menstrual bleeding, infertility, obesity, oily skin, seborrhea, cystic acne, or hirsutism. These symptoms may have a significant negative effect on a woman's quality of life and may cause psychological anguish that jeopardizes both her femininity and physical health. As a result, the illness may lead to issues at work, dysfunctional family dynamics, and changing perceptions of oneself[2].

The pathophysiology of PCOS is known to be affected by both genetic and environmental factors, but its exact cause is unclear[3]. Despite the fact that the molecular mechanism behind PCOS pathogenesis is still largely unknown, there is a vast amount of evidence indicating that hyperandrogenism is crucial to the progression and consequences of PCOS[4]. Whether PCOS is a single clinical entity or a conglomerate of many diseases with a common clinical appearance is unclear as of yet. Researchers have reported heightened risks of insulin resistance, diabetic mellitus, cardiovascular disease, metabolic syndrome, endometrial dysfunction, and pregnancy complications in PCOS patients[5-7].

The risk of cancer among PCOS-affected women has been a topic of debate over the years[8]. Given the high incidence of PCOS, any link to cancer would be crucial from the standpoint of public health. Identifying the component that significantly increases the risk of developing cancer is extremely challenging due to the multifactorial nature of the syndrome, along with its heterogeneous presentation[9]. In more developed countries, endometrial cancer is one of the most common reproductive cancers, but it is ranked as the second most common in less developed countries[10]. Stein and Leventhal initially described PCOS in 1935, but the first study of PCOS and the risk of endometrial cancer was only published after 14 years later[11,12]. Although the direct association of PCOS with breast cancer and ovarian cancer has yet to be established, the incidence of these cancers in PCOS patients has been observed over the years in several studies[13-15]. Androgens are believed to be involved in the pathogenesis of ovarian cancer in PCOS patients, while the level of estrogen is related to the progression of endometrial cancer[16-18]. One of the key drugs used in PCOS management is oral contraceptives. Studies have suggested that oral contraceptives interfere with cancer-associated regulations and reduce the risk of developing cancer in PCOS women[19,20]. The association of breast cancer with PCOS is unclear, but incessant research efforts have been made to find a strong link between them. Although the effects of androgen on breast cancer development in PCOS have not been fully understood, studies have shown evidence of the role of estrogen in breast cancer development[21-23]. Several studies are also underway to identify the genes responsible for cancer pathogenesis in PCOS[24,25].

We reviewed the current state of knowledge in a comprehensive manner, specifically the overall potential of hazardous cancers that could occur in patients with PCOS, along with the underlying factors and mechanism. The information provided in this article would help create awareness among the younger generation, thus ameliorating these problems. Early detection and proper treatment can lessen the burden of clinical symptoms and concomitant psychological anguish, thus reducing the risk of developing cancer and the effect on PCOS patients' quality of life in terms of their health. This article might also aid in the efforts of identifying novel therapeutic targets to counteract shared dysregulation in PCOS and cancer.

2. Pathophysiology of PCOS

PCOS is a multifaceted syndrome and its complex pathophysiology is yet to be fully understood. The two most notable phenotypes of PCOS are hyperandrogenism and ovarian dysfunction. Increased blood levels of free (unbound) testosterone, a crucial hormone involved in the pathogenesis of PCOS, are indicative of hyperandrogenism[26-28]. Figure 1 illustrates the pathophysiology of PCOS.

Androgens, including dehydroepiandrosterone, dehydroepiandrosterone sulfate, testosterone, dihydrotestosterone, and androstenedione, are found in serum in decreasing order of concentration[29,30]. The hypothalamic-pituitary-ovarian axis is thought to be imbalanced in PCOS as a result of neuroendocrine dysregulation. This leads to increased frequency of gonadotropin-releasing hormone pulses. The increase in frequency of GnRH pulses promotes luteinizing hormone (LH) rather than follicle-stimulating hormone (FSH) production, leading to an increase in LH: FSH ratio in PCOS, which causes hyperandrogenism[31,32]. The anti-Müllerian hormone (AMH) produced by ovarian granulosa cells is a regulatory factor of GnRH release. In PCOS patients, high AMH stimulates LH production through AMH receptor on the hypothalamus and pituitary, increasing the secretion of androgen by ovarian theca cells. At the same time, high AMH suppresses FSH receptor and aromatase production in granulosa cells; a low level of FSH prevents testosterone from being converted to estrogen, thus resulting in androgen excess. Elevated testosterone, in turn, promotes the direct and indirect release of AMH from granulosa cells[33,34].

In PCOS, there are disruptions in the interactions and coordination between LH, FSH, insulin-like growth factor 1 (IGF-1), AMH, androgen conversion enzymes, and additional variables, resulting in arrested ovarian follicular development. Ovarian hyperandrogenism, hyperinsulinemia from insulin resistance, and intra-ovarian paracrine signaling are all factors in PCOS that disrupt follicle growth. The reduced FSH level inhibits ovarian follicular development, which may cause amenorrhea, anovulation, and polycystic morphology[33,35-36].

Although insulin receptor gene alterations are uncommon, hyperinsulinemia, and insulin resistance are two prominent clinical conditions of PCOS in women. Insulin resistance, a disordered physiological state caused by impaired glucose transport and utilization, is a result of the biological effects of insulin being reduced when it is

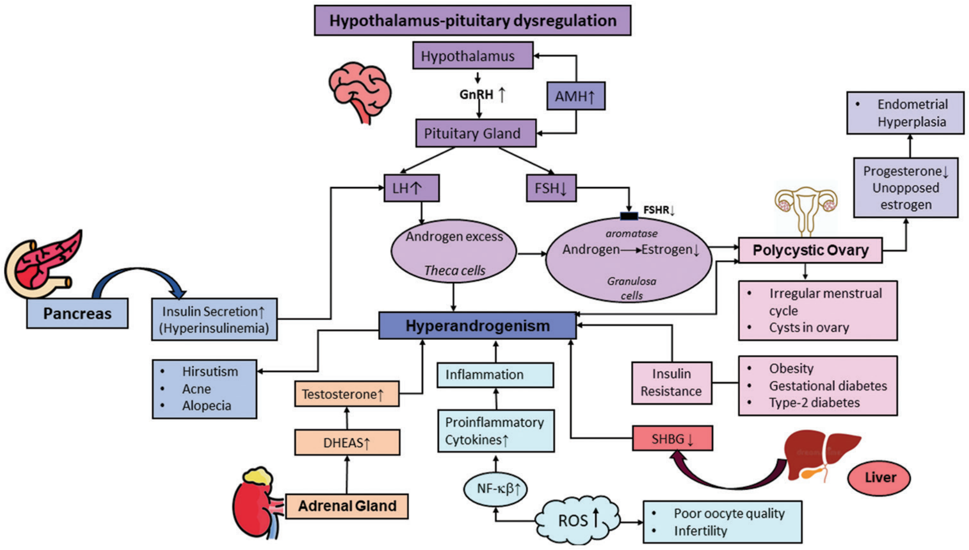


Figure 1. Pathophysiology of PCOS. Hyperandrogenism and ovarian dysfunction are the main phenotypes of PCOS. Neuroendocrine dysregulation disrupts the balance of the hypothalamic-pituitary-ovarian axis, resulting in an increase in frequency of GnRH release. The increased GnRH pulse frequency causes an elevated LH: FSH ratio, leading to hyperandrogenism. Several factors such as elevated AMH level, insulin resistance, hyperinsulinemia, low SHBG level, and increased ROS generation contribute to hyperandrogenism in PCOS. Clinical features of hyperandrogenism include hirsutism, acne, and androgenic alopecia, while decreased FSH level causes amenorrhea, anovulation, and polycystic morphology. Low progesterone, along with unopposed estrogen, leads to endometrial hyperplasia.

AMH: Anti-Müllerian hormone; DHEAS: Dehydroepiandrosterone sulfate; FSHR: Follicle-stimulating hormone receptor; GnRH: Gonadotropin-releasing hormone; LH: Luteinizing hormone; NF-κB: Nuclear factor kappa B; PCOS: Polycystic ovary syndrome; ROS: Reactive oxygen species; SHBG: Sex hormone-binding globulin.

present at high amounts[35,37]. Insulin resistance is higher in peripheral tissues, particularly in skeletal muscles. Burghen *et al*., who noted that hyperinsulinemia is related to insulin resistance, made the initial argument that insulin plays a role in ovarian function in women with hyperandrogenemia[38,39]. In particular, it has been shown that insulin secreted from pancreatic beta cells acts directly through its own receptor at physiological concentrations on cultivated polycystic ovary theca cells and stimulates androgen production, which is noticeably higher than in ovarian theca cells from normal women. Importantly, insulin can work in synergy with LH to boost androgen biosynthesis, while increasing androstenedione production on its own[38,40,41].

Hirsutism, acne, and androgenic alopecia are clinical signs and symptoms of hyperandrogenism caused by hyperinsulinemia[42]. The majority of women with PCOS, in fact, have insulin resistance and compensatory hyperinsulinemia, which is partially attributed to an innate insulin resistance mechanism, especially in those who are overweight or obese, or have diabetes[8,43,44]. In PCOS women, the hypersecretion of adrenocorticotropic hormone causes the overproduction of androgens from the adrenal glands[45].

Hyperandrogenism is mainly manifested by free or unbound testosterone in the blood. Only 1–2% of the testosterone in the blood is unbound, while the remaining 98% is predominantly bound to sex hormone-binding globulin (SHBG). Growing levels of testosterone are considered a characteristic of puberty in teenage females. PCOS may develop if this condition worsens and there is an overproduction of testosterone[46,47]. In addition to directly increasing ovarian androgen production, hyperinsulinemia also increases the fraction of free testosterone in PCOS by lowering the synthesis of liver SHBG[48]. Hence, low serum SHBG level in PCOS patients leads to hyperandrogenism.

The production of a large amount of ROS can be considered a contributing factor in the pathophysiology of PCOS. The imbalance between free radicals and antioxidants in the body occurs from the overproduction of ROS, leading to oxidative stress[45]. Oxidative stress is more frequent in obese PCOS patients who develop early insulin resistance[49]. The vast amount of ROS causes increased production of pro-inflammatory cytokines[50]. The pro-inflammatory cytokines that are produced inside the endometrium can easily hinder the mechanism of action of insulin in PCOS, resulting in insulin resistance and eventually causing hyperandrogenism[51,52].

The detailed mechanism of the entire process of oxidative stress causing inflammation and eventually leading to hyperandrogenism is still under investigation, with several pathways and factors considered to be potential areas of study[49,53]. Among the factors, nuclear factor-κB (NF-κB) is a potential effector of inflammation induced by hyperandrogenism. The expression of NF-κB is enhanced through increased level of phosphorylation that induces ROS production and inflammatory responses[36,49,54].

3. Polycystic ovary syndrome and cancer

PCOS is an intrinsic endocrine and metabolic condition that seriously affects menstrual and reproductive functions, causing detrimental effects on a woman's health throughout her lifetime. Although PCOS has been linked to hyperandrogenism, genetic, and epigenetic factors, the exact cause of PCOS remains unknown[55,56]. Between 6% and 20% of premenopausal women have PCOS, possibly making it the most prevalent endocrine and metabolic condition among women of reproductive age[57-59]. The risk of cancer among PCOS-affected women has been a debatable topic over the years[8]. There remains, however, absence of adequate conclusive research linking PCOS to the risk of various cancers due to uncertainties in its etiopathology, dubious diagnostic standards, and the complex endocrine and metabolic dysfunctions[60]. PCOS is conjectured to be linked to various types of reproductive cancers, including endometrial cancer, ovarian cancer, breast cancer, uterine cancer, and others. One of the primary clinical signs of PCOS, chronic oligo- or anovulation, induced by continuous estrogen exposure without counteracting progesterone, was linked to endometrial cancer in initial reports of the relationship between PCOS and cancer[61,62].

3.1. Risk and underlying factors of developing reproductive cancer in polycystic ovary syndrome

3.1.1. Endometrial cancer

Endometrial cancer predominantly affects 2–3% of postmenopausal women and is the most prevalent female genital tract malignancy worldwide[63,64]. According to studies, women with PCOS may have a higher chance of developing endometrial cancer than those without the condition[65]. Women with PCOS have multiple endometrial cancer risk factors and their risk for developing endometrial cancer may be higher[66]. Prolonged and unchallenged exposure of estrogen to the endometrium of PCOS patients, low progesterone, obesity, hyperinsulinemia, insulin resistance, IGF, diabetes, nulliparity, cyclin D1, and glutathione S-transferase constitute the clinical, metabolic, and molecular risk factors for endometrial cancer development[66-68].

Six studies have looked into the link between PCOS and endometrial cancer without taking body mass index (BMI) into consideration (Table 1). In a registry-based cohort study of 12,070 women with PCOS, Gottschau *et al.* observed that these patients had 4 times increased risk of endometrial cancer, with a standardized incidence ratio (SIR) of 3.9 (95% confidence index [CI] = 2.2 – 6.3)[69]. Nearly one-third of these women were between the ages of 15 and 24 at the time of their first admission or visit for PCOS, and about 50% were between the ages of 25 and 34[69]. In another population-based, retrospective, and cohort study, a higher mean adjusted hazard ratio (HR) of uterine cancer was found in PCOS patients[70]. A report by Haoula *et al.* concluded that endometrial cancer is 3 times more likely to develop in patients with PCOS in comparison with women without it[71]. A meta-analysis has reported increased risk of endometrial cancer in PCOS patients, with an odds ratio (OR) of 2.79 at CI = 1.31 – 5.95 (P < 0.008)[61]. The risk estimate increased when women aged <54 were excluded from the study[61]. A case–control study by Fearnley *et al.* showed a link between PCOS and an increased risk of endometrial cancer in women under the age of 50[72]. Pillay *et al.* also observed a higher prevalence of PCOS in endometrial cancer patients aged below 50[67].

3.1.2. Ovarian cancer

Ovarian cancer is the seventh most common type of neoplasm in women worldwide[73]. Around 239,000 new cases and 152,000 fatalities worldwide are reported each year, with Eastern and Central Europe having the highest incidence[60]. It has been proposed that PCOS increases the risk of ovarian cancer by increasing androgen exposure[74].

Cancer risk and infertility have been linked, albeit the debatable association with ovarian cancer. The most common reasons of infertility in female are hormonal conditions that interfere with ovulation, including PCOS. Ovarian cancer has been linked to a number of endocrine and reproductive changes. As a result, nulliparity, early menarche, and a later onset of menopause have all been linked to an increased risk of ovarian cancer[75-78]. Four studies have explicitly looked at the relationship between PCOS and ovarian cancer[61,69,70,79]. However, most these studies have presented mainly negative results although one recent meta-analysis of three studies did imply an elevated risk (OR = 1.4; 95% CI = 0.9 – 2.2)[61]. According to a Danish registry-based study, women with PCOS had a non-significantly higher risk of developing ovarian cancer than the general Danish female population (SIR = 1.8; 95% CI = 0.8 – 3.2); however, the study was constrained by its small sample size (10 ovarian cases)[69]. No correlation was found between PCOS and ovarian cancer in a recent Taiwanese investigation with only 11 occurrences of the disease (HR = 1.0; 95% CI = 0.2 – 4.6)[70]. Following a retrospective cohort study of women with PCOS, Brinton *et al.* found that patients with secondary infertility had a higher risk of developing ovarian cancer than those with

Table 1. Prevalence of reproductive cancer in polycystic ovary syndrome patients

Type of study	Location	Total participants	Cancer type (reproductive)	Number of cancer patients	Adjusted HR/RR/SIR/OR (95% CI value)	References
Cohort	Denmark	12,070	Endometrium	16	3.9 (2.2 – 6.3)	[69]
			Ovarian	10	1.8 (0.8 – 3.2)	
			Breast	59	1.1 (0.8 – 1.4)	
			Cervix uteri	14	0.8 (0.5 – 1.4)	
Registry-based cohort	Taiwan	3,566	Ovarian	2	0.997 (0.214 – 4.636)	[70]
			Breast	14	1.976 (1.035 – 3.722)	
			Uterine	15	8.420 (1.615 – 43.888)	
Meta-analysis	USA Japan Australia Europe	919	Endometrium	5	2.79 (1.31 – 5.95)	[61]
			Ovarian	3	1.41 (0.93 – 2.15)	
			Breast	3	0.95 (0.64 – 1.39)	
Meta-analysis	United Kingdom	938 (cancer patients)	Endometrium	938	2.9 (1.5 – 5.5)	[71]
Retrospective cohort	USA	2,560	Uterine corpus	5	1.28 (0.49 – 3.35)	[79]
			Ovarian	2	0.42 (0.10 – 1.75)	
			Breast	23	0.81 (0.53 – 1.25)	
Case-control	Australia	1,399	Endometrium	156	3.64 (1.76 – 7.52)	[72]
Cross-sectional	United Kingdom	128	Endometrium	11	1.0 (0.4 – 2.7)	[67]
Cohort	USA	116,671	Breast	2267	0.8 (0.6 – 1.0)	[83]

HR: Hazard ratio; OR: Odds ratio; RR: Relative risk; SIR: Standardized incidence ratio

primary infertility (RR = 0.42; 95% CI = 0.10 – 1.75)[79]. In-depth research is required to ascertain if PCOS is related to ovarian cancer.

3.1.3. Breast cancer

Data from several organizations have shown that breast cancer is the most common cancer in women[80]. According to statistics from 2018, 627,000 women died from breast cancer, accounting for 15% of all cancer fatalities in women. Although breast cancer rates are greater in more developed areas, they are rising practically in every area of the world[61]. There is a theory that the dysregulated genes observed in PCOS patients and the genes linked to breast cancer overlap. Three such potential genes, discovered by Xu *et al.*, include alpha polypeptide, platelet-derived growth factor receptor, and hydroxysteroid (17-beta) dehydrogenase[81].

However, there have been studies done that did not find any connection between PCOS patients and a higher risk of breast cancer[82]. With 59 breast cancer cases observed and 56 expected, a recent Danish registry-based study has found no association between PCOS and breast cancer risk, with an SIR of 1.1 (95% CI = 0.8 – 1.4)[69]. Similarly, a retrospective and cohort study in Taiwan has also found no association between PCOS and breast cancer risk (HR = 1.6, 95% CI = 0.9 – 2.8)[70]. According to a meta-analysis, no increased risk was seen in women with PCOS (OR = 1.0; 95% CI = 0.6 – 1.4)[61]. When compared to women who had not been treated for infertility, women with ovulatory problems had a considerably lower risk of breast cancer (RR = 0.8; 95% CI = 0.6 – 1.0)[83]. In a study conducted by Brinton *et al.*, the patients with PCOS did not appear to have an increased risk of breast cancer (RR = 0.81; 95% CI = 0.53 – 1.25)[79].

3.2. Risk and underlying factors of developing non-reproductive cancer in polycystic ovary syndrome

PCOS patients have direct and indirect risks of developing cancers in their later life. If left untreated, the risk of developing atypical endometrial hyperplasia and carcinoma is high due to irregular menstruation with protracted exposure to unobstructed estrogen[84]. These risks are significantly higher among women with obesity or who are overweight. Only a few studies on the association of PCOS with non-reproductive cancers are available (Table 2). A registry-based cohort study by Gottschau *et al.* reported elevated risk of colon, kidney, and brain cancers, with SIR of 2.1, 3.9, and 2.2, respectively (95% CI = 1.1 – 3.8, 1.4 – 8.4, and 1.3 – 3.5, respectively), among PCOS patients but no significant risk of other types of cancers, such as lung cancer, melanoma, and other types of skin cancers[69].

4. Mechanism of cancer progression in polycystic ovary syndrome

Many reports exploring PCOS and the risk of cancer have been published. Hyperinsulinemia, high estrogen level, and chronic inflammation are some of the factors associated with normal PCOS progression[35,37], and there is an evidence that these chronic conditions may be contributing factors to oncogenesis and cancer progression in PCOS[85,86]. Hyperinsulinemia directly initiates and regulates the initiation of cancer development, while inflammation affects various pro-tumorigenic pathways that ultimately lead to angiogenesis and carcinogenesis, thus promoting cancer cell development in certain sites[87,88]. Recent evidence has suggested that the effect of sympathetic hyperactivity is also a risk factor for cancer progression in PCOS women. Sympathetic hyperactivity leads to the secretion of norepinephrine, in which increased norepinephrine acts as a biochemical switch for tumor angiogenesis[85,89-91]. Besides that, low progesterone level is directly associated with the development of endometrial cancer, and many studies have supported that low progesterone level is an indication of hyperandrogenism, thus providing a link between the two[92,93] (Figure 2).

A recent study has identified several PCOS-related genes (PRGs), based on literature review and genomics analysis, that showed significant genomic alterations in endometrial, ovarian, and breast cancers[94]. Interestingly, these PRGs included several cancer driver genes, such as *PTEN*, *ESR1*, and *TP53* in case of endometrial cancer, *PTEN* and *TP53* for ovarian cancer, and *ERBB2*, *NCOR1*, *ESR1*, *TP53*, *PTEN*, and *AKR1C3* for breast cancer[94]. Among these identified genes, the tumor suppressor gene *PTEN* showed the highest number of mutations in endometrial cancer[94]. Therefore, it is presumable that mutations in the cancer driver genes that are included in PRGs might drive PCOS patients toward cancer progression (Figure 2).

Based on clinical behavior and morphological feature, endometrial cancer can be generically divided into two distinct categories: Type I endometrial carcinoma, which is an estrogen-related malignancy with a favorable prognosis; and Type II endometrial carcinoma, which is not related to estrogen and carries a poor prognosis[95]. Studies have found increased endometrial expression of insulin signaling-related genes, such as *IGF1*, *IGFBP1*, and *PTEN*, both in PCOS and endometrial cancer patients[86,94]. *PTEN*, *KRAS*, *CTNNB1*, and *PIK3CA* mutations are

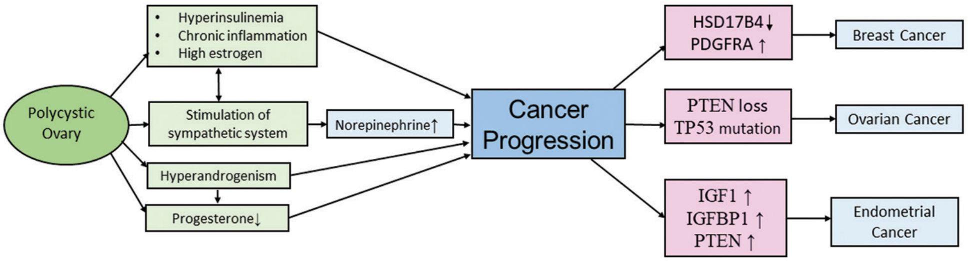


Figure 2. Proposed mechanism of cancer progression in polycystic ovary syndrome (PCOS). Chronic conditions such as hyperinsulinemia, high estrogen level, sympathetic hyperactivity, low progesterone level, hyperandrogenism, and chronic inflammation in PCOS might be the contributing factors to cancer progression. The chronic condition in PCOS might cause significant genomic alterations in the cancer driver genes that are included in PCOS-related genes, such as *PTEN*, *TP53*, *IGF1*, *IGFBP1*, *HSD17B4*, *PDGFRA*, and more, thus raising the possibility of developing cancer.

Table 2. Prevalence of non-reproductive cancer in polycystic ovary syndrome patients

Type of study	Location	Total participant	Cancer type (non-reproductive)	Number of cancer patients	Adjusted HR/RR/SIR/OR (95% CI value)	References
Cohort	Denmark	12,070	Lung	9	1.1 (0.5 – 2.0)	[69]
			Colon	11	2.1 (1.1 – 3.8)	
			Thyroid	8	1.5 (0.7 – 3.0)	
			Kidney	6	3.9 (1.4 – 8.4)	
			Brain	18	2.2 (1.3 – 3.5)	
			Melanoma	22	0.9 (0.5 – 1.3)	
Retrospective cohort	USA	2,560	Melanoma	5	1.17 (0.45 – 3.02)	[79]
			Thyroid	5	2.68 (1.24 – 10.63)	
			Lymphatic	5	1.99 (0.64 – 4.64)	

HR: Hazard ratio; OR: Odds ratio; RR: Relative risk; SIR: Standardized incidence ratio some of the common genetic mutations associated with Type I endometrial carcinoma[96]. These observations are indicative of how patients with PCOS might be more prone to developing endometrial cancer.

Elevated androgen level is positively correlated with the development of both PCOS and breast cancer[81]. Xu *et al.* identified 53 potent key genes that contribute to the onset of PCOS[81]. Among these genes, *HMGB2*, *PDGFRA*, and *HSD17B4* are involved in the development of male sexual characteristics. Xu *et al.* suggested that PCOS patients with downregulated HSD17B4 and upregulated PDGFRA may have more risk of developing breast cancer[81].

Ovarian cancer and PCOS are strongly associated; however, the underlying molecular mechanism remains largely unknown[97]. Ovarian cancer can be divided into two separate categories based on clinical behavior and molecular genetic abnormalities[98]. Low-grade endometriosis, borderline serous tumors, low-grade serous carcinomas, mucinous, and clear cell carcinomas are examples of Type I tumors, while examples of Type II ovarian cancer include undifferentiated tumors, carcinosarcomas, and high-grade serous carcinomas. High levels of genomic instability are present in Type II tumors. *TP53* is a PCOS-related gene that is highly mutated in high-grade serous carcinoma[94]. However, ovarian cancer, as observed in a mouse model, was not triggered by *TP53* mutation alone but, rather, was cooperatively associated with *PTEN* loss[99].

5. Conclusion

PCOS is becoming a growing public health concern globally because of its high incidence. Although there is a growing body of evidence indicating the inclination of PCOS patients to develop cancer due to shared metabolic and endocrine abnormalities, the overall association remains dubious. In our review, we extensively explored the association of PCOS with various types of cancers, along with the potential risk factors and molecular mechanism. We found that PCOS patients are more susceptible to developing endometrial cancer than other reproductive cancers, but there is contradictory evidence linking PCOS to either ovarian cancer or breast cancer. Due to the lack of studies, it is difficult to draw a strong link between PCOS and non-reproductive cancers. Disrupted hormonal balance, hyperinsulinemia, unchallenged high estrogen level, chronic inflammation, sympathetic hyperactivity, hyperandrogenism, and low progesterone level are all associated with normal PCOS progression and can also contribute to oncogenesis and cancer progression in PCOS. PRGs have shown significant genomic alterations in endometrial cancer, ovarian cancer, and breast cancer. These PRGs include several cancer driver genes, thereby indicating that mutations in the cancer driver genes that are included in the PRGs might drive PCOS patients toward cancer progression. Ethnically diverse and larger clinical trials along with molecular and bioinformatics approaches in an integrated manner are required to fully understand the association of PCOS with cancer.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Rubiat Afrin Ayon, Md. Ibrahim Khalil, and Md. Ashraful Hasan

Supervision: Md. Ibrahim Khalil and Md. Ashraful Hasan

Writing – original draft: Rubiat Afrin Ayon, Md. Azmain Faike, Sumiya Zaman Ononna, Aminul Hassan, Shoumik Kundu, and Md. Ashraful Hasan

Writing – review & editing: Rubiat Afrin Ayon, Farhana Akhter, Mohammad Mahfuz Ali Khan Shawan, Md. Mozammel Hossain, Md. Ibrahim Khalil, and Md. Ashraful Hasan

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Shermin S, Noor A, Jahan S, 2019, Polycystic ovary syndrome: A brief review with recent updates. *Delta Med Coll J*, 7(2): 84–99.

   https://doi.org/10.3329/dmcj.v7i2.45567

2. Hasan M, Sultana S, Sohan M, *et al.*, 2022, Prevalence and associated risk factors for mental health problems among patients with polycystic ovary syndrome in Bangladesh: A nationwide cross-sectional study. *PLoS One*, 17(6): e0270102.

   https://10.1371/journal.pone.0270102

3. Guzick DS, 2004, Polycystic ovary syndrome. *Obstet Gynecol*, 103(1): 181–193.

https://doi.org/10.1097/01.AOG.0000104485.44999.C6

4. Lim JJ, Lima PD, Salehi R, *et al.*, 2017, Regulation of androgen receptor signaling by ubiquitination during folliculogenesis and its possible dysregulation in polycystic ovarian syndrome. *Sci Rep*, 7(1): 10272.

   https://doi.org/10.1038/s41598-017-09880-0

5. Ruan X, Kubba A, Aguilar A, *et al.*, 2017, Use of cyproterone acetate/ethinylestradiol in polycystic ovary syndrome: Rationale and practical aspects. *Eur J Contracept Reprod Health Care*, 22(3): 183–190.

   https://doi.org/10.1080/13625187.2017.1317735

6. Kamrul-Hasan AB, Aalpona FT, Mustari M, *et al.*, 2020, Divergences in clinical, anthropometric, metabolic, and hormonal parameters among different phenotypes of polycystic ovary syndrome presenting at endocrinology outpatient departments: A multicenter study from Bangladesh. *J Hum Reprod Sci*, 13(4): 277–284.

   https://doi.org/10.4103/jhrs.JHRS_34_20

7. Fauser BC, Tarlatzis BC, Rebar RW, *et al.*, 2012, Consensus on women's health aspects of polycystic ovary syndrome (PCOS): The Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus workshop Group. *Fertil Steril*, 97(1): 28–38.e25.

   https://doi.org/10.1016/j.fertnstert.2011.09.024

8. Corbould A, Kim YB, Youngren JF, *et al.*, 2005, Insulin resistance in the skeletal muscle of women with PCOS involves intrinsic and acquired defects in insulin signaling. *Am J Physiol Endocrinol Metab*, 288(5): E1047–E1054.

   https://doi.org/10.1152/ajpendo.00361.2004

9. Balen A, 2001, Polycystic ovary syndrome and cancer. *Hum Reprod Update*, 7(6): 522–525.

   https://doi.org/10.1093/humupd/7.6.522

10. Siegel RL, Miller KD, Jemal A, 2016, Cancer statistics, 2016. *CA Cancer J Clin*, 66(1): 7–30.

    https://doi.org/10.3322/caac.21332

11. Stein IF, Leventhal ML, 1935, Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol*, 29(2): 181–191.

    https://doi.org/10.1016/S0002-9378(15)30642-6

12. Speert H, 1949, Carcinoma of the endometrium in young women. *Surg Gynecol Obstet*, 88(3): 332–336.

13. Harris HR, Terry KL, 2016, Polycystic ovary syndrome and risk of endometrial, ovarian, and breast cancer: A systematic review. *Fertil Res Pract*, 2(1): 1–9.

    https://doi.org/10.1186/s40738-016-0029-2

14. Cirillo PM, Wang ET, Cedars MI, *et al.*, 2016, Irregular menses predicts ovarian cancer: Prospective evidence from the child health and development studies. *Int J Cancer*, 139(5): 1009–1017.

    https://doi.org/10.1002/ijc.30144

15. Li Y, Chen C, Ma Y, *et al.*, 2019, Multi-system reproductive metabolic disorder: Signifcance for the pathogenesis and therapy of polycystic ovary syndrome (PCOS). *Life Sci*, 228: 167–175.

    https://doi.org/10.1016/j.lfs.2019.04.046

16. Gadducci A, Gargini A, Palla E, *et al.*, 2005, Polycystic ovary syndrome and gynecological cancers: Is there a link? *J Gynaecol Endocrinol*, 20(4): 200–208.

    https://doi.org/10.1080/09513590400021201

17. Syed V, Ulinski G, Mok SC, *et al.*, 2001, Expression of gonadotropin receptor and growth responses to key reproductive hormones in normal and malignant human ovarian surface epithelial cells. *Cancer Res*, 61(18): 6768–6776.

18. Modugno F, 2004, Ovarian cancer and polymorphisms in the androgen and progesterone receptor genes: A HuGE review. *Am J Epidemiol*, 159(4): 319–335.

    https://doi.org/10.1093/aje/kwh046

19. Ding DC, Chen W, Wang JH, *et al.*, 2018, Association between polycystic ovarian syndrome and endometrial, ovarian, and breast cancer: A population-based cohort study in Taiwan. *Medicine (Baltimore)*, 97(39): e12608.

    https://doi.org/10.1097/MD.0000000000012608

20. Beral V, Doll R, Hermon C, *et al.*, 2008, Ovarian cancer and oral contraceptives: Collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet*, 371(9609): 303–314.

    https://doi.org/10.1016/S0140-6736(08)60167-1

21. Anderson KE, Sellers TA, Chen PL, *et al.*, 1997, Association of Stein Leventhal syndrome with the incidence of postmenopausal breast carcinoma in a large prospective study of women in Iowa. *Cancer*, 79(3): 494–499.

22. Shobeiri F, Jenabi E, 2016, The association between polycystic ovary syndrome and breast cancer: A meta-analysis. *Obstet Gynecol Sci*, 59(5): 367–372.

    https://doi.org/10.5468/ogs.2016.59.5.367

23. Fanta M, 2013, Is polycystic ovary syndrome, a state of relative estrogen excess, a real risk factor for estrogen-dependant malignancies? *Gynecol Endocrinol*, 29(2): 145–147.

    https://doi.org/10.3109/09513590.2012.730575

24. Urbanek M, Legro Driscoll DA, *et al.*, 1999, Thirty-seven candidate genes for polycystic ovary syndrome: Strongest evidence for linkage is with follistatin. *Proc Natl Acad Sci U S A*, 96(15): 8573–8578.

    https://doi.org/10.1073/pnas.96.15.8573

25. Liu Y, Gold EB, Lasley BL, *et al.*, 2004, Factors affecting

menstrual cycle characteristics. *Am J Epidemiol*, 160(2): 131–140.

https://doi.org/10.1093/aje/kwh188

26. Rani R, Hajam YA, Kumar R, *et al.*, 2022, A landscape analysis of the potential role of polyphenols for the treatment of Polycystic Ovarian Syndrome (PCOS). *Phytomed Plus*, 2(1): 100161.

    https://doi.org/10.1016/j.phyplu.2021.100161

27. Bulsara J, Patel P, Soni A, *et al.*, 2021, A review: Brief insight into polycystic ovarian syndrome. *Endocrinol Metab Sci*, 3: 100085.

    https://doi.org/10.1016/j.endmts.2021.100085

28. Ibáñez L, Oberfield SE, Witchel S, *et al.*, 2017, An international consortium update: Pathophysiology, diagnosis, and treatment of polycystic ovarian syndrome in adolescence. *Horm Res Paediatr*, 88(6): 371–395.

    https://doi.org/10.1159/000479371

29. Dumesic DA, Oberfield SE, Stener-Victorin E, *et al.*, 2015, Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocr Rev*, 36(5): 487–525.

    https://doi.org/10.1159/000479371

30. Kanbour SA, Dobs AS, 2022, Hyperandrogenism in women with polycystic ovarian syndrome: Pathophysiology and controversies. *Androgens Clin Res Ther*, 3(1): 22–30.

    https://doi.org/10.1089/andro.2021.0020

31. Cheung AP, Cog F, 2010, Polycystic ovary syndrome: A contemporary view. J *Obstet Gynaecol Can*, 32(5): 423–425.

    https://doi.org/10.1016/S1701-2163(16)34493-0

32. Walters KA, Gilchrist RB, Ledger WL, *et al.*, 2018, New perspectives on the pathogenesis of PCOS: Neuroendocrine origins. *Trends Endocrinol Metab*, 29(12): 841–852.

    https://doi.org/10.1016/j.tem.2018.08.005

33. Lv PP, Jin M, Rao JP, *et al.*, 2020, Role of anti-Müllerian hormone and testosterone in follicular growth: A cross-sectional study. *BMC Endocr Disord*, 20(1): 101.

    https://doi.org/10.1186/s12902-020-00569-6

34. Ashraf S, Nabi M, Rashid F, *et al.*, 2019, Hyperandrogenism in polycystic ovarian syndrome and role of CYP gene variants: A review. *Egypt J Med Hum Genet*, 20(1): 25.

    https://doi.org/10.1186/s43042-019-0031-4

35. Yau TT, Ng NY, Cheung LP, *et al.*, 2017, Polycystic ovary syndrome: A common reproductive syndrome with long-term metabolic consequences. *Hong Kong Med J*, 23(6): 622–634.

    https://doi.org/10.12809/hkmj176308

36. Yilmaz M, Bukan N, Ayvaz G, *et al.*, 2005, The effects of rosiglitazone and metformin on oxidative stress and homocysteine levels in lean patients with polycystic ovary syndrome. *Hum Reprod*, 20(12): 3333–3340.

    https://doi.org/10.1093/humrep/dei258

37. Moller DE, Flier JS, 1988, Detection of an alteration in the insulin-receptor gene in a patient with insulin resistance, acanthosis nigricans, and the polycystic ovary syndrome (Type A insulin resistance). *N Engl J Med*, 319(23): 1526–1529.

    https://doi.org/10.1056/NEJM198812083192306

38. Diamanti-Kandarakis E, Papavassiliou AG, 2006, Molecular mechanisms of insulin resistance in polycystic ovary syndrome. *Trends Mol Med*, 12(7): 324–332.

    https://doi.org/10.1016/j.molmed.2006.05.006

39. Burghen GA, Givens JR, Kitabchi AE, 1980, Correlation of hyperandrogenism with hyperinsulinism in polycystic ovarian disease. *J Clin Endocrinol Metab*, 50(1): 113–116.

    https://doi.org/10.1210/jcem-50-1-113

40. Armanini D, Boscaro M, Bordin L, *et al.*, 2022, Controversies in the pathogenesis, diagnosis and treatment of PCOS: Focus on insulin resistance, inflammation, and hyperandrogenism. *Int J Mol Sci*, 23(8): 4110.

    https://doi.org/10.3390/ijms23084110

41. Willis D, Franks S, 1995, Insulin action in human granulosa cells from normal and polycystic ovaries is mediated by the insulin receptor and not the Type-I insulin-like growth factor receptor. *J Clin Endocrinol Metab*, 80(12): 3788–3790.

    https://doi.org/10.1210/jcem.80.12.8530637

42. Azziz R, Carmina E, Dewailly D, *et al.,* 2009, The androgen excess and PCOS society criteria for the polycystic ovary syndrome: The complete task force report. *Fertil Steril*, 91(2): 456–488.

    https://doi.org/10.1016/j.fertnstert.2008.06.035

43. Gambineri A, Patton L, Altieri P, *et al.*, 2012, Polycystic ovary syndrome is a risk factor for Type 2 diabetes: Results from a long-term prospective study. *Diabetes*, 61(9): 2369–2374.

    https://doi.org/10.2337/db11-1360

44. Bremer AA, Miller WL, 2008, The serine phosphorylation hypothesis of polycystic ovary syndrome: A unifying mechanism for hyperandrogenemia and insulin resistance. *Fertil Steril*, 89(5): 1039–1048.

    https://doi.org/10.1016/j.fertnstert.2008.02.091

45. Siddiqui S, Mateen S, Ahmad R, *et al.*, 2022, A brief insight into the etiology, genetics, and immunology of polycystic ovarian syndrome (PCOS). *J Assist Reprod Genet,* 39(11): 2439–2473.

    https://doi.org/10.1007/s10815-022-02625-7

46. Moll GW Jr., Rosenfield RL, 1983, Plasma free testosterone in the diagnosis of adolescent polycystic ovary syndrome.

J Pediatr, 102(3): 461–464.

https://doi.org/10.1016/S0022-3476(83)80678-7

47. Van Hooff MH, Voorhorst FJ, Kaptein MB, *et al.*, 2000, Polycystic ovaries in adolescents and the relationship with menstrual cycle patterns, luteinizing hormone, androgens, and insulin. *Fertil Steril*, 74(1): 49–58.

    https://doi.org/10.1016/S0015-0282(00)00584-7

48. Diamanti-Kandarakis E, Dunaif A, 2012, Insulin resistance and the polycystic ovary syndrome revisited: An update on mechanisms and implications. *Endocr Rev*, 33(6): 981–1030.

    https://doi.org/10.1210/er.2011-1034

49. Zuo T, Zhu M, Xu W, 2016, Roles of oxidative stress in polycystic ovary syndrome and cancers. *Oxid Med Cell Longev*, 2016: 8589318.

    https://doi.org/10.1155/2016/8589318

50. Sabuncu T, Vural H, Harma M, *et al.*, 2001, Oxidative stress in polycystic ovary syndrome and its contribution to the risk of cardiovascular disease. *Clin Biochem*, 34(5): 407–413.

    https://doi.org/10.1016/S0009-9120(01)00245-4

51. Patel S, 2018, Polycystic ovary syndrome (PCOS), an inflammatory, systemic, lifestyle endocrinopathy. *J Steroid Biochem Mol Biol*, 182: 27–36.

    https://doi.org/10.1016/j.jsbmb.2018.04.008

52. Xue Z, Li J, Feng J, *et al.,* 2021, Research progress on the mechanism between polycystic ovary syndrome and abnormal endometrium. *Front Physiol*, 12: 788772.

    https://doi.org/10.3389/fphys.2021.788772

53. Yang Y, Qiao J, Li R, *et al.,* 2011, Is interleukin-18 associated with polycystic ovary syndrome? *Reprod Biol Endocrinol*, 9(1): 7.

    https://doi.org/0.1186/1477-7827-9-7

54. González F, Nair KS, Daniels JK, *et al.*, 2012, Hyperandrogenism sensitizes leukocytes to hyperglycemia to promote oxidative stress in lean reproductive-age women. *J Clin Endocrinol Metab*, 97(8): 2836–2843.

    https://doi.org/10.1210/jc.2012-1259

55. Rosenfield RL, Ehrmann DA, 2016, The pathogenesis of polycystic ovary syndrome (PCOS): The hypothesis of PCOS as functional ovarian hyperandrogenism revisited. *Endocr Rev*, 37(5): 467–520.

    https://doi.org/10.1210/er.2015-1104

56. Day F, Karaderi T, Jones MR, *et al.*, 2018, Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet*, 14(12): e1007813.

    https://doi.org/10.1371/journal.pgen.1007813

57. Escobar-Morreale HF, 2018, Polycystic ovary syndrome: Definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinol*, 14(5): 270–284.

    https://doi.org/10.1038/nrendo.2018.24

58. Azziz R, Carmina E, Chen Z, *et al.*, 2016, Polycystic ovary syndrome. *Nat Rev Dis Primers*, 2(1): 16057.

    https://doi.org/10.1038/nrdp.2016.57

59. Lizneva D, Suturina L, Walker W, *et al.*, 2016, Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril*, 106(1): 6–15.

    https://doi.org/10.1016/j.fertnstert.2016.05.003

60. Meczekalski B, Pérez-Roncero GR, López-Baena MT, *et al.*, 2020, The polycystic ovary syndrome and gynecological cancer risk. *J Gynaecol Endocrinol*, 36(4): 289–293.

    https://doi.org/10.1080/09513590.2020.1730794

61. Barry JA, Azizia MM, Hardiman PJ, 2014, Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: A systematic review and meta-analysis. *Hum Reprod Update*, 20(5): 748–758.

    https://doi.org/10.1093/humupd/dmu012

62. de França Neto AH, Rogatto S, Do Amorim MM, *et al.*, 2010, Oncological repercussions of polycystic ovary syndrome. *J Gynaecol Endocrinol*, 26(10): 708–711.

    https://doi.org/10.3109/09513590.2010.490607

63. Parker SL, 1997, Cancer statistics, 1997. *CA Cancer J Clin*, 47: 5–27.

    https://doi.org/10.3322/canjclin.47.1.5

64. National Cancer Institute of Canada, 2003, Canadian Cancer Statistics. Toronto, Canada: National Cancer Institute.

65. Daniilidis A, Dinas K, 2009, Long term health consequences of polycystic ovarian syndrome: A review analysis. *Hippokratia*, 13(2): 90–92.

66. Hardiman P, Pillay OS, Atiomo W, 2003, Polycystic ovary syndrome and endometrial carcinoma. *Lancet*, 361(9371): 1810–1812.

    https://doi.org/10.1016/S0140-6736(03)13409-5

67. Pillay OC, Te Fong, LW, Crow JC, *et al.*, 2006, The association between polycystic ovaries and endometrial cancer. *Hum Reprod*, 21(4): 924–929.

    https://doi.org/10.1093/humrep/dei420

68. Atiomo W, Khalid S, Parameshweran S, *et al.*, 2009, Proteomic biomarkers for the diagnosis and risk stratification of polycystic ovary syndrome: A systematic review. *BJOG*, 116(2): 137–143.

    https://doi.org/10.1111/j.1471-0528.2008.02041.x

69. Gottschau M, Kjaer SK, Jensen A, *et al.*, 2015, Risk of cancer among women with polycystic ovary syndrome: A Danish cohort study. *Gynecol Oncol,* 136(1): 99–103.

https://doi.org/10.1016/j.ygyno.2014.11.012

70. Shen CC, Yang AC, Hung JH, *et al.*, 2015, A nationwide population-based retrospective cohort study of the risk of uterine, ovarian and breast cancer in women with polycystic ovary syndrome. *Oncologist*, 20(1): 45–49.

    https://doi.org/10.1634/theoncologist.2014-0311

71. Haoula Z, Salman M, Atiomo W, 2012, Evaluating the association between endometrial cancer and polycystic ovary syndrome. *Hum Reprod*, 27(5): 1327–1331.

    https://doi.org/10.1093/humrep/des042

72. Fearnley EJ, Marquart L, Spurdle AB, *et al.*, 2010, Polycystic ovary syndrome increases the risk of endometrial cancer in women aged less than 50 years: An Australian case-control study. *Cancer Causes Control*, 21(12): 2303–2308.

    https://doi.org/10.1007/s10552-010-9658-7

73. Reid BM, Permuth JB, Sellers TA, 2017, Epidemiology of ovarian cancer: A review. *Cancer Biol Med,* 14(1): 9–32.

    https://doi.org/10.20892/j.issn.2095-3941.2016.0084

74. Risch HA, 1998, Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *J Natl Cancer Inst,* 90(23): 1774–1786.

    https://doi.org/10.1093/jnci/90.23.1774

75. Hunn J, Rodriguez GC, 2012, Ovarian cancer: Etiology, risk factors, and epidemiology. *Clin Obstet Gynecol*, 55(1), 3–23.

    https://doi.org/10.1097/GRF.0b013e31824b4611

76. Ali AT, 2014, Reproductive factors and the risk of endometrial cancer. *Int J Gynecol Cancer,* 24(3): 384–393.

    https://doi.org/10.1097/IGC.0000000000000075

77. Rojas K, Stuckey A, 2016, Breast cancer epidemiology and risk factors. *Clin Obstet Gynecol*, 59(4): 651–672.

    https://doi.org/10.1097/GRF.0000000000000239

78. Pérez-López FR, 2019, Sporadic ovarian and fallopian tube cancer in postmenopausal women. In: Postmenopausal Diseases and Disorders. Cham: Springer. p79–100.

    https://doi.org/10.1007/978-3-030-13936-0_5

79. Brinton LA, Moghissi KS, Westhoff CL, *et al.*, 2010, Cancer risk among infertile women with androgen excess or menstrual disorders (including polycystic ovary syndrome). *Fertil Steril*, 94(5): 1787–1792.

    https://doi.org/10.1016/j.fertnstert.2009.10.012

80. Bray F, Ferlay J, Soerjomataram I, *et al.*, 2018, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6): 394–424.

    https://doi.org/10.3322/caac.21492

81. Xu H, Han Y, Lou J, *et al.*, 2017, PDGFRA, HSD17B4 and HMGB2 are potential therapeutic targets in polycystic ovarian syndrome and breast cancer. *Oncotarget*, 8(41): 69520–69526.

    https://doi.org/10.18632/oncotarget.17846

82. Cooney LG, Dokras A, 2018, Beyond fertility: Polycystic ovary syndrome and long-term health. *Fertil Steril*, 110(5): 794–809.

    https://doi.org/10.1016/j.fertnstert.2018.08.021

83. Terry KL, Willett WC, Rich-Edwards JW, *et al.*, 2006, A prospective study of infertility due to ovulatory disorders, ovulation induction, and incidence of breast cancer. *Arch Intern Med*, 166(22): 2484–2489.

    https://doi.org/10.1001/archinte.166.22.2484

84. Ledger WL, 2003, Non-reproductive consequences of polycystic ovary syndrome. *Curr Obstet Gynecol*, 13(6): 350–354.

    https://doi.org/10.1016/S0957-5847(03)00072-6

85. Mravec B, Tibensky M, 2020, Increased cancer risk in polycystic ovary syndrome: An (un) sympathetic connection? *Med Hypotheses*, 134: 109437.

    https://doi.org/10.1016/j.mehy.2019.109437

86. Shafiee MN, Chapman C, Barrett D, *et al.*, 2013, Reviewing the molecular mechanisms which increase endometrial cancer (EC) risk in women with polycystic ovarian syndrome (PCOS): Time for paradigm shift? *Gynecol Oncol,* 131(2): 489–492.

    https://doi.org/10.1016/j.ygyno.2013.06.032

87. Arcidiacono B, Iiritano S, Nocera A, *et al.*, 2012, Insulin resistance and cancer risk: An overview of the pathogenetic mechanisms. *Exp Diabetes Res,* 2012: 789174.

    https://doi.org/10.1155/2012/789174

88. Grivennikov SI, Greten FR, Karin M, 2010, Immunity, inflammation, and cancer. *Cell J*, 140(6): 883–899.

    https://doi.org/10.1016/j.cell.2010.01.025

89. Zahalka AH, Arnal-Estapé A, Maryanovich M, *et al.*, 2017, Adrenergic nerves activate an angio-metabolic switch in prostate cancer. *Science,* 358(6361): 321–326.

    https://doi.org/10.1126/science.aah5072

90. Hara MR, Kovacs JJ, Whalen EJ, *et al.*, 2011, A stress response pathway regulates DNA damage through β2-adrenoreceptors and β-arrestin-1. *Nature,* 477(7364): 349–353.

    https://doi.org/10.1038/nature10368

91. Magnon C, Hall SJ, Lin J, *et al.*, 2013, Autonomic nerve development contributes to prostate cancer progression. *Science,* 341(6142): 1236361.

    https://doi.org/10.1126/science.1236361

92. Papaioannou S, Tzafettas J, 2010, Anovulation with or without PCO, hyperandrogenaemia and hyperinsulinaemia as promoters of endometrial and breast cancer. *Best Pract Res Clin Obstet Gynaecol*, 24(1): 19–27.

    https://doi.org/10.1016/j.bpobgyn.2008.11.010

93. Dumesic DA, Lobo RA, 2013, Cancer risk and PCOS. *Steroids*, 78(8): 782–785.

    https://doi.org/10.1016/j.steroids.2013.04.004

94. Yumiceba V, López-Cortés A, Pérez-Villa A, *et al.*, 2020, Oncology and pharmacogenomics insights in polycystic ovary syndrome: An integrative analysis. *Front Endocrinol (Lausanne)*, 11: 585130.

    https://doi.org/10.3389/fendo.2020.585130

95. Setiawan VW, Yang HP, Pike MC, *et al.*, 2013, Type I and II endometrial cancers: Have they different risk factors? *Clin Oncol,* 31(20): 2607–2018.

    https://doi.org/10.1200%2FJCO.2012.48.2596

96. Samarnthai N, Hall K, Yeh IT, 2010, Molecular profiling of endometrial malignancies. *Obstet Gynecol,* 2010: 162363.

    https://doi.org/10.1155/2010/162363

97. Zou J, Li Y, Liao N, *et al.*, 2022, Identification of key genes associated with polycystic ovary syndrome (PCOS) and ovarian cancer using an integrated bioinformatics analysis. *J Ovarian Res,* 15(1): 1–16.

    https://doi.org/10.1186/s13048-022-00962-w

98. Toss A, Tomasello C, Razzaboni E, *et al.*, 2015, Hereditary ovarian cancer: Not only BRCA 1 and 2 genes. *Biomed Res Int*, 2015: 341723.

    https://doi.org/10.1155/2015/341723

99. Perets R, Wyant GA, Muto KW, *et al.*, 2013, Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. *Cancer Cell*, 24(6): 751–765.

    https://doi.org/10.1016/j.ccr.2013.10.013

ACCSCIENCE PUBLISHING

Tumor Discovery

ORIGINAL RESEARCH ARTICLE

An approach for classification of lung nodules

Naveen HM[1]*, Naveena C[1], and Manjunath Aradhya VN[2]

[1]Department of Computer Science Engineering, SJB Institute of Technology, Bangalore, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

[2]Department of Master Computer Application, JSSTU, Mysuru, Affiliated to JSS Science and Technology University, Mysuru, Karnataka, India

Abstract

The main objective of the proposed work is to develop an automated computer-aided detection (CAD) system to classify lung nodules using various classifiers from computed tomography (CT) images. One of the most important steps in lung nodule detection is the classification of nodule and non-nodule patterns in CT. The early detection of the condition helps lower the mortality rate. The developed CAD systems consist of segmentation, feature extraction, and classification. In this work, a filter method is used to segment the infected region. Later, we extracted features through and fed into classifiers such as Decision Stump (DS), Random Forest (RF), and Back Propagation Neural Network (BPNN). The experimentation was conducted on LIDC-IDRI dataset, and the results with BPNN outperformed those with DS and RF classifiers.

Keywords: Decision stump; Random forest; AdaBoost-Decision stump; AdaBoost-Random forest; Back propagation neural network

***Corresponding author:**
Naveen HM
(naveenhm056@gmail.com)

Citation: Naveen HM, Naveena C, and Aradhya VNM, 2023, An approach for classification of lung nodules. *Tumor Discov,* 2(1): 317. https://doi.org/10.36922/td.317

Received: December 28, 2022
Accepted: February 17, 2023
Published Online: March 8, 2023

1. Introduction

The second most frequent cancer in both men and women is thought to be lung cancer. It is the main factor in cancer-related fatalities. According to the most recent estimates, there are around 7.6 million cancer-related deaths globally each year, according to the most recent numbers supplied by the World Health Organization[1]. Furthermore, it is anticipated that the number of deaths from lung cancer would keep increasing, reaching almost 17 million in 2030. Successful treatment of lung cancer depends greatly on early detection. Significant data suggest that early identification of lung cancer will reduce mortality rates[2]. Lung cancer in an early stage manifests itself as a pulmonary nodule, which grows rapidly and later becomes a tumor. The characteristics of pulmonary nodules are based on calcification, internal structure, sphericity, speculation, subtlety, and texture. Nodules usually appear smaller in medical images. Hence, detection of pulmonary nodule is one of the most challenging tasks[3].

Various imaging techniques, including radiography, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography-CT (PET-CT), among others, can be used to detect pulmonary nodules. Radiologists face a challenging problem when trying to find lung nodules on radiographs, because nodules present behind the rib cages are hidden and the miss rate could increase up to 30%[4,5]. MRI and PET-CT techniques are more expensive and time-consuming. The

CT imaging is less expensive and produces a variety of cross-sectional images of complete chest within a single breath hold. For the analysis and early detection of lung nodules, it is now regarded as the best imaging technique. Majority of the pulmonary nodules are benign; however, a small populace of them grow to be malignant. The radiologists examine the CT scan to conclude whether a nodule presents a chance for malignancy. The radiologist finds it challenging and time-consuming to detect some nodules in the CT because of non-pathological features[6]. To get around this, radiologists opt for a computer-assisted approach as a backup method to validate their interpretation. Computer-based processes like computer-aided detection (CAD) act as the radiologist's "second pair of eyes" to analyze medical pictures for any problematic regions. CAD is a topical technique designed to improve the radiologists' ability to find even the smallest lung nodules at their earliest stages.

The main objective of CAD is to improve disease identification by lowering the false negative (FN) rate brought on by observational omission. CAD was created with the purpose of improving systematic clinical decision-making and the performance of detection in medical imaging modalities. The aim of this paper is to focus on the architecture of different stages of CAD design with fruitful results to assist the radiologists in detecting lung nodule at early stage.

2. Literature survey

In the field of lung cancer and in relation to the work of this study, there are a number of existing models and algorithms. The stages of CAD systems can be used to categorize the existing work.

Various techniques, such as region growing[7,8], watershed segmentation[9], fuzzy logic, active contours, intensity-based thresholding, graph search algorithm[10], etc., were used for region of interest (ROI) image segmentation Segmentation is an essential step in nodule detection. The survey in various existing segmentation techniques for lung nodule is discussed in detail. Dai *et al.*[11] developed a segmentation algorithm known as segmentation by registration. The authors compared their work with algorithms such as automatic region growing, interactive region growing and voxel classification. Daneshmand *et al.*[12] demonstrated a precise technique designed for toning lung nodules in chest CT scans. The region growing, optimal thresholding and optimal cube registration were used in this system Dawoud *et al.*[13] proposed an adaptive border marching which is a geometric-based algorithm used for segmenting the lung. De Nunzio *et al.*[14] performed phantom experiment for CT image. The density of the nodule and its size was measured using the point spread function. Deep *et al.*[15] presented an algorithm for segmenting different nodule types such as juxtravascular, pleura tail, juxtapleural, solid, and non-solid nodules.

Segmentation technique such as region growing and some hybrid fuzzy connectivity models was implemented. Another segmentation scheme proposed by Dehmeshki *et al.*[16] used two different datasets taken from the LIDC database. Two different techniques such as dynamic programming model and multidirection fusion techniques are used to know the information, relationship between adjacent slices and to reduce segmentation error. Dehmeshki *et al.*[17] utilized a three-step procedure to find lung nodules. Initially, lung regions are segmented using an adaptive threshold algorithm. Second, lung vessel was removed using active contour model (ACM) and finally, the suspicious nodules were located using a Hessian matrix (selected shape filter). Delogu *et al.*[18] developed a fully automated segmentation method for detecting the pulmonary nodule. The authors used region growing approach for a set of 130 CT images. Only 84 images produced satisfactory result. Dheepak *et al.*[19] proposed a methodology for segmentation of juxtapleural lung nodules. The authors used two techniques for detecting the nodules from the lung. They are region growing and shape curvature-based techniques. Doi *et al.*[20] developed a method for segmenting the lung with the help of CT data. The method was fully automatic and was composed of two steps. They are robust active shape model (RASM) and optimal surface finding method. Dolejsi *et al.*[21] proposed a CAD to reduce the lung volume and juxtapleural nodule from thoracic CT images. For segmenting the lung volume and nodule in juxtapleural, region growing and a 3D-mass-spring mdel (MSM) was used. Elizabeth *et al.*[22] proposed a region-based ACM based on local divergence energies. This model was designed for blurred boundary and noisy images. The author used regularization function to smoothen the boundary from different noise level. The system performance was evaluated with Chan-Vese's (CV) model, region scalable fitting and local Gaussian fitting. Enquobahrie *et al.*[23] developed an edge detection model, which precisely diffuses the edge space. Farag *et al.*[24] proposed a newly developed nodule segmentation algorithm which was stable, accurate, and automated. Farag *et al.*[25] developed a segmentation model to segment the juxtavascular and ground glass (GG) nodules. The authors proposed parametric mixture model for juxtavascular nodules and ACM for detecting leakage boundary.

A nonlinear level set method proposed by Farag *et al.*[26] used adaptive velocity function and edge stopping function to employ a noise-free segmentation model. Gambhir *et al.*[27] focused on segmenting the lung nodule

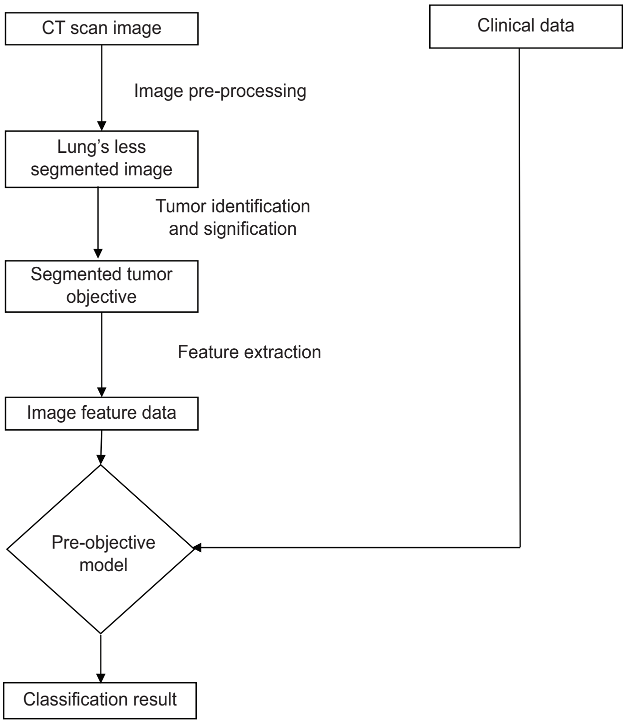


Figure 1. General workflow of the process of developing and using predictive models.

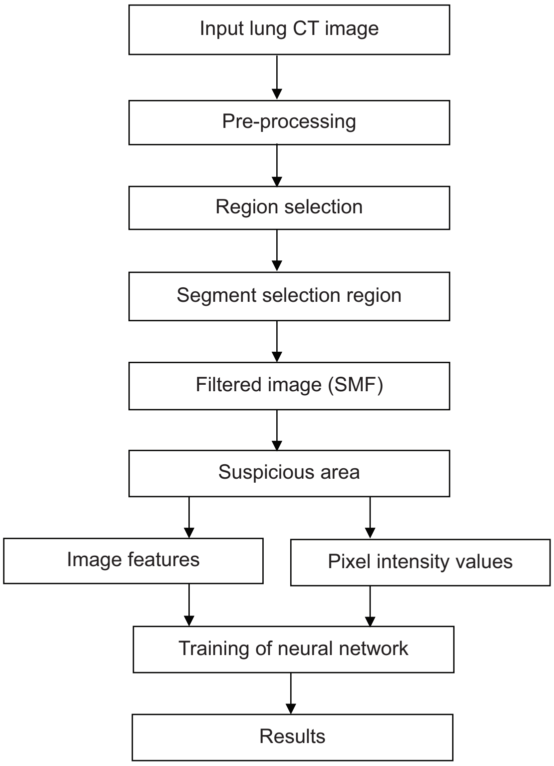


Figure 2. Pre-processing selective median filter.

with lesser number of false positive (FP) findings. Garro *et al.*[28] proposed a method to segment the juxtapleural nodule and lung vessels from the CT image. Golosio *et al.*[29] segmented the pleural and vessels in lung CT. Gomathi *et al.*[30] developed a segmentation method to improve nodule detection accuracy. The authors mainly focused on juxtapleural nodule for image segmentation. A parameter-free algorithm such as bidirectional chain coding method was used to smoothen the lung border. Gomathi *et al.*[31] presented a segmentation algorithm to produce efficient and accurate result. An improved graph that cuts algorithm along with Gaussian mixture models (GMMs) was proposed to segment the lung nodule. Gon alves *et al.*[32] developed a hybrid segmentation technique which combined the fully automatic and semi-automatic global segmentation technique. Gould *et al.*[33] formulated central medialness adaptive principle, a Hessian-based strategy, to segment the lung nodule in CT images. Multi-resolution contour let transform Grigorescu *et al.*[34] can also be used to extract the features. These features are used for further processing in the classification, which is the final stage of the CAD system. Gu *et al.*[35] proposed a technique to detect the nodule using template-based model. The minimum and maximum Hounsfield density (HU) was obtained from the intensity of nodule data. Shape-based or shape-texture-based methods resulted in an overall detection process with the lowest accuracy. The existing segmentation techniques produced low accuracy, high error rate, reduced similarity coefficient, long computation time, etc. Medical image segmentation is difficult due to complexity and diversity of anatomical structures on one hand and particular properties such as noise and low contrast (non-solid nodules), on the other hand.

Gudise *et al.*[36] comparative study is made on the computational requirements of the PSO and BP as training algorithms for neural networks. Hua *et al.*[37] presents an automatic algorithm for pathological lung CT image segmentation that uses a graph search driven by a cost function combining the intensity, gradient, boundary smoothness, and the rib information. Jacobs *et al.*[38], a CAD system that combines the output of two prototype CAD systems aimed at detection of ground glass nodules and solid nodules, respectively, could lead to efficient detection of the entire spectrum of lung nodules in chest CT scans.

Shen *et al.*[39] proposes a parameter-free lung segmentation algorithm with the aim of improving lung nodule detection accuracy, focusing on juxtapleural nodules. A bidirectional chain coding method combined with a support vector machine (SVM) classifier is used to selectively smooth the lung border while minimizing the over-segmentation of adjacent regions. Shen *et al.*[40] proposed a robust segmentation technique based on an extension to the traditional fuzzy c-means (FCM) clustering algorithm.

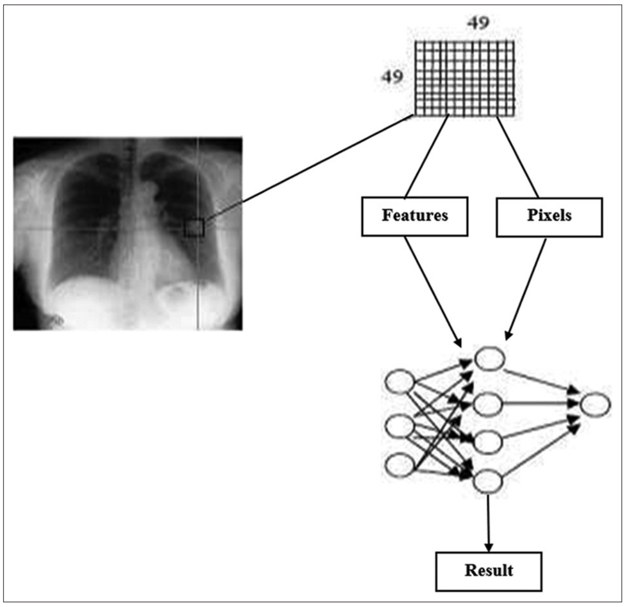


Figure 3. Layer-based region segmentation.

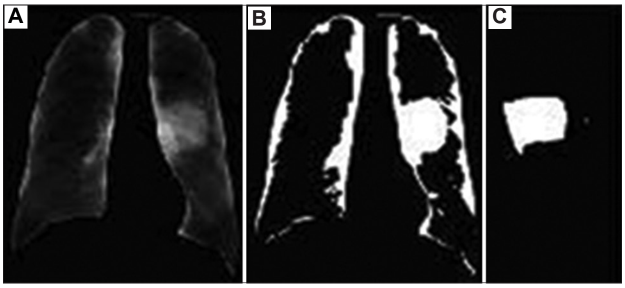


Figure 4. (A) Image after growing, (B) thresholding, and (C) segmented nodule.

Shi *et al.*[41] presents the Optimized Kalman Particle Swarm (OKPS) filter. This filter results from two years of research and improves the Swarm Particle Filter (SPF). Shih-Chung *et al.*[42] presented to predict long term survival versus short term survival. Forty adenocarcinoma diagnostic lung computed tomography (CT) scans from Moffitt Cancer Center were analyzed for survival prediction. A decision tree classifier was able to predict the survival group with an accuracy of 77.5%.

Yuan[43] proposed model can handle blurry boundaries and noise problems. In addition, the regularity of the level set function is intrinsically preserved by the level set regularization term to ensure accurate computation. Zhou S *et al.*[44] proposed a fast and fully automatic scheme based on iterative weighted averaging and adaptive curvature threshold is proposed in this study to facilitate accurate lung segmentation for inclusion of juxtapleural nodules and pulmonary vessels and ensure the smoothness of the lung boundary.

In this work, we proposed efficient method for segmentation and classification of lung nodules.

3. Geometric features

Geometric features are considered the first set of features. The vital structural information of tumor can be easily analyzed with the 2D and 3D geometric features. The evaluation of the geometric features is very useful in quantifying and analyzing the biomedical images like CT scans (2D, 3D)[45].

Image object is formed by the numerous pixels and is rescaled using unit information. The one unit is the area of the single pixel, which denotes that the number of pixels forming the image is the area of the image. If we have the unit information of the image data provided, then the area of the whole image is equal to the product of the area covered by one pixel and the number of pixels unit in the image object. In this chapter, a fully automatic method is described by the authors to detect the cancer in the lungs. This method comprises three sequential steps. The first step is to implement the gray level thresholding method to separate lung region from the image. The second step is the detection of the anterior and posterior junctions to separate the left lung and right lung region. The final step is the smoothening on the boundary of lung along the mediastinum. There are some differences between our proposed and the previous works.

The authors demonstrated that an automated texture mapping methods. The proposed work is experimental in nature: we propose an efficient technique to discover the gray scale qualities of an HRCT dataset with the co-training paradigm. We utilize an effective technique to enhance classifiers that are prepared with not very many posterior and anterior intersection lines between the marked illustrations utilizing a huge pool of concealed right and left lungs.

Finally, to get more cases, there are two or more disjoint functions called views. Processing time and stable results even leaving the lung. It has also been shown that the structures named by experts are smooth with the lung and can be connected stepwise within the frame with irregular boundaries along the mediastinal pathway (Kawane *et al.*[46]). The outcomes are likewise analyzed against "density mask," as of now a standard approach utilized for emphysema recognition in medicinal picture analysis and other automated procedures utilized for arrangement of emphysema in the literature. The new framework can group diffuse districts of emphysema beginning from a bullous setting.

The classifiers worked at various iterations additionally seem to demonstrate an intriguing relationship with various levels of emphysema, which merits more investigation[47]. The authors suggested using form features to distinguish between obstructive lung infections, and the results showed increased classification sensitivity when compared to features based solely on texture. Be that as it may, their proposed framework is reliant on the region size, for example, 16 × 16, 32 × 32, and 64 × 64 pixels. Gathering region pictures from CT picture is not a simple errand particularly with settled size of area. To expand the proficiency while safeguarding the high sensitivity, another component is wanted[48]. In this paper, a novel feature known as continuous local histogram (CLH) is presented. CLH coordinates three fundamental sorts of features, which are brightness, texture feature, and shape feature, to build the separation.

4. Proposed method

The proposed method is the new dynamic multi-level CAD framework to automatically identify the defects in the lung CT image. This work also employed the enhanced selective median filter (SMF) to increase the quality of the image with clear view and noise reduction. A new neural network multi-level classifier segmentation method was used on the quality-improved CT image to eliminate the suspicious region. The last process classification was done on the CT scan images using the new neural network-based multi-level classifier with the extracted textural features as shown in Figure 1.

The intention of the extraction is to discover the tumor area on the lung area. In some tumor area with tissues and after the binarization, the tumor area would not be included in the lung area. The image morphology can be used to reduce image quality. This paper demonstrated an innovative multi-level boundary repair procedure to enhance the lung CT images using SMF for enhancing the image quality. The phases of the process are as follows: utilizing the less and more circumstances of erosion and expansion operation with respect to the real images. We can just utilize the less circumstances of erosion and operations to achieve the less time of testing.

5. Implementation steps

The initial process will always be the image processing routine of separation of lung from the other internal organs on the chest CT images and finding the area suspected for the presence of nodule. During first stage, the method extracts square areas of 32 × 32 with the suspicious part at the center. Since it is a multi-level region-based and pixel-based technique the inputs to the system are in the square. The pixel's intensity values falling within the suspicious region that are separated and stored in a database will be used in training the system at the next stages. The training of a new neural network multi-level segmentation process is the second stage. It depends on the two types of input, namely, dynamic feature-based inputs and pixel-based inputs. In dynamic feature-based inputs, the first and second order dynamic features are taken into account. In pixel-based inputs, the pixel's intensity values are used in detecting suspicious region. A SMF is used in pre-processing to enhance the quality of the image by enhancing the poor contrast due to noise and effect due to poor lightning conditions while capturing the image and glare. The generation of low-frequency image is done by placing the median pixel value in every pixel value location. The median value of the pixel is calculated on the square area of 8 × 8 pixels centered at the pixel location. The methods used for enhancing the contrast of the images are sharpening and histogram equalization as shown in Figure 2.

6. Lung region segmentation

Due to the active shape models' availability in the database, lung masks were constructed using them. When segmenting the lung region in CT scans, the user can locate the scope by choosing the questionable locations.

For the purpose of selecting the suspicious zone, a 49 × 49 square mask was created using nodules with a resolution of 96 pixels per inch and a diameter of 13 mm. The image database contains nodules with sizes ranging from 8.9 mm to 29.1 mm, with an average of 17.4 mm. The lung nodule is considered to be between the size of 5mm and 20 mm, and it is detected at the initial stage. The input patterns for the classification stage come from the feature vector from the extraction procedure and the selection stage. The dataset on lung CT images were utilized in training the classifiers and for performance evaluation as shown in Figure 3. Basically, region growing is the conceptually better and the simplest approach for image processing. The segmented region is formed by combining the pixel units of same intensity values in this algorithm. A couple of quantized pixels of the same amplitude will be paired together to form a group called atomic region in the initial phase of the process. The process of combining the weak and combining boundaries between the regions is done in the second evaluation as shown in Figure 4.

Table 1. Geometrical features

Features	Value
Area	2815
Perimeter	226.85
Diameter	59.686
Irregularity index	0.69

6.1. Lung nodule segmentation

In region growing, the procedure called labeling is done to put the negative number representing the label of the region to which the pixel has a place. The labeling process keeps tracks of the record of pixels which are yet to be replaced by the labeling. With this reference list, the operation of insertion and removal will be carried out. Removal is done by eliminating the pixels from the front list, and insertion is by inserting the pixels at the end of the list.

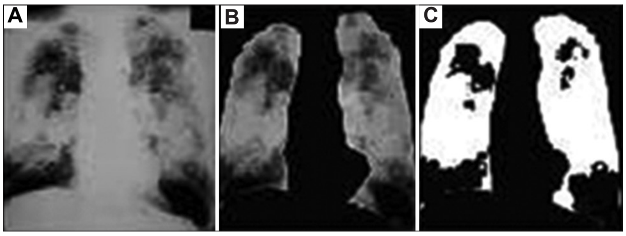


Figure 5. (A) Original image, (B) separated lung fields, and (C) separated cancerous portion

Table 2. Calculated SMF of the filtered images for all types of filters

Filter type	Output SMF
Average filter	7.3133
Weighted average filter	4.2478
Gaussian filter	6.4443
Selective median filter	8.5816
Wavelet filter	8.8732
Wiener filter	13.3969

SMF: Selective median filter

7. Feature extraction

The diameter, area, irregularity index, and perimeters are some of the geometrical features that will be evaluated from the separated lung nodules. The number of pixels in the picture array with the value 1 will be used to determine the size of the segmented tumor image. The area is estimated using the established technique using bit quads, or 2-by-2pixel patterns. The quantity of boundary pixels provides an estimate of the tumor image's perimeter. In terms of its morphology, the tumor has a circular form. The circulatory index will be calculated to identify the irregularity in the circular shape using the expression $I = 4\pi A/p2$, where A is the area of the tumor and P is the perimeter of the tumor in pixels. Table 2 shows the geometrical features for Figure 5. The important features used in the lung cancer classification are texture or contrast features.

The 1st order statistic and 2nd order statistic are the categories under which the contrast features are classified. The shorter processing time and lower minimum cost are the advantages in feature extraction using wavelet due to the solid depiction of the wavelet transform.

8. Classification of lung nodule

Classification is the process of determining whether the nodules belong to a specified class or not. In supervised classification, samples and anticipated classes are known before a classifier is trained on a set of data. It is evident that supervised classification-based methods consistently outperform other types of traditional classification methods[43]. Classification of nodule in the lung is accomplished using

Table 3. Performance metrics of solid feature for various inertia weights

Classifi ers	Accuracy y	Sensiti vity	Specificit y	PPV	NPV	F-measu re	G-mean	MC
DS	88	92	84	85.18	91.30	87.81	87.9	0.76
RF	88	92	84	85.18	91.30	87.81	87.9	0.76
Ada-DS	86	92	80	82.14	90.9	85.58	85.79	0.72
Ada-RF	90	100	80	83.33	100	88.89	89.44	0.81
BPNN	98	96	100	100	96.15	97.95	97.97	0.96

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stump, RF: Random Forest, BPNN: Back Propagation Neural Network

Table 4. Performance metrics of part-solid feature for various inertia weights

Classifie rs	Accura cy	Sensitiv ity	Specifi city	PPV	NPV	F-measur e	G-mean	C
DS	77.6	60.19	95.14	92.53	70.5	73.54	76.17	0.59
RF	86.2	85.4	86.4	86.27	85.57	85.49	85.49	0.72
Ada-DS	87.9	83.5	92.2	91.48	84.82	86.81	87.74	0.61
Ada-RF	86.4	85.4	87.4	91.48	84.82	87.89	86.39	0.76
BPNN	93.68	91.26	96.1	95.91	96.11	93.43	93.46	0.87

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stumpe, RF: Random Forest, BPNN: Back Propagation Neural Network

Table 5. Performance metrics of non-solid feature for various inertia weights

Classifier s	Accura cy	Sensitiv ity	Specific ity	PPV	NPV	F-measur e	G-mean	MC
DS	77.8	100	55.6	69.23	100	71.46	74.56	0.59
RF	91.7	83.3	100	100	85.71	90.88	91.26	0.83
Ada-DS	94.4	88.9	100	100	90	94.12	94.28	0.88
Ada-RF	94.4	100	88.9	88.88	100	94.12	94.28	0.88
BPNN	97.2	100	88.9	94.73	100	97.11	97.15	0.94

PPV: Positive predictive value, NPV: Negative predictive value, DS: Decision Stump, RF: Random Forest, BPNN: Back Propagation Neural Network

different classifiers in this work. Classifiers such as Decision Stump (DS), Random Forest (RF), AdaBoost-Decision Stump, AdaBoost-Random Forest, and Back Propagation Neural Network (BPNN) are used. Classifiers are trained to distinguish the true nodule from false nodule.

8.1. Confusion matrix

A confusion matrix is a chart that is used to describe the classifier's performance on a set of predicted condition for which the actual conditions are known. If the radiologist identifies a patient as disease present and the proposed CAD indicates the presence of disease, the detection test result would yield true positive (TP); if the radiologist identifies a patient as nodule absent and the proposed CAD indicates the presence of nodule, the detection test result would yield FP; if the radiologist identifies a patient as nodule present and the proposed CAD indicates the absence of nodule, the detection test result would yield FN; and if the radiologist identifies a patient as nodule absent and the proposed CAD indicates the absence of nodule, the detection test result would yield true negative (TN). Using TP, FP, FN, and TN, various performance metrics such as classification accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F-measure, and G-mean are calculated.

8.2. Accuracy

Accuracy can be defined as the ratio between the sum of TP and TN to the total sum of attributes used. Accuracy relies mainly on the classification rate of the classifier.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{AN}{TOTAL} \tag{I}$$

8.3. Sensitivity

Sensitivity can be defined as the ratio between TP to the sum of actual positive. It defines how the nodule is correctly diagnosed.

$$\text{Sensitivity} = \frac{TP}{AP} \tag{II}$$

8.4. Specificity

Specificity is the ratio between TN to the sum of actual negative. It defines how well the absence of nodule is correctly diagnosed.

$$\text{Specificity} = \frac{TN}{AN} \tag{III}$$

8.5. Receiver operating characteristics

Receiver operating characteristics (ROC) is graphical representation between FP rate (FPR) and TP rate (TPR). FPR can also be defined as (1-specificity). In ideal situation, the sensitivity and specificity of diagnostic result will be 100% and this is called perfect classification.

8.6. Positive predictive value

The performance of proposed CAD and ground truth should predict correctly the prevalence of disease. Mathematically, positive predictive value (PPV) can be expressed as:

$$PPV = \frac{TP}{PP} \tag{IV}$$

8.7. Negative predictive value

The performance of proposed CAD and ground truth should predict correctly the absence of disease. Mathematically, negative predictive value (NPV) can be expressed as:

$$NPV = \frac{TN}{PN} \tag{V}$$

8.8. F-measure

F-measure can be defined as the weighted mean value of precision and recall.

$$\text{F-measure} = \frac{2^*(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \tag{VI}$$

8.9. G-mean

G-mean maintains a balance between the positive class and negative class classification accuracies. The classification precision of a positive class is defined by sensitivity. In addition, specificity determines how well a negative class is classified. The value nearer to 100% represents the perfect classification accuracy.

$$G-mean=\sqrt{\left(sensitivity*specificity\right)} \qquad \text{(VII)}$$

8.10. Mathew's correlation

In Mathew's correlation (MC), the actual and predicted condition takes the value between 0 and 1. The value of 1 corresponds to perfect correlation, whereas the value of 0.5 corresponds to random prediction.

$$\frac{\left(TP*TN\right)-\left(FP*FN\right)}{\left[\left(TP+FP\right)\left(TP+FN\right)\left(TN+FP\right)\left(TN+FN\right)\right]^{1/2}} \qquad \text{(VIII)}$$
$$=\frac{\left(TP*TN\right)-\left(FP*FN\right)}{\left[AP*AN*PP*PN\right]^{1/2}}$$

9. Results and discussion

The images used for examining the proposed methodology were taken from the LIDC-IDRI, SPIE-AAPM Lung CT challenge, and hospitals. Nodule size between 3 mm and 30 mm were considered in this work. Most specifically, solid, part-solid, and non-solid nodules were chosen. In the LIDC-IDRI database, 71 exams are chosen. Out of 71 exams, 246 nodule case and 240 non-nodule cases were selected. In the SPIE-AAPM database, out of 70 only 35 exams were used. Among them, 28 nodule cases and 34 non-nodule cases are selected. About 36 CT images were acquired from hospitals. A total of 584 images were considered in this work, of which 292 belong to nodule and the rest belong to non-nodule cases. The input datasets are grouped into training set and testing set with 292 datasets each. All these databases were aimed to promote the development of the proposed CAD system.

The performance measure of each classifier for different inertia weights can be measure using accuracy, sensitivity, and specificity. Using confusion matrix, these measures can be calculated. Accuracy of each classifier can be obtained correctly by determining the ratio of the correctly classified and total number of samples. Sensitivity can be measured from the misclassified rate of nodule case to the total number of nodule case used. Specificity can be measured from the misclassified rate of non-nodule case to the total number of non-nodule case used. Tables 1-3 describe briefly the performance of classifiers based on confusion matrix. For solid features, the accuracy, sensitivity, specificity, PPV, NPV, F-measure, G-mean, and MC of various classifiers are noted.

10. Conclusions

In this work, we proposed an automated CAD system for classification of lung nodules using various classifiers from CT images. The classification of nodule and non-nodule patterns in CT is one of the most significant processes during the detection of lung nodule. The developed CAD systems consist of segmentation, feature extraction and classification. For segmentation, we used filters for effective extraction infected region. Later, we extracted features through features and fed into classifiers such as DS, RF, and BPNN. The experimentation was conducted on LIDC-IDRI dataset (Tables 3-5), and the results with BPNN outperformed those with DS and RF classifiers. The performance was measured using sensitivity, specificity, PPV, NPV, F-measure, and G-Mean.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization: Naveen HM
Investigation: Naveen HM
Methodology: Naveen HM
Formal analysis: Naveena C, Manjunath Aradhya VN
Writing – original draft: Naveen HM
Writing – review & editing: Naveen HM

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Akbari R, Ziarati K, 2011, A rank based particle swarm optimization algorithm with dynamic adaptation. *J Comput Appl Math.*, 235(8): 2694–2714.

   https://doi.org/10.1016/j.cam.2010.11.021

2. Aoyama M, Li Q, Katsuragawa S, *et al.*, 2003, Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images.

Med Phys, 30(3): 387–394.

https://doi.org/10.1118/1.1543575

3. Armato SG, Giger ML, Moran CJ, *et al.*, 1999, Computerized detection of pulmonary nodules on CT scans. Radiographics, 19(5): 1303–1311.

https://doi.org/10.1148/radiographics.19.5.g99se181303

4. Armato SG 3rd, Hadjiiski L, Tourassi GD, *et al.*, 2015, Guest editorial: Lungx challenge for computerized lung nodule classification: Reflections and lessons learned. *J Med Imaging*, 2(2): 020103.

https://doi.org/10.1117/1.JMI.2.2.020103

5. Armato SG, McLennan G, Bidaut L, *et al.*, 2011, The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med Phys*, 38(2): 915–931.

https://doi.org/10.1118/1.3528204

6. Arumugam MS, Rao MV, 2006, On the performance of the particle swarm optimization algorithm with various inertia weight variants for computing optimal control of a class of hybrid systems. *Discrete Dyn Nat Soc.*, 2006: 079295.

https://doi.org/10.1155/DDNS/2006/79295

7. Choi WJ, Choi TS, 2014, Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput Methods Programs Biomed*, 113(1): 37–54.

https://doi.org/10.1016/j.cmpb.2013.08.01

8. Criminisi A, Shotton J, Bucciarelli S, 2009, Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI Workshop on Probabilistic Models for Medical Image Analysis. vol. 1. Rochester, Minnesota: MICCAI Society.

9. Cross GR, Jain AK, 1983, Markov random field Texture models. *IEEE Trans Pattern Anal Mach Intell*, 5(1): 25–39.

https://doi.org/10.1109/tpami.1983.4767341

10. Da Silva Sousa JR, Silva AC, de Paiva AC, *et al.*, 2010, Methodology for automatic detection of lung nodules in computerized tomography images. *Comput Methods Programs Biomed*, 98(1): 1–14.

https://doi.org/10.1016/j.cmpb.2009.07.006

11. Dai S, Lu K, Dong J, *et al.*, 2015, A novel approach of lung segmentation on chest CT images using graph cuts. *Neurocomputing*, 168: 799–807.

https://doi.org/10.1016/j.neucom.2015.05.044

12. Daneshmand F, Mehrshad N, Massinaei M, 2013, A new approach for froth image segmentation using fuzzy logic. In: First Iranian Conference on Pattern Recognition and Image Analysis (PRIA). New York City: IEEE.

13. Dawoud A, 2011, Lung segmentation in chest radiographs by fusing shape information in iterative thresholding. *IET Comput Vision*, 5(3): 185–190.

14. De Nunzio G, Tommasi E, Agrusti A, *et al.*, 2011, Automatic lung segmentation in CT images with accurate handling of the Hilar Region. J Digit Imaging, 24(1): 11–27.

https://doi.org/10.1007/s10278-009-9229-1

15. Deep G, Kaur L, Gupta S, 2013, Lung nodule segmentation in CT images using rotation invariant local binary pattern. *Int J Signal Image Process*, 4(1): 20.

16. Dehmeshki J, Amin H, Valdivieso M, *et al.*, 2008, Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach. *IEEE Trans Med Imaging*, 27(4): 467–480.

https://doi.org/10.1109/TMI.2007.907555

17. Dehmeshki J, Ye X, Lin X, *et al.*, 2007, Automated detection of lung nodules in CT images using shape-based genetic algorithm. *Computer Med Imaging Graph*, 31(6): 408–417.

https://doi.org/10.1016/j.compmedimag.2007.03.002

18. Delogu P, Cheran S, De Mitri I, *et al.*, 2005, Preprocessing methods for nodule detection in lung CT. In: International Congress Series. vol. 1281. Netherlands: Elsevier.

19. Dheepak G, Premkumar S, Ramachandran R, 2015, Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Method. *Int J Power Control Comput,* 7: 24–28.

20. Doi K, 2007, Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computer Med Imaging Graph*, 31(4): 198–211.

21. Dolejsi M, Kybic J, Polovincak M, *et al.*, 2009, The lung time: Annotated lung nodule dataset and nodule detection framework. In: SPIE Medical Imaging. Washington USA: International Society for Optics and Photonics.

22. Elizabeth D, Nehemiah H, Raj CR, *et al.*, 2012, Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image. *IET Image Processing*, 6(6): 697–705.

23. Enquobahrie AA, Reeves AP, Yankelevitz DF, *et al.*, 2007, Automated detection of small pulmonary nodules in whole lung CT scans. *Acad Radiol*, 14(5): 579–593.

24. Farag A, Abdelmunim H, Graham J, *et al.*, 2012, An AAM based detection approach of lung nodules from LDCT scans. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI). New York City: IEEE.

25. Farag A, Ali A, Graham J, *et al.*, 2011, Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. New York City: IEEE.

26. Farag AA, Abdelmunim H, Graham J, *et al.*, 2011b, Variational approach for segmentation of lung nodules. In:

IEEE International Conference on Image Processing (ICIP). New York City: IEEE.

27. Gambhir S, Shepherd J, Shah B, *et al.*, 1998, Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol*, 16(6): 2113–2125.
28. Garro BA, Vazquez RA, 2015, Designing artificial neural networks using particle swarm optimization algorithms. *Comput Intell Neurosci*, 2015: 369298.

    https://doi.org/10.1155/2015/369298
29. Golosio B, Masala GL, Piccioli A, 2009, A novel multithreshold method for nodule detection in lung CT. *Med Phys*, 36(8): 3607–3618.

    https://doi.org/10.1118/1.3160107
30. Gomathi M, Thangaraj P, 2010a, A computer aided diagnosis system for detection of lung cancer nodules using extreme learning machine. *Int J Eng Sci Technol*, 2(10), 5770–5779.
31. Gomathi M, Thangaraj P, 2010b, A computer aided diagnosis system for lung cancer detection using support vector machine. *Am J Appl Sci*, 7(12): 1532.

    https://doi.org/10.3844/ajassp.2010.1532.1538
32. Alves LG, Novo J, Campilho A, 2016, Hessian based approaches for 3d lung nodule segmentation. *Expert Syst Appl*, 61: 1–15.

    https://doi.org/10.1016/j.eswa.2016.05.024
33. Gould MK, Donington J, Lynch WR, *et al.*, 2013, Evaluation of individuals with pulmonary nodules: When is it lung cancer. *Chest*, 143(5 Suppl): 93S–120S.

    https://doi.org/10.1378/chest.12-2351
34. Grigorescu SE, Petkov N, Kruizinga P, 2002, Comparison of texture features based on Gabor filters. *IEEE Trans Image Process*, 11(10): 1160–1167.

    https://doi.org/10.1109/TIP.2002.804262
35. Gu Y, Kumar V, Hall LO, *et al.*, 2013, Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recogn*, 46(3): 692–702.
36. Gudise VG, Venayagamoorthy VK, 2003, Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. In: Swarm Intelligence Symposium. New York City: IEEE.
37. Hua P, Song, M, Sonka EA, *et al.*, 2011, Segmentation of pathological and diseased lung tissue in CT images using a graph-search algorithm. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. New York City: IEEE.
38. Jacobs C, Murphy K, Twellmann T, *et al.*, 2011, Computer-aided detection of solid and ground glass nodules in thoracic CT images using two independent cad systems. In: The Fourth International Workshop on Pulmonary Image Analysis.
39. Shen S, Bui AA, Cong J, *et al.*, 2015, An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Comput Biol Med*, 57: 139–149.

    https://doi.org/10.1016/j.compbiomed.2014.12.008
40. Shen S, Sandham W, Granat M, *et al.*, 2005, MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Trans Inform Technol Biomed*, 9(3): 459–467.

    https://doi.org/10.1109/titb.2005.847500.
41. Shi, Y, Eberhart R, 1998, A modified particle swarm optimizer. In Evolutionary Computation Proceedings, 1998. In: IEEE World Congress on Computational Intelligence, The 1998 IEEE International Conference. New York City: IEEE.
42. Shih-Chung BL, Freedman MT, Lin JS, Mun SK, 1993, Automatic lung nodule detection using profile matching and back-propagation neural network techniques. *J Digit Imaging*, 6(1): 48–54.
43. Yuan J, 2013, Active contour driven by local divergence energies for ultrasound image segmentation. *IET Image Processing*, 7(3): 252–259.

    https://doi.org/10.1049/iet-ipr.2012.0120
44. Zhou S, Cheng Y, Tamura S, 2014, Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images. *Biomed Signal Process Control*, 13: 62–67.

    https://doi.org/10.1016/j.bspc.2014.03.010
45. Flohr TG, Schaller S, Stierstorfer K, *et al.*, 2005, Multi-detector row CT systems and image-reconstruction techniques. *Radiology*, 235(3): 756–773.

    https://doi.org/10.1148/radiol.2353040037
46. Kawane K, Fukuyama H, Yoshida H, *et al.*, 2003, Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nat Immunol*, 4(2): 138–144.

    https://doi.org/10.1038/ni881
47. Shi J, Malik J, 2000, Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Machine Intell*, 22(8): 888–905.

    https://doi.org/10.1109/34.868688
48. Bercoff J, Tanter M, Fink M, 2004, Supersonic shear imaging: A new technique for soft tissue elasticity mapping. *IEEE Trans Ultrason Ferroelectr Freq Control*, 51(4): 396–409.

    https://doi.org/10.1109/tuffc.2004.1295425

Tumor Discovery

ORIGINAL RESEARCH ARTICLE

Malignant versus normal breast tissue: Optical differentiation exploiting hyperspectral imaging system

Mohamed Hisham Aref[1]*, Ibrahim H. Aboughaleb[1], Abdallah Abdelkader Hussein[2], Ayman Mohammed Farag[3], Sara Abd El-Ghaffar[4], and Yasser H. El-Sharkawy[5]

[1]Biomedical Engineering Researcher, Egyptian armed forces, Cairo, Egypt

[2]Department of Histopathology, "Kobri El-Koba" Complex Hospital, Egyptian Armed Forces, Cairo, Egypt

[3]Department of Radiology, Military Medical Academy, Cairo, Egypt

[4]Department of Radiology, Maadi Armed Forces Medical Compound, Cairo, Egypt

[5]Department of Optoelectronics, Military Technical College, Cairo, Egypt

Abstract

Breast malignancy is a critical problem that severely affects women's health globally with a high-frequency rate, necessitating fast, effective, and early diagnostic methods. The present study aims to measure the breast tissue's optical properties by capturing the spectral signatures from malignant and normal breast tissue for therapeutic and diagnostic applications. The optical imaging system incorporates a hyperspectral (HS) camera to capture the spectral signatures for both the malignant and normal breast tissues within 400 ~ 1000 nm. The system was subdivided into two exploratory (reflection/transmission) to measure the tissue's diffuse reflectance (R_d) and light transmission (T_r), respectively. The study involved 30 breast tissue (normal/tumor) samples from 30 females in the age range of 46 ~ 72 years, who were optically inspected in the visible and near-infrared (VIS-NIR) spectra. Then, the inverse adding doubling (IAD) method for breast tissue characterization and descriptive analysis (T-test) was exploited to verify the significant difference between the various types of breast tissues and select the optimum wavelength. Finally, comparing the study outcome with the histopathological examination to evaluate the system's effectiveness by calculation (sensitivity, specificity, and accuracy). The average outcome values demonstrated that the optimal spectral bands distinguishing between the normal and the tumor tissues regarding the reflectance approach were 600 ~ 680 nm and 750 ~ 960 nm at the VIS and NIR spectrum, respectively. Then, for the transmission technique, the optimal spectral bands were 560 ~ 590 nm and 760 ~ 810 nm at the VIS and NIR spectra, respectively. Later, the T-test and the IAD verified that the highest R_d values for discrimination were 600 ~ 640 nm and 800 ~ 840 nm at the VIS and NIR spectra, respectively. On the other side, the highest T_r values were 600 ~ 640 nm and 760 ~ 800 nm at the VIS and NIR spectra, respectively. The investigation's average reading accuracy, sensitivity, and specificity were 85%, 81.88%, and 88.8%, respectively. The experimental trials revealed that the system could identify the optimal wavelength for therapeutic and diagnostic applications through the light interaction behavior of the breast tissue's optical properties.

Keywords: Breast cancer detection; Hyperspectral imaging system; Tissue characterization; Breast biopsy; Optical properties; Visible and near-infrared spectroscopy

***Corresponding author:**
Mohamed Hisham Aref
(MH-Aref@ieee.org)

Citation: Aref MH, Aboughaleb IH, Hussein AA, *et al.*, 2023, Malignant versus normal breast tissue: Optical differentiation exploiting hyperspectral imaging system. *Tumor Discov*, 2(1): 258. https://doi.org/10.36922/td.258

Received: November 15, 2022
Accepted: February 21, 2023
Published Online: March 21, 2023

1. Introduction

Breast malignancy (BM) is a threatening disease in women[1,2]. BM has a high occurrence and mortality rates globally[2]. In line with the World Health Organization's recommendation, it is vital to have an effective healthcare system for BM diagnosis and therapy[3]. Of all women's tumors, 16% are diagnosed with breast cancer[4]. It is the primary source of malignancy-based death in Africa, with an elevated rate in low- and middle-income nations[5,6]. Even though the sign of cancer is not selective to females, the incidence rate of breast cancer in women is multiple times higher than that in men[7]. The malignant-based death, in one way, is credited to late diagnosis as the malignancy has grown in size or metastasized. Consequently, early diagnosis is an excellent approach to improving the survival rate, which can go up to 95%[8].

Early detection of BM was the objective of multiple studies involving imaging modalities[9], such as mammograms[10], ultrasound (US)[11], and breast magnetic resonance imaging (BMRI)[12]. On the other hand, it is limited to several components, for example, the contrast noise ratio[13], longitudinal resolution, and signal-to-noise ratio of every machine[14]. Researchers have investigated numerous breast cancer diagnostic techniques, incorporating X-ray mammogram, MRI, US, positron emission tomography scan, computerized tomography (CT), and tissue removal (biopsy), which are detailed in Table S1 (Supplementary File), for the regular breast cancer diagnostic approaches and their constraints[15].

At present, tissue sampling is the standard method for BM identification. However, it requires acquiring a tumor biopsy and then an investigation by a pathologist[16]. In addition, tissue sampling involves staining, sample cutting, and microscopic investigations, which are time-consuming and costly[17].

None of the abovementioned imaging modalities is flawless for intraoperative resection of malignant tumors[18]. For example, 37% of the women had breast apportioning, and the tumor was allocated at the edge of the biopsy[19]. In oncology, failure to completely eradicate malignant develops cancer recurrence which leads to the spread of the disease[20]. Presently, a pathologist, who investigates the tissue with a digital microscope, surveys the resection edge a couple of days after an operation and following this, the pathologist can give a direct examination before commencing the treatment[21].

The US Food and Drug Administration approved the whole slide imaging (WSI) for the fundamental analysis of histopathological slides[22,23]. Computerized radiology enables immediate investigation of WSI in routine pathology practice. However, there are critical contrasts in motivations for appropriating computerized innovation in pathology when contrasted with radiology. In digital radiology, sensors can straightforwardly catch the data from imaging sources. This advanced cycle obviates the need for ordinary films, harmful compound preparation, and X-ray record rooms. By contrast, slide imaging for pathology necessitates that the tissues are prepared in the typical way (embedded in paraffin, cut, set on glass slides, and stained). Afterward, the pathologist investigates the samples visually thru the microscope, which often consumes time and effort and relies on clinician experience[24].

Several studies compared the frozen section analysis and the imprint cytology involving hematoxylin and eosin (H&E) staining examination alone. *Motomura et al.* reported that compared to H&E, imprint cytology analysis has an overall accuracy, sensitivity, and the specificity of 96%, 90.9%, and 98.5%, respectively[25]. Flett *et al.* stated that frozen section analysis accurately predicts axillary node status with 95%[26]. Moreover, Van Diest *et al.* stated that frozen section examinations were better than imprint cytology analysis (sensitivity 91% versus 63%), and specificity for both methods was 100%[27].

Hyperspectral imaging (HSI) is also called imaging spectrometry and chemical imaging[28]. HSI has advantages over multispectral and RGB imaging because it captures tens to hundreds (continuous spectrum) along the electromagnetic spectrum. HSI collects spatial information (x, y), and a spectral wavelength called a data cube, with each pixel, provides a specific spectral signature based on the reflection, transmission, and absorption of electromagnetic radiation for each material under examination[29]. HSI coordinates ordinary imaging and spectroscopy modalities to provide spatial and spectral information about the object of interest[30,31]. Spectrometry imaging had been exploited in land scanning remotely since 1985[32].

As an imaging methodology for clinical applications, HSI offers extraordinary potential for noninvasive diagnostic and surgical navigation[33]. Several studies highlight the novelty of the HSI in medical applications, *For example,* Fabelo *et al.* used an intraoperative imaging system utilizing HSI to assist in brain tumor delineation to differentiate between normal and cancer tissue in the brain during a neurosurgical operation[34]. In addition, Goto *et al.* implemented a study to recognize gastric tumors in *ex vivo* human tissues[35]. Furthermore, *Regeling et al.* utilized flexible endoscopy combined with an HSI system to discriminate laryngeal cancer to acquire HS cubes in the region within 390 ~ 680 nm[36]. Moreover, HSI had been exploited in monitoring the thermal ablation in the biological tissue[37,38], sketching of the blood vessels in the arm to aid with Phlebotomy and tissue oxygen[39,40], and breast cancer investigation and malignancy detection[41-43].

The expansion of optical systems in current medical sectors in therapeutic, diagnosis, and surgery regions has motivated the research on optical properties of numerous biological tissues. At the same time, the effectiveness of laser therapy varies by photon propagation and spreading fluence rate inside irradiated tissues[44]. Moreover, the ability to image a biological sample deeply is limited by light penetration depth inside the biological tissues, as distinguished by high turbidity[45]. Regarding the optical properties of the biological tissues (comprising blood, lymph, and other biological fluids), it had been classified into two classes: (i) opaque tissues (intensely scattering) such as the brain, skin, blood, and vascular walls; and (ii) translucent tissues (inadequately scattering) such as the cornea and anterior eye chamber lens[46]. The light interaction (reflection, scattering, and absorption) with the investigated biological soft tissue varies concerning the optical properties variation of its fundamental characteristics, is presented in Figure S1 (Supplementary File).

In this study, we designed an optical imaging system incorporating the hyperspectral (HS) camera to acquire a fast and effective method for breast tissue characterization by capturing the spectral signatures of the malignant and normal breast tissues for both investigative and therapeutic objectives. The exploited optical imaging had been divided into two separate setups (Reflection/Transmission) with spectral range of 380 – 1050 nm to measure the tissue's diffuse reflectance (R_d) and light transmission (T_r), and then the sample absorption coefficient (μ_a) was calculated from T_r. Then, from the measurements of the previously stated parameters for both the normal and the malignant breast tissues, we exploited the inverse adding doubling (IAD) method for breast tissue characterization. Furthermore, the T-test was utilized to verify the significant difference between the various types of breast tissues and select the optimum wavelength for diagnosis and therapy applications. Finally, the proposed methods with histopathological examination were compared to evaluate the system's effectiveness in terms of sensitivity, specificity, and accuracy.

2. Materials and methods

2.1. Primary system interconnections

The primary system interconnections include the following:

- The design and implementation of the optical imaging system
- The optical phantoms and system calibration
- The investigation and patient criteria for breast tissue samples selection and preparation
- Capturing the HS image for the *ex vivo* breast samples
- The measurement of the sample's diffuse reflection (R_d) for both the cancerous and the non-cancerous regions
- The measurements for sample T_r
- Calculating the sample absorption coefficient (μ_a) from the measured T_r
- The statistical analysis to select the optimum wavelength for the diagnostic and therapeutic applications
- Calculating the system efficiency (average reading accuracy, sensitivity, and specificity).

2.2. Design and implementation of the optical imaging system

The principal structure of the proposed framework is partitioned into two different configurations. The first configuration (reflection approach) utilizes a polychromatic source light (Derungs, 20P SX -20 Watt, Germany) with a spectral range (400 ~ 950 nm) to measure the R_d of the investigated *ex vivo* breast samples, as illustrated in Figure 1A for the schematic diagram and Figure 1B for the actual setup[47]. The second configuration (transmission approach) employs the same light source. However, underneath the investigated samples for light T_r measurement, these measurements yield the calculation of the μ_a, as presented in Figure 1C for the schematic diagram and Figure 1D for the actual setup.

Both configurations exploit the HS camera (Surface Optics, SOC710, USA) at 400 ~ 1000 nm, with a spectral resolution of 4.69 nm and a bit depth of 12, which is equipped with an objective lens (Schneider, 400 – 1000 nm, Germany). The employed HS camera is a push broom imager with scanned cube 128 frames which has a built-in translation sensor capable of directly collecting information for the entire spatial image of the whole object. The camera was settled at a height of 20 cm, and the light source was 16 cm from the breast samples. The light was settled at the same distance under the samples in the transmission configuration. The signal analysis measurements were analyzed with software (SOC's Hyperscanner and SRAnalysis, USA) accompanied by (DADiSP, SE 6.7, USA) on a personal laptop (DELL, INSPIRON 5584, Intel Core I7, 16 GB RAM, Windows 10, USA) where the actual setup with all of its components is displayed in Figure 1.

2.3. Optical phantom preparation and system calibration

Initially, we prepared liquid optical phantoms for system calibration. Then, we used demineralized water as a matrix material and added milk (whole milk, lactose-free, and fat-free) as the scattering material with three different

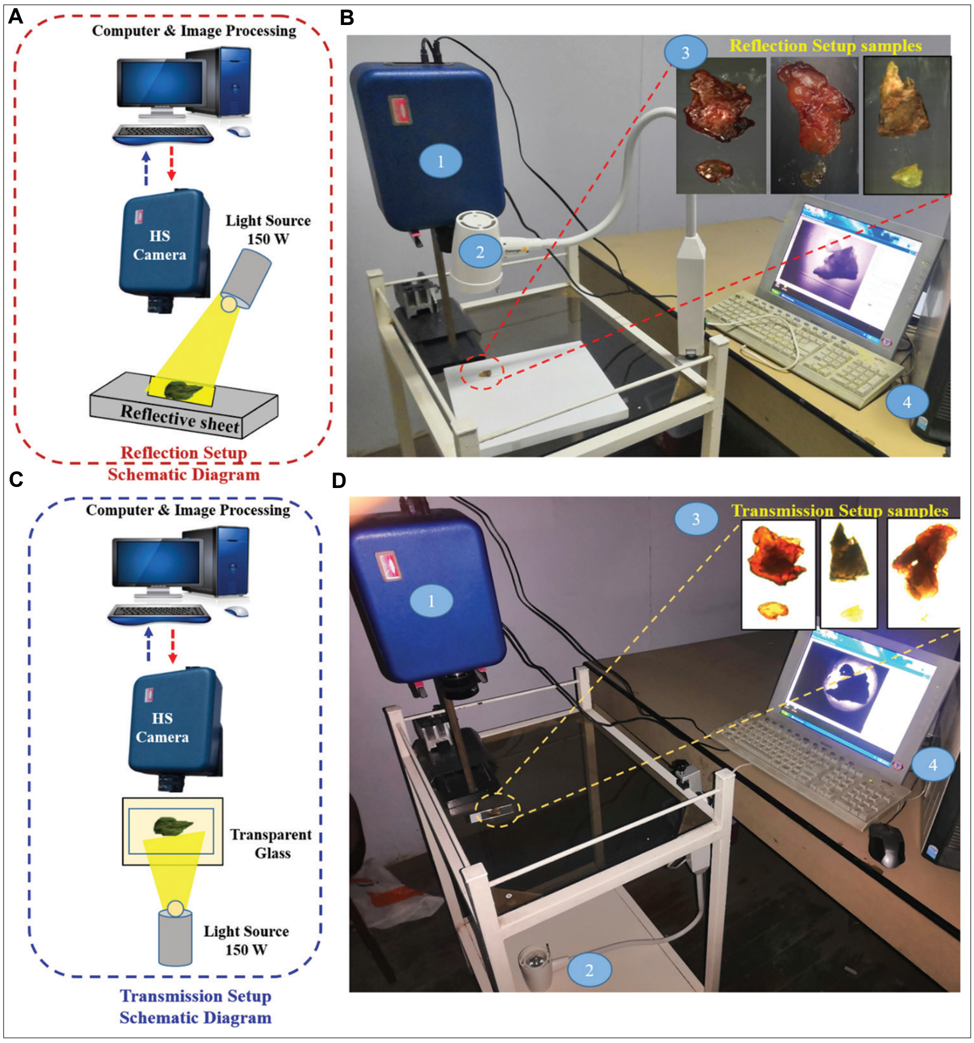


Figure 1. (A) The schematic diagram of the reflection setup. (B) The actual reflection setup to measure the tissue's diffuse reflectance (R_d): (1) the HS camera, (2) the polychromatic source light, (3) the investigated breast samples, and (4) the computer and the image software analysis. (C) The schematic diagram of the transmission setup. (D) The actual transmission setup to measure the tissue's light transmission (T_r).

fat concentrations (3%, 1.75%, and 0%). Samples were arranged in a 300 mL beaker, mixing the milk with the water to achieve various lipid concentrations (0.1% ~ 1%, with increments of 0.05%). Moreover, the exploited absorption material was a red Indian ink (Speedball, Statesville, USA) with whole fat milk (3% of fat concentration). Mixture was prepared by adding 200 mL from the ink with 0.2 mL of milk at concentration range (0.005% ~ 0.2%). The ink concentration was constant at 0.1% with 200 µL dilution in 100 mL of demineralized water. Milk was added to the mixture with a syringe of 5 mL of lipid at various concentrations (0.1% ~ 0.5%)[48,49].

2.4. Patient criteria and sample preparation

The current trial examination was approved by the Institutional Ethical Committee. In addition, all patients read and signed the consent form before data collection began. The study was conducted from May 2020 to October 2021 with a total of 30 female patients who were diagnosed with breast cancer by two different imaging methods (mammogram/US) and had underwent mastectomy; the patient data for the present study are summarized in Table S2 (Supplementary File).

The breast tumor samples, which had been classified by the pathologists, were processed at the pathological center and then attached with the complete histology lab report for each patient. The breast samples were cut by the pathologist into various sizes around 10 ~ 12 cm^2 with a thickness of 4 ~ 5 mm. One of the team members transported samples in an icebox. During the experimental investigation, samples were dipped in phosphate-buffered saline (pH = 7.4) to remove blood. After scanning, the samples were stored in a freezer at a temperature of −75°C. The experimental trials were conducted in a lab with room temperature of ~25°C.

2.5. Principal theory and procedure equations

The proliferation of incident light inside tissue is a significant issue in clinical applications and the improvement of diagnostic techniques. This way, this segment is committed to a concise audit of the light tissue collaboration systems, optical cycles included in HSI, and valuable diagnostic and therapeutic data provided by HSI[33]. Light entering biological breast tissue goes through multiple scattering and absorption as it proliferates across the tissue[50]. Biological tissues are assorted in composition through spectral distinctions in optical properties. Scattering appears where there is a spectral distinction in the refractive index[51].

The diffusion profundity of light into biological tissues depends on how unequivocally the tissue absorbs light. Most tissues are adequately powerless absorbers to allow substantial light diffusion inside the therapeutic window, going from 600 to 1300 nm. Inside the therapeutic window, scattering is higher than absorption, so the spreading light gets diffuse[50,51]. The primary block diagram of the two applied system approaches (reflection and transmission) is illustrated in Figure 2.

Light proliferation in investigated tissue depends on the transport hypothesis[52,53]. Transport theory depends on the superposition of energy flux, so the wave properties of light (polarization and interference) are not considered in transport theory. Where the radiant power of the light transferred to the surface is displayed in Equation I:

$$\text{Ŗ} = \int \text{Ƒ}.\text{n}\, dA \tag{I}$$

Where (Ƒ) is the flux vector, (Ŗ) is the radiant power transferred through a surface with the area (A).

As the surface of the biological tissue is not homogeneous leading to light proliferation. However, it is crucial to understand a few of the significant optical parameters which are exploited in modeling of the light proliferation, such as the propagation of photons, fluence ratio, radiance, and flux[54].

The photon allocation function $\text{Ŋ}(\text{ŗ},\hat{\text{ş}})$ is defined as the number of photons for each unit volume moving in the course of a unit vector $\widehat{\text{V}}$, in the component of fixed angle incorporating $\widehat{\text{V}}$ at a specified spot ŗ divided by this component. The power of photons (β) that proliferate through minute area $d\text{Ă}$ in the minute fixed angle (dw) in the course of $\widehat{\text{V}}$, with energy hν and speed Ś_t is shown in Equation II:

$$d\beta(\text{ŗ},\hat{\text{ş}})[W] = \text{Ŋ}(\text{ŗ},\hat{\text{ş}})\; d\text{Ą}\; d\text{w}\; \text{Ś}_t\; \text{h}\nu \tag{II}$$

Where (Ś_t) is the speed of light in tissue and $(d\text{Ă})$ is perpendicular to $\widehat{\text{V}}$.

Very often, the operating mechanism of medical techniques exploits the relations of light spreading through tissue. The quantity of light could be stated as the irradiance $\check{\text{Ȩ}}_0$, which is the radiant energy flux of the surface element divided by surface area. Part of the incident light is reflected, and the others are attenuated with the tissue by diffuse reflection and absorption according to the Beer's law, as shown in Equations III and IV.

$$\omega(\underset{\wedge}{\text{Ļ}}) = \check{\text{Ȩ}}_0\; (1 - \text{Ŗ})\left(e^{(-(\mu_a - \mu_s)\underset{\wedge}{\text{Ļ}})}\right) \tag{III}$$

Where $\omega\ (\underset{\wedge}{\text{Ļ}})$ is the fluence ratio for the unscattered beam at location $(\underset{\wedge}{\text{Ļ}})$ $\check{\text{Ȩ}}_0$ is the irradiance, and Ŗ is the surface reflection (Fresnel).

$$\psi = \frac{1}{(\ddot{\text{R}}\text{d} + \mu_a)} = \frac{1}{\mu_t} \tag{IV}$$

Where R_d is the scattering (diffuse reflection) coefficient, μ_a is the absorption coefficient, μ_t is the total attenuation coefficient, and ψ is the penetration depth.

While light travels within the tissue, its intensity gradually weakens, in a phenomenon known as light absorption and expressed by μ_a, which is described as the probability of photon absorption after being proliferated per unit length. The light absorption follows the Lambert-Beer law. Therefore, when there are only light absorption phenomena of tissue, it could be expressed by Equation V. Additionally, the optical homogenous scattering phenomena follows Lambert-Beer law and could also be expressed by Equation VI[55].

$$I = I_{0\, e^{-\mu_a d}} \tag{V}$$

$$I = I_{0\, e^{-\mu_s d}} \tag{VI}$$

Where μ_a is known as the absorption coefficient, I_0 is the incident light, I is the light intensity after passage through the medium or tissue, and (μ_s) is the scattering coefficient.

The incident light beam interaction with the biological tissues is evaluated in terms of T_r, R_d, and calculated absorption coefficient (μ_a)[54], using Equations VII and VIII:

$$\text{Tr} = \frac{\text{Ţ}_{\ddot{\text{R}}\text{d}} - \text{Ţ}_0}{\text{Ţ}_{\ddot{\text{R}}\text{d}t} - \text{Ţ}_0} \tag{VII}$$

$$\mu_a = \frac{\mu_{a\ddot{\text{R}}\text{d}} - \mu_{a_0}}{\mu_{a\ddot{\text{R}}\text{d}t} - \mu_{a_0}} \tag{VIII}$$

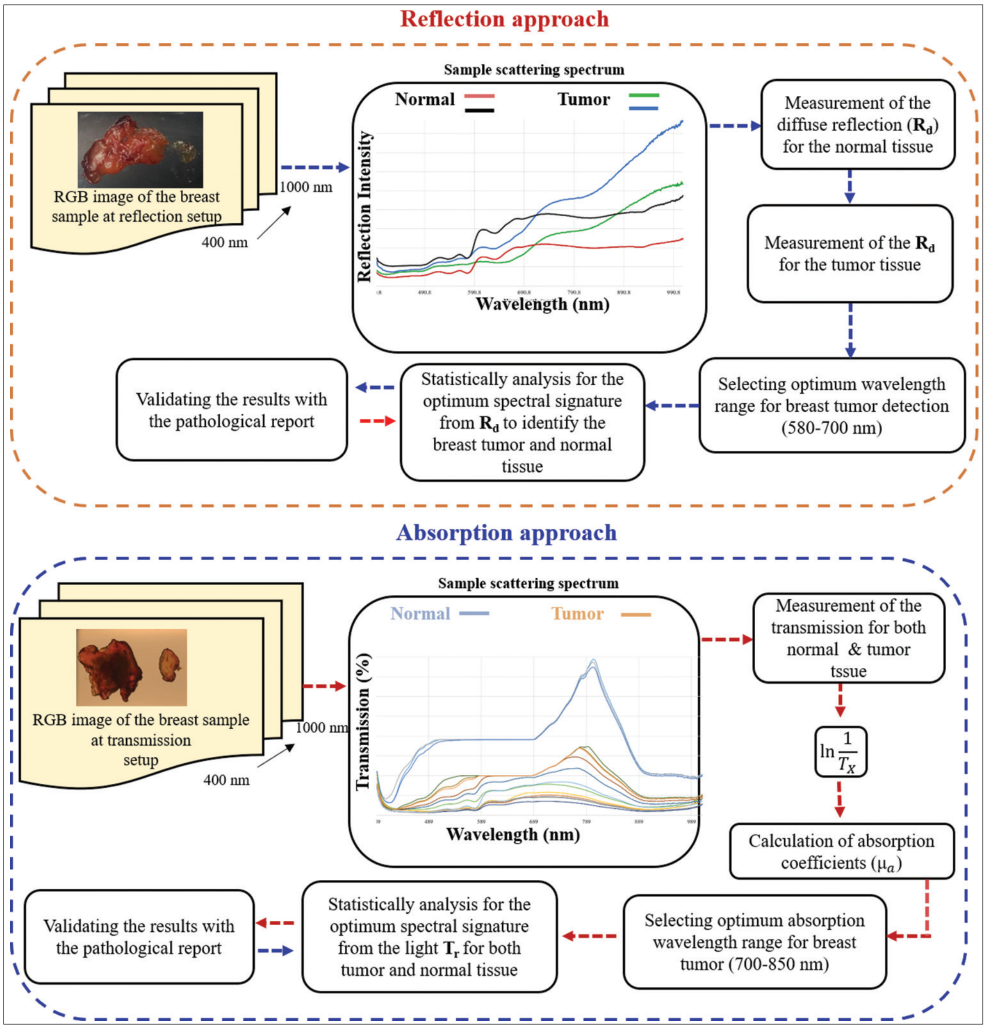


Figure 2. The main block diagram of the two applied system approaches (reflection/transmission) to measure the tissue's diffuse reflectance (R_d) and light transmission (T_r), which highlight the signal analysis for the optical properties and capture the spectral signatures of the investigated *ex vivo* breast tissue samples.

Where $\frac{\mu a \ddot{\mathrm{R}}\mathrm{d}}{\mathrm{T}s}$ and $\frac{\mu a \ddot{\mathrm{R}}\mathrm{d}t}{\mathrm{T}st}$ are the measured light intensity by R_d/T_r with the investigated breast tissue and the typical white plate reflection, respectively. The $\frac{\mu a_0}{\mathrm{T}_0}$ demonstrates the background light intensity discriminated by R_d/T_r without the sample reflection on the plate.

2.6. Spectral arrangement

The hyper spectral images were taken for every investigated breast tissue sample from the two different arrangements as explained previously with spatial pixels (max/nominal) = 1040/520, spectral channels = 128, HS cube captured 100 frames/s within 6.96 s/cube. They applied statistical analysis for the spectral signature for selecting the optimum spectral range to discriminate the normal from the abnormal tissue.

It is essential to capture two data cubes each time a breast sample is imaged: the first one is the dark data cube which represents the image sensor's dark current, and the second one is for a standard white reference for spectral calibration[56]. First, the acquired dark cube was taken by closing the lens of the HS camera with its cap. The data of the dark and white cubes were exploited to adjust the reflected sample image to remove the noise impacts of the investigated sample tissue, as expressed in Equation IX:

$$\ddot{\mathrm{R}}_f(\vartheta) = \frac{\tilde{\mathrm{I}}_m(\vartheta)-\tilde{\mathrm{I}}_d(\vartheta)}{\tilde{\mathrm{I}}_w(\vartheta)-\tilde{\mathrm{I}}_d(\vartheta)} \times 100\% \qquad \text{(IX)}$$

Where $\ddot{\mathrm{R}}_f(\vartheta)$ is the relative reflectance of the investigated sample image, $\tilde{\mathrm{I}}_m(\vartheta)$ is the seized image, $\tilde{\mathrm{I}}_d(\vartheta)$ is the opaque image cube, and $\tilde{\mathrm{I}}_w(\vartheta)$ is the attained image of the white reflectance board.

2.7. Statistical analysis

So far, the commonly used methods to calculate the scattering and absorption properties of the various mediums are:

(i) Monte Carlo (MC), which is a general class of computational algorithms to achieve a numerical outcome by relying on random sampling. MC in the optical field is effective for a broad range of light μa, μs, and photon paths[57].

(ii) Diffuse approximation (DA), which is an alternative calculation method to calculate the scattering and absorption properties of turbid mediums[58].

(iii) IAD, which is an extended method from the adding double (AD) method, and exploited to solve the radiative transport equation in the optical field sector related to the light's interaction with the tissues in a slab geometry[59].

The radiative transport equation could be used to achieve light intensity distribution for the physical mediums, as shown in Equation X[60].

$$\frac{dI(r,s)}{ds} = -(\mu_a + \mu_s)I(r,s) + \frac{\mu_s}{4\pi}\int_{4\pi} p(s,s')I(r,s')d\theta \tag{X}$$

Where $I\,(r,s)$ is the measured intensity per unit length, r is the target location, s is the unit direction vector, $p(s,s')$ is the phase function, and θ is the solid angle.

Although there is no analytical solution for Equation X, it is achievable by exploiting the MC technique[61,62]. Moreover, the IAD method is used to solve the radiative transport equation. IAD technique and MC model have offered more precise approximations of optical properties for the biological tissue (μ_a, μ_s, g) better than other methods. Two dimensionless quantities are exploited in the whole process of the IAD, that is, the albedo (a) and the optical depth ($\mathcal{T}$), which are well-defined in Equations XI and XII:

$$a = \frac{\mu_s}{\mu_s + \mu_a} \tag{XI}$$

$$\mathcal{T} = t(\mu_s + \mu_a) \tag{XII}$$

Where is the sample's thickness (mm), the measured values of R_d, the total diffuse transmittance (T_d), and the unscattered collimated transmittance (T_c) are applied to the IAD process to calculate the (μ_a, μ_s).

Due to the minimum computational time and high accuracy comparable to both DA and MC methods, we exploited in our system the IAD method for breast tissue characterization and descriptive analysis (T-test) to verify the significant differences among the various types of breast tissues and to select the optimum wavelength.

2.8. System efficiency analysis

The efficiency of the presented system analysis methods is achieved by comparing the outcomes with the histological investigations. Regarding the results of these comparisons, three numerical values (sensitivity, specificity, and accuracy) could be measured to evaluate the various spectral analysis methods, as shown in Equations XIII, XIV, and XV:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{XIII}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{XIV}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total no. of samples}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{XV}$$

Where true positive (TP) is the cases detected by the proposed system as actual masses (tumors); false negative (FN) is the cases of the system which had not been detected and have masses; true negative (TN) is the cases detected by the proposed system as normal, and they are normal cases; and false positive (FP) is the cases mistakenly detected by the presented system as abnormal masses, and they are normal cases.

3. Results

Our primary goal of these assessments is to investigate the optical properties of the *ex vivo* breast samples (normal/tumor) by identifying the spectral signatures through incorporating the HS camera capabilities to provide the essential data for diagnostic and therapeutic applications concerning breast cancer. We set up two diverse frameworks (reflection/transmission) methods using the HS camera at wavelength range (400 ~ 1000 nm) and with a polychromatic light source in the VIS-NIR range for this examination.

We initially exploited the first framework (reflection method) to measure the investigated *ex vivo* breast tissue sample's light R_d; one of the investigated cases is presented in Figure 3. The solid red line represents the measured tumor tissue's R_d spectrum, and the solid blue line identifies the normal tissue's R_d spectrum over the VIS-NIR range. From the measured light R_d for the investigated samples, we could visually highlight the spectrum peaks, which distinguish between the normal and the tumor tissues at wavelength range 600 ~ 680 nm and 750 ~ 960 nm at the VIS range and NIR range, respectively.

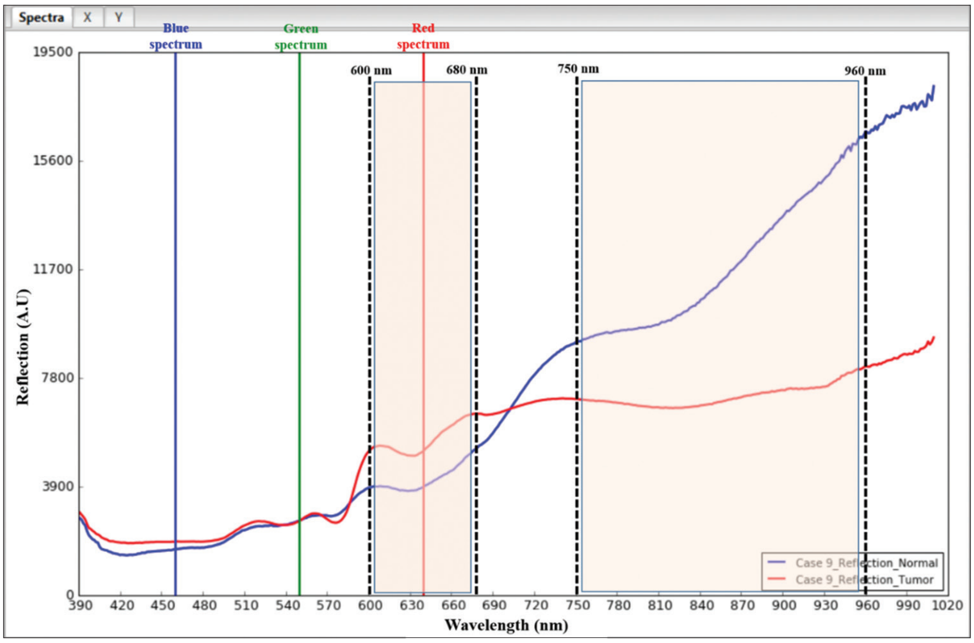


Figure 3. The measured light R_d spectrum for one of the investigated samples (patient ID 1009). The solid red line is for the measured tumor tissue's light R_d spectrum, and the solid blue line identifies the measured normal tissue's R_d spectrum highlighting the peaks which could visually identify between both normal and tumor tissues at wavelength ranges of 600 ~ 680 nm and 750 ~ 960 nm at the visible and near-infrared spectra, respectively.

Moreover, regarding the second framework (transmission method) to measure the investigated *ex vivo* breast tissue sample's light T_r, one of the investigated cases is presented in Figure 4. The solid red line represents the measured tumor tissue's T_r spectrum, and the solid blue line identifies the normal tissue's T_r spectrum over the VIS-NIR range. From the measured light T_r for the investigated samples, we could visually highlight the spectrum peaks that distinguish between the normal and the tumor tissues at wavelength range 560 ~ 590 nm and 760 ~ 810 nm at the VIS range and NIR range, respectively.

From the two frameworks, we could measure the tissue's R_d and light T_r, and then calculate the sample μ_a from the measured T_r. Then, from the measurements of the previously stated parameters for both the normal and the malignant breast tissues, we could identify a spectral signature for each tissue type by investigating the optical spectroscopy in the VIS-NIR range for the measured light R_d and T_r. The measured tissue's R_d, T_r, and calculated μ_a of the investigated *ex vivo* breast samples are illustrated in Figure S2A-C (Supplementary File), respectively.

Later, we exploited the IAD method for breast tissue characterization and the descriptive analysis (T-test) to verify the significant difference between the various types of breast tissues and select the optimum wavelength. From the T-test and the IAD regarding the measured R_d, we could verify that the highest R_d values for discrimination were 600 ~ 640 nm at the VIS range and 800 ~ 840 nm at the NIR range, with the minimum tolerance error, as shown in Figure 5.

On the other side, the highest T_r values were 560 ~ 600 nm at the VIS range. Since the tolerance error was high in tumor measurements, it was better to select the wavelength range of 600 ~ 640 nm at the VIS range. Meanwhile, the wavelength range of 760 ~ 800 nm at the NIR range was with the minimum tolerance error, as shown in Figure 6.

Finally, we compared the system outcome with the pathological reports to evaluate the system efficiency and calculate the three numerical values (sensitivity, specificity, and accuracy). Where, the corrected prediction of the system (TP and TN) is compared with the pathological report for each region (malignant in red color and non-malignant in blue color), as shown in Figure 7A. Figure 7B shows the receiver operating characteristic (ROC) curve data illustrating the normal and tumor data to highlight the effect of the cut-off point on decision-making concerning the designed machine learning model. Figure 7C shows the ROC curve to highlight the test's sensitivity (TP rate) and specificity (FP rate) at various cutoff values.

4. Discussion

BM is the second most common cancer in women after skin cancer globally[63]. BM is a threatening disease in both incidence and mortality rates. Therefore, early

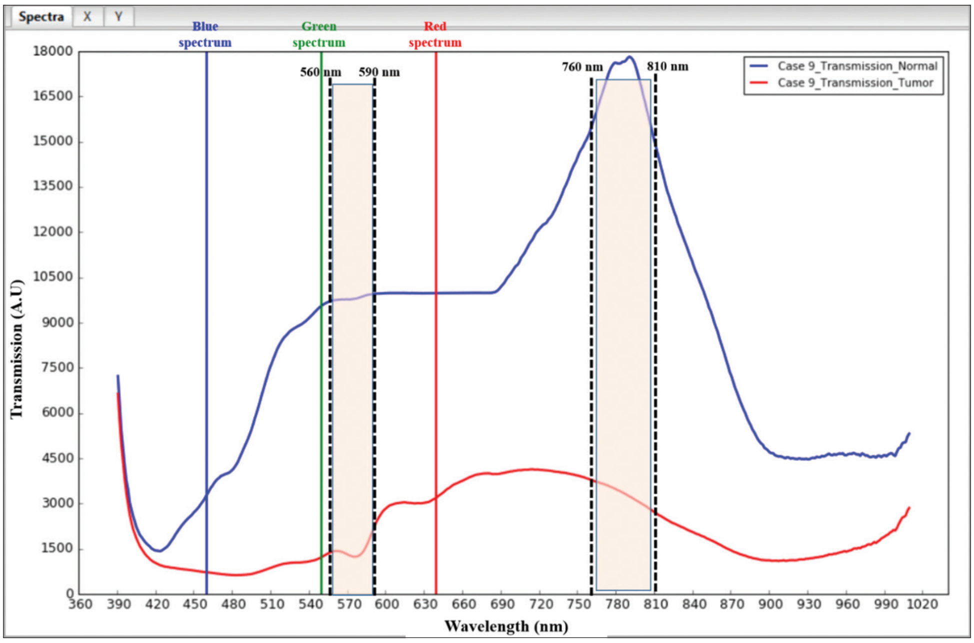


Figure 4. The measured light T_r spectrum for one of the investigated samples (patient ID 1009). The solid red line is for the measured tumor tissue's light T_r spectrum, and the solid blue line identifies the measured normal tissue's T_r spectrum highlighting the peaks which could visually identify between both normal and tumor tissues at wavelength ranges of 560 ~ 590 nm and 760 ~ 810 nm at the visible and near-infrared spectra, respectively.

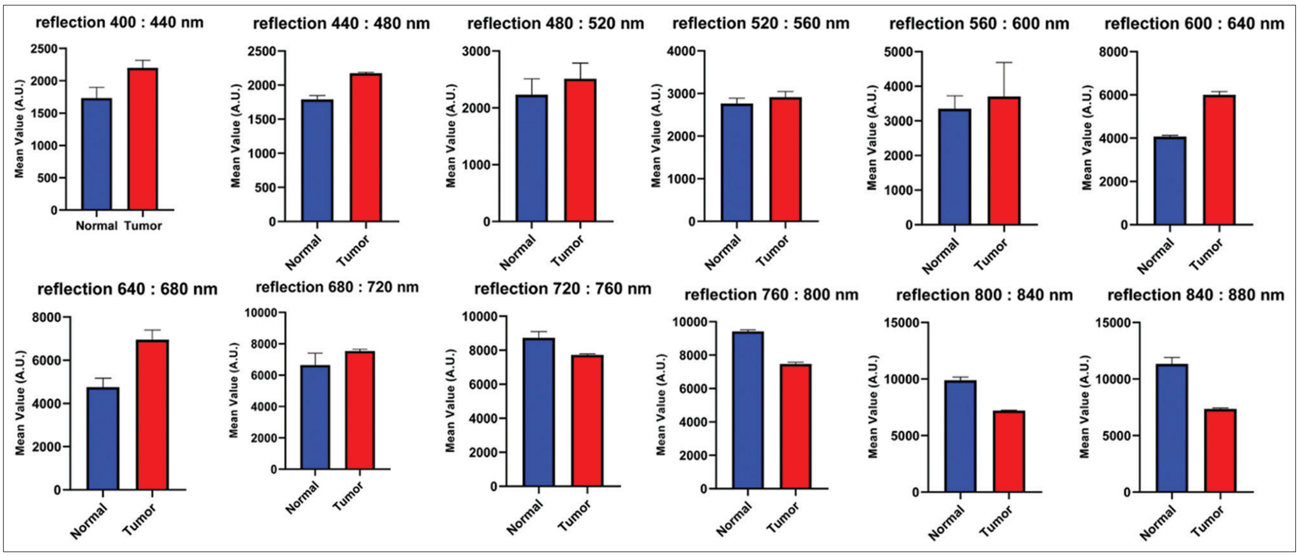


Figure 5. The chart analysis of the inverse adding doubling method associated with the T-test for the measured diffuse reflection (R_d) signatures of the investigated *ex vivo* breast samples from both the normal and breast tumor at the visible and near-infrared (VIS-NIR) spectrum range (400 ~ 1000 nm), with resolution of 40 nm over twelve groups, where the highest R_d values for discrimination were 600 ~ 640 nm at the VIS range and 800 ~ 840 nm at the NIR range, with the minimum tolerance error.

diagnosis is vital for life-saving[64,65], which could go up to 95%[8]. Breast-conserving surgery (BCS) is a common therapy. However, clear surgical margins are vital to avoid cancer recurrence. Additionally, intraoperative pathologic diagnostic techniques, such as imprint cytology and frozen section analysis, are well-known essential tools in BCS. In addition to the traditional methods, there are also modern methods, which have various advantages and disadvantages (Table 1)[66].

Recently, there has been a rapid development in the research of optical methods in the biomedical field sector, leading to a growing number of commercial diagnostic and therapeutic methods. The HSI is a capable non-

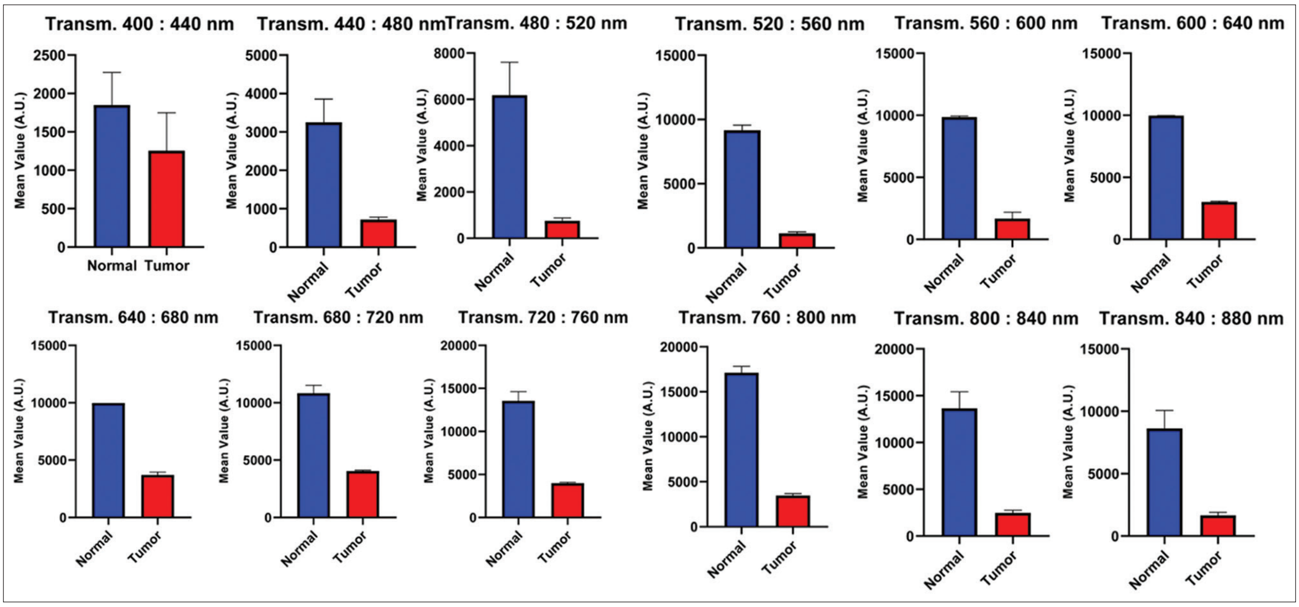


Figure 6. The chart analysis of the inverse adding doubling method associated with the T-test for the measured light transmission (T_r) signatures of the investigated *ex vivo* breast samples from both the normal and breast tumor at the visible and near-infrared (VIS-NIR) spectrum range (400 ~ 1000 nm), with resolution of 40 nm over twelve groups, where the highest T_r values for discrimination were 600 ~ 640 nm at the VIS range and 760 ~ 800 nm at the NIR range, with the minimum tolerance error.

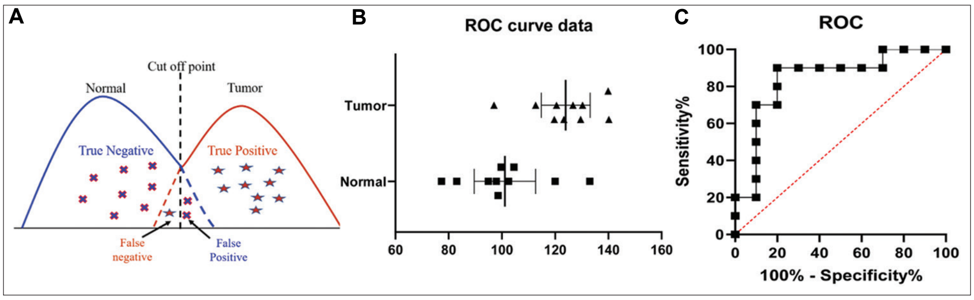


Figure 7. (A) The system outcome where the tumor samples illustrated in red font and normal samples in blue font versus the estimation differences to show the correct prediction and the missed diagnosis in each trial. (B) The receiver operating characteristic (ROC) curve data which represent the normal and tumor data to highlight the effect of the cutoff point on decision-making concerning the designed machine learning model. (C) The ROC curve demonstrating the sensitivity (true positive rate) and specificity (false positive rate) of the test at various cut-off values.

invasive, non-ionizing method that encourages fast achievement and analysis of diagnostic data in the medical field[67]. McCormack *et al.* proposed an optical imaging system incorporating the HS camera, which was capable to map oxygen saturation, vessel density and branching in cancer microvasculature with high resolution to successfully computing malignant microvascular response on anti-cancer therapy over a long period of time[68]. In addition, Kim *et al.* proposed an algorithm to extract the region of interest (ROI) from the HS images of BM samples instead of the visual or manual inspection[69]. Moreover, Pourreza-Shahri *et al.* suggested a classification algorithm to identify the BM margins in the HS images with a sensitivity of 98% and a specificity of 99%[70].

Globally, up to 40% of the BCS necessitate additional surgical procedures due to positive resection margins. Therefore, numerous researchers propose techniques to reduce this value by assessing the resection margins in real time using the HSI system during surgery[71]. The novel techniques in breast cancer investigation exploiting the HSI are briefly depicted in Table S3.

Figure 3 illustrates the first approach of measuring the R_d of the investigated *ex vivo* breast tissue samples. The solid red line represents the measured tumor tissue's R_d spectrum, and the solid blue line represents the normal tissue's R_d spectrum. We noticed from the graph plot that we could visually differentiate between the tumor and the normal tissue at wavelength range of 600 ~ 680 nm and 750

Table 1. Summary of the conventional pathologic methods for the breast-conserving surgery (BCS) and a simple illustration of the proposed new techniques with respect to the gold standard BCS (frozen section analysis).

Technique	Complexity	Process speed	Cost	Margin evaluation	Morphology
Imprint cytology [72]	Simple	Fast	Cost-effective	√	×
Frozen section [66]	Complex	Time-consuming	Expensive	√	√
H&E associated with artificial intelligence [73]	Complex	Fast	Cost-effective	×	√
Standard specimen radiography [74]	Simple	Fast	Cost-effective	√	×
Optical coherence tomography [75]	Simple	Fast	Expensive	×	×
Ultraviolet-photoacoustic microscopy [76]	Simple	Time-consuming	Expensive	×	√
Intraoperative ultrasonography [77]	Simple	Fast	Cost-effective	√	×
Micro-computed tomography [78]	Simple	Fast	Expensive	×	×
Radiofrequency spectroscopy [79]	Simple	Fast	Cost-effective	√	×
Bioimpedance spectroscopy [80]	Simple	Fast	Cost-effective	√	×
HSI attached with standard microscope and deep learning [81]	Complex	Fast	Cost-effective	√	×

~ 960 nm at the VIS and NIR ranges, respectively. Then, from the second framework, we could measure the light T_r, and from the graph, we could visually distinguish between the normal and the tumor tissues at wavelength range of 560 ~ 590 nm and 760 ~ 810 nm in the VIS range and NIR range, respectively.

From the two methods (reflection/transmission), we could calculate the sample μ_a from the measured T_r. Then, we could identify a spectral signature for each tissue type in the VIS-NIR range from the measurements of the previously stated parameters for both the normal and the malignant breast tissues. The measured light tissue's R_d, T_r, and calculated μ_a of the investigated *ex vivo* breast samples are illustrated in Figure S2A-C (Supplementary File), respectively.

Furthermore, we exploited the IAD method for breast tissue characterization and the T-test to verify the significant difference among the various types of breast tissues and to select the optimum wavelength. From the T-test and the IAD regarding the measured R_d, we could verify that the highest R_d values for discrimination were 600 ~ 640 nm at the VIS range and 800 ~ 840 nm at the NIR range, with the minimum tolerance error, as shown in Figure 5. On the other side, the highest T_r values were 590 ~ 600 nm at the VIS range. Since the tolerance error was high in tumor measurements, it was better to select the wavelength range of 600 ~ 640 nm at the VIS range. Meanwhile, the wavelength range of 760 ~ 800 nm at the NIR range was with the minimum tolerance error, as shown in Figure 6.

Finally, to evaluate the system efficiency, we compared the system outcome versus the pathological reports to calculate the corrected prediction (TP) and the incorrect prediction (FP), alongside the FN and TN. Regarding these values, we could plot the ROC curve and determine the system performance (accuracy: 85%; sensitivity: 81.88%; and specificity: 88.8%).

5. Conclusions

The current study showed that the potential and capabilities of HS camera in providing a rapid and non-invasive method to measure the breast tissue's optical properties by capturing the spectral signatures from the malignant and normal breast tissue and to distinguish between them in both the diagnostic and therapeutic applications. The light traveling through tissue was exposed according to two parameters, R_d and μ_a, which depend on the optical properties of the breast tissue. Finally, we conclude from the Rd measurements of the investigated breast samples, that wavelength ranges 600 ~ 640 nm and 800 ~ 840 nm are the optimum ranges to identify the cancerous and non- cancerous regions regarding the diagnostic purpose at the VIS, and NIR spectrum. However, from the Tr values (therapeutic applications), the ideal wavelength ranges were 600 ~ 640 nm, and 760 ~ 800 nm. In the present study, the average accuracy, sensitivity, and specificity were 85%, 81.88%, and 88.8%, respectively. In future work, we intend to explore this direction in infrared band using a commercial and low-cost spectral detector to calculate the optimum wavelength with the highest R_d value to develop an alternative low-cost and rapid diagnostic technique of breast cancer.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization: Mohamed Hisham Aref, Ibrahim H. Aboughaleb, Yasser H. El-Sharkawy
Investigation: Mohamed Hisham Aref, Ibrahim H. Aboughaleb
Resources: Abdallah Abdelkader Hussein, Ayman Mohammed Farag, Sara Abd El-Ghaffar
Supervision: Ayman Mohammed Farag, Sara Abd El-Ghaffar, Yasser H. El-Sharkawy
Writing – original draft: Mohamed Hisham Aref, Ibrahim H. Aboughaleb
Writing – review & editing: Yasser H. El-Sharkawy

Ethics approval and consent to participate

No experimental investigation was performed on any individuals within this study and all patients read and signed two copies of a consent form before the beginning of data collection.

Consent for publication

All the authors declare that all the participated patients had been well informed about the investigation. All patients read and signed two copies of a consent form before the beginning of data collection, patient data had been coded and summarized in Table S2 in Supplementary file.

Availability of data

The authors stated and declare that all the datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Ferlay J, Colombet M, Soerjomataram I, *et al.*, 2021, Cancer statistics for the year 2020: An overview. *Int J Cancer*, 149(4): 778–789.

   https://doi.org/10.1002/ijc.33588

2. Sung H, Ferlay J, Siegel RL, *et al.*, 2021, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 71(3): 209–249.

   https://doi.org/10.3322/caac.21660

3. World Health Organization, 2021, World Health Statistics 2021. Geneva: World Health Organization.

4. Fitzmaurice C, Dicker D, Pain A, *et al.*, 2015, The global burden of cancer 2013. *JAMA Oncol*, 1(4): 505–527.

   https://doi.org/10.1001/jamaoncol.2015.0735

5. Ferlay J, Soerjomataram I, Dikshit R, *et al.*, 2015, Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, 136(5): E359–E386.

   https://doi.org/10.1002/ijc.29210

6. Mody GN, Nduaguba A, Ntirenganya F, *et al.*, 2013, Characteristics and presentation of patients with breast cancer in Rwanda. *Am J Surg*, 205(4): 409–413.

   https://doi.org/10.1016/j.amjsurg.2013.01.002

7. Faust O, Acharya UR, Meiburger KM, *et al.*, 2018, Comparative assessment of texture features for the identification of cancer in ultrasound images: A review. *Biocybern Biomed Eng*, 38(2): 275–296.

8. Etehadtavakol M, Ng EY, 2013, Breast thermography as a potential non-contact method in the early detection of cancer: A review. *J Mech Med Biol*, 13(2): 1330001.

9. Nofiele JT, Czarnota GJ, Cheng HL, 2014, Noninvasive manganese-enhanced magnetic resonance imaging for early detection of breast cancer metastatic potential. *Mol Imaging*, 13(1): 1–10.

   https://doi.org/10.2310/7290.2013.00071

10. Gøtzsche PC, Jørgensen KJ, 2013, Screening for breast cancer with mammography. *Cochrane Database Syst Rev*, 2013(6): CD001877.

    https://doi.org/10.1002/14651858.CD001877.pub5

11. Brem RF, Lenihan MJ, Lieberman J, *et al.*, 2015, Screening breast ultrasound: Past, present, and future. *Am J Roentgenol*, 204(2): 234–240.

    https://doi.org/10.2214/AJR.13.12072

12. Hambly NM, Liberman L, Dershaw DD, *et al.*, 2011, Background parenchymal enhancement on baseline screening breast MRI: Impact on biopsy rate and short-interval follow-up. *Am J Roentgenol*, 196(1): 218–224.

    https://doi.org/10.2214/AJR.10.4550

13. Mehnati P, Tirtash MJ, 2015, Comparative efficacy of four imaging instruments for breast cancer screening. *Asian Pac J Cancer Prev*, 16(15): 6177–6186.

    https://doi.org/10.7314/apjcp.2015.16.15.6177

14. Gutierrez RM, Cerquera EA, Mañana G, 2012, MPGD for breast cancer prevention: A high resolution and low dose radiation medical imaging. *J Instrum*, 7(7): C07007.

    https://doi.org/10.1088/1748-0221/7/07/C07007

15. Wang L, 2017, Early diagnosis of breast cancer. *Sensors (Basel)*, 17(7): 1572.

    https://doi.org/10.3390/s17071572

16. Breastcancer.org (2006). Your Guide to the Breast Cancer Pathology Report. Available from: https://www.breastcancer.org [Last accessed on 2022 Nov 15].

17. Sahu A, 2012, Hyperspectral Imaging to Discern Malignant and Benign Canine Mammary Tumors a Thesis Proposal Submitted to the Temple University Graduate Board By Amrita Sahu.

18. Ouyang Y, Tsui P, Wu S, *et al*., 2019, Classification of benign and malignant breast tumors using H-scan ultrasound imaging. *Diagnostics (Basel)*, 9(4): 182.

https://doi.org/10.3390/diagnostics9040182

19. Merrill AL, Coopey SB, Tang R, *et al*., 2016, Implications of new lumpectomy margin guidelines for breast-conserving surgery: Changes in reexcision rates and predicted rates of residual tumor. *Ann Surg Oncol*, 23(3): 729–734.

https://doi.org/10.1245/s10434-015-4916-2

20. Vos EL, Jager A, Verhoef C, *et al*., 2015, Overall survival in patients with a re-excision following breast conserving surgery compared to those without in a large population-based cohort. *Eur J Cancer*, 51(3): 282–291.

https://doi.org/10.1016/j.ejca.2014.12.003

21. Kho E, Dashtbozorg B, de Boer LL, *et al*., 2019, Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information. *Biomed Opt Express*, 10(9): 4496.

https://doi.org/10.1364/boe.10.004496

22. Boyce BF, 2017, An update on the validation of whole slide imaging systems following FDA approval of a system for a routine pathology diagnostic service in the United States. *Biotech Histochem*, 92(6): 381–389.

https://doi.org/10.1080/10520295.2017.1355476

23. Evans AJ, Bauer TW, Bui MM, *et al*., 2018, US Food and Drug Administration approval of whole slide imaging for primary diagnosis: A key milestone is reached and new questions are raised. *Arch Pathol Lab Med*, 142(11): 1383–1387.

https://doi.org/10.5858/arpa.2017-0496-CP

24. Flotte TJ, Bell DA, 2018, Anatomical pathology is at a crossroads. *Pathology*, 50(4): 373–374.

https://doi.org/10.1016/j.pathol.2018.01.003

25. Motomura K, Inaji H, Komoike Y, *et al*., 2000, Intraoperative sentinel lymph node examination by imprint cytology and frozen sectioning during breast surgery. *Br J Surg*, 87(5): 597–601.

https://doi/org/10.1046/j.1365-2168.2000.01423.x

26. Flett MM, Going JJ, Stanton PD, *et al*., 1998, Sentinel node localization in patients with breast cancer. *Br J Surg*, 85(7): 991–993.

https://doi.org/10.1046/j.1365-2168.1998.00746.x

27. Van Diest PJ, Torrenga H, Borgstein PJ, *et al*., 1999, Reliability of intraoperative frozen section and imprint cytological investigation of sentinel lymph nodes in breast cancer. *Histopathology*, 35(1): 14–18.

https://doi.org/10.1046/j.1365-2559.1999.00667.x

28. Aref M, Youssef AB, Hussein AA, *et al*., 2022, Custom fluorescence imaging system exploiting hyperspectral camera to characterize and diagnose RNA breast cancer. *Biointerface Res Appl Chem*, 12(4): 5548–5566.

https://doi.org/10.33263/BRIAC124.55485566

29. Welch AJ, Van Gemert MJ, 2011, Optical-Thermal Response of Laser-Irradiated Tissue Vol. 2. Berlin: Springer.

30. Bioucas-Dias JM, Plaza A, Camps-Valls G, *et al*., 2013, Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci Remote Sens Magazine*, 1(2): 6–36.

31. Adão T, Hruška J, Pádua L, *et al*., 2017, Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens*, 9(11): 1110.

https://doi.org/10.3390/rs9111110

32. Robles-Kelly A, Huynh CP, 2013, Imaging spectroscopy for scene analysis. In: *Imaging Spectroscopy for Scene Analysis*. Berlin: Springer.

https://doi.org/10.1007/978-1-4471-4652-0

33. Lu G, Fei B, 2014, Medical hyperspectral imaging: A review. *J Biomed Opt*, 19(1): 10901.

https://doi.org/10.1117/1.JBO.19.1.010901

34. Fabelo H, Ortega S, Lazcano R, *et al*., 2018, An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation. *Sensors (Basel)*, 18(2): 430.

https://doi.org/10.3390/s18020430

35. Goto A, Nishikawa J, Kiyotoki S, *et al*., 2015, Use of hyperspectral imaging technology to develop a diagnostic support system for gastric cancer. *J Biomed Opt*, 20(1): 16017.

https://doi.org/10.1117/1.JBO.20.1.016017

36. Regeling B, Thies B, Gerstner AO, *et al*., 2016, Hyperspectral imaging using flexible endoscopy for laryngeal cancer detection. *Sensors (Basel)*, 16(8): 1288.

https://doi.org/10.3390/s16081288

37. Aref MH, Youssef AB, Aboughaleb IH, *et al*., 2021, Prospective study for commercial and low-cost hyperspectral imaging systems to evaluate thermal tissue effect on bovine liver samples. *J Spectral Imaging*, 10: a5.

38. Aref MH, Aboughaleb IH, Youssef AB, *et al*., 2021, Hyperspectral image-based analysis of thermal damage for *ex-vivo* bovine liver utilizing radiofrequency ablation. *Surg Oncol*, 38: 101564.

https://doi.org/10.1016/j.suronc.2021.101564

39. El-Sharkawy YH, Aref MH, Elbasuney S, *et al*., 2022, Oxygen saturation measurements using novel diffused reflectance with hyperspectral imaging: Towards facile COVID-19

diagnosis. *Opt Quantum Electron*, 54(5): 322.

https://doi.org/10.1007/s11082-022-03658-z

40. Aref MH, Sharawi AA, El-Sharkawy YH, 2021, Delineation of the arm blood vessels utilizing hyperspectral imaging to assist with phlebotomy for exploiting the cutaneous tissue oxygen concentration. *Photodiagnosis Photodyn Ther*, 33: 102190.

    https://doi.org/10.1016/j.pdpdt.2021.102190

41. Aref M, Youssef AB, Hussein AA, *et al.*, 2021, Custom fluorescence imaging system exploiting hyperspectral camera to characterize and diagnose RNA breast cancer. *Biointerface Res Appl Chem*, 2: 5548–5566.

42. Aref MH, Abbass MA, Youssef AB, *et al.*, 2022, Optical Characterization of Biological Tissues in Visible and Near-Infrared Spectra. In: *2022 13th International Conference on Electrical Engineering (ICEENG)*. IEEE. p. 159–163.

43. Aboughaleb IH, Matboli M, Shawky SM, *et al.*, 2021, Transcriptome spectral analysis using hyperspectral imaging for hepatocellular carcinoma detection. *QJM Int J Med*, 114(Suppl 1): hcab088-004.

44. Bashkatov AN, Genina EA, Kochubey VI, *et al.*, 2005, Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *J Phys D Appl Phys*, 38(15): 2543–2555.

    https://doi.org/10.1088/0022-3727/38/15/004

45. Ostantini IR, Icchi RI, Ilvestri LU, *et al.*, 2019, *In-vivo* and *ex-vivo* optical clearing methods for biological tissues : Review. *Biomed Opt Express*, 10(10): 5251–5267.

    https://doi.org/10.1364/BOE.10.005251

46. Tuchin VV, 1997, Light scattering study of tissues. *Phys Uspekhi*, 40(5): 495–515.

    https://doi.org/10.1070/pu1997v040n05abeh000236

47. Aref MH, Youssef AB, Aboughaleb IH, *et al.*, 2021, Characterization of normal and malignant breast tissues utilizing hyperspectral images and associated differential spectrum algorithm. *J Biomed Photon Eng*, 7(2): 1–12.

    https://doi.org/10.18287/JBPE21.07.020302

48. Fajardo C, Solarte E, 2020, Optical properties of a simple model of soft biological tissue. *J Phys Conf Ser*, 1547: 012026.

    https://doi.org/10.1088/1742-6596/1547/1/012026

49. Pogue BW, Patterson MS, 2022, Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry. *J Biomed Opt*, 11: 041102.

    https://doi.org/10.1117/1.2335429

50. Patterson MS, Wilson BC, Wyman DR, 1991, The propagation of optical radiation in Tissue I. Models of radiation transport and their application. Lasers Med Sci, 6: 155–168.

51. Vo-Dinh T, 2014, *Biomedical Photonics Handbook: Biomedical Diagnostics*. United States: CRC Press.

52. DiPerna RJ, Lions PL, 1989, Ordinary differential equations, transport theory and Sobolev spaces. *Invent Math*, 98(3): 511–547.

    https://doi.org/10.1007/BF01393835

53. Rinzema K, Murrer LH, Star WM, 1998, Direct experimental verification of light transport theory in an optical phantom. *J Opt Soc Am A*, 15(8): 2078–2088.

54. Yavari N, 2006, Optical spectroscopy for tissue diagnostics and treatment control. *Transport*, 760: 141–157.

    https://doi.org/10.1007/978-1-61779-176-5

55. Xie D, Guo W, 2020, Measurement and calculation methods on absorption and scattering properties of turbid food in Vis/NIR range. *Food Bioprocess Technol*, 13(2): 229–244.

    https://doi.org/10.1007/s11947-020-02402-3

56. Noor SS, Michael K, Marshall S, *et al.*, 2017, Hyperspectral image enhancement and mixture deep-learning classification of corneal epithelium injuries. *Sensors*, 17(11): 2644.

    https://doi.org/10.3390/s17112644

57. Palmer GM, Ramanujam N, 2006, Monte Carlo-based inverse model for calculating tissue optical properties. Part I: Theory and validation on synthetic phantoms. *Appl Opt*, 45(5): 1062–1071.

58. Alexandrakis G, Farrell TJ, Patterson MS, 1998, Accuracy of the diffusion approximation in determining the optical properties of a two-layer turbid medium. *Appl Opt*, 37(31): 7401–7409.

    https://doi.org/10.1364/ao.37.007401

59. Prahl SA, van Gemert MJ, Welch AJ, 1993, Determining the optical properties of turbid media by using the adding-doubling method. *Appl Opt*, 32(4): 559–568.

60. Wan S, Anderson RR, Parrish JA, 1981, Analytical modeling for the optical properties of the skin with *in vitro* and *in vivo* applications. *Photochem Photobiol*, 34(4): 493–499.

61. Hourdakis CJ, Perris A, 1995, A monte carlo estimation of tissue optical properties for use in laser dosimetry. *Phys Med Biol*, 40(3): 351–364.

    https://doi.org/10.1088/0031-9155/40/3/002

62. Jacques SL, Wang L, 1995, Monte Carlo modeling of light transport in tissues. In: *Optical-Thermal Response of Laser-Irradiated Tissue*. Berlin: Springer. p. 73–100.

63. Peng L, Chen W, Zhou W, *et al.*, 2016, An immune-inspired semi-supervised algorithm for breast cancer diagnosis. *Comput Methods Programs Biomed*, 134: 259–265.

    https://doi.org/10.1016/j.cmpb.2016.07.020

64. Hossain S, Mohammadi FA, 2016, Tumor parameter estimation considering the body geometry by thermography. *Comput Biol Med*, 76: 80–93.

https://doi.org/10.1016/j.compbiomed.2016.06.023

65. Hossain S, Mohammadi FA, Nejad ET, 2011, Neural network approach for the determination of heat source parameters from surface temperature image. In: *2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE. p. 1109–1112.

66. Pradipta AR, Tanei T, Morimoto K, *et al.*, 2020, Emerging technologies for real-time intraoperative margin assessment in future breast-conserving surgery. *Adv Sci (Weinh)*, 7(9): 1901519.

    https://doi/org/10.1002/advs.201901519

67. Halicek M, Fabelo H, Ortega S, *et al.*, 2019, *In-vivo* and *ex-vivo* tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer. *Cancers*, 11(6): 756.

    https://doi.org/10.3390/cancers11060756

68. McCormack DR, Walsh AJ, Sit W, *et al.*, 2014, *In vivo* hyperspectral imaging of microvessel response to trastuzumab treatment in breast cancer xenografts. *Biomed Opt Express*, 5(7): 2247–2261.

    https://doi.org/10.1364/BOE.5.002247

69. Kim B, Kehtarnavaz N, LeBoulluec P, Liu H, *et al.*, 2013, Automation of ROI Extraction in Hyperspectral Breast Images. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. p. 3658–3661.

70. Pourreza-Shahri R, Saki F, Kehtarnavaz N, Leboulluec P, *et al.*, 2013, Classification of *ex-vivo* Breast Cancer Positive Margins Measured by Hyperspectral Imaging. In: *2013 IEEE International Conference on Image Processing, ICIP 2013-Proceedings*. p. 1408–1412.

    https://doi.org/10.1109/ICIP.2013.6738289

71. Sterenborg HJ, Kho E, de Boer LL, *et al.*, 2014, Hyperspectral imaging for intraoperative margin assessment during breast cancer surgery. *Opt InfoBase Conf Papers*, 2016: 6–9.

    https://doi.org/10.1364/ACPC.2016.AF1K.6

72. Bakhshandeh M, Tutuncuoglu SO, Fischer G, *et al.*, 2007, Use of imprint cytology for assessment of surgical margins in lumpectomy specimens of breast cancer patients. *Diagn Cytopathol*, 35(10): 656–659.

73. Coudray N, Ocampo PS, Sakellaropoulos T, *et al.*, 2018, Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med*, 24(10): 1559–1567.

74. Kim SH, Cornacchi SD, Heller B, *et al.*, 2013, An evaluation of intraoperative digital specimen mammography versus conventional specimen radiography for the excision of nonpalpable breast lesions. *Am J Surg*, 205(6): 703–710.

    https://doi/org/10.1016/j.amjsurg.2012.08.010

75. Ha R, Friedlander LC, Hibshoosh H, *et al.*, 2018, Optical coherence tomography: A novel imaging method for post-lumpectomy breast margin assessment-a multi-reader study. *Acad Radiol*, 25(3): 279–287.

    https://doi.org/10.1016/j.acra.2017.09.018

76. Wong TT, Zhang R, Hai P, *et al.*, 2017, Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy. *Sci Adv*, 3(5): e1602168.

    https://doi.org/10.1126/sciadv.1602168

77. Moschetta M, Telegrafo M, Introna T, *et al.*, 2015, Role of specimen US for predicting resection margin status in breast conserving therapy. *G Chir*, 36(5): 201–204.

    https://doi.org/10.11138/gchir/2015.36.5.201

78. Tang R, Saksena M, Coopey SB, *et al.*, 2016, Intraoperative micro-computed tomography (micro-CT): A novel method for determination of primary tumour dimensions in breast cancer specimens. *Br J Radiol*, 89(1058): 20150581.

79. Schnabel F, Boolbol SK, Gittleman M, *et al.*, 2014, A randomized prospective study of lumpectomy margin assessment with use of MarginProbe in patients with nonpalpable breast malignancies. *Ann Surg Oncol*, 21(5): 1589–1595.

    https://doi.org/10.1245/s10434-014-3602-0

80. Dixon JM, Renshaw L, Young O, *et al.*, 2016, Intra-operative assessment of excised breast tumour margins using ClearEdge imaging device. *Eur J Surg Oncol*, 42(12): 1834–1840.

    https://doi.org/10.1016/j.ejso.2016.07.141

81. Ortega S, Halicek M, Fabelo H, *et al.*, 2020, Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images. *Proc SPIE Int Soc Opt Eng*, 11320: 206–214.

ACCSCIENCE PUBLISHING

Tumor Discovery

ORIGINAL RESEARCH ARTICLE

A novel gene prognostic signature lymphocyte cytosolic protein 2 regulates melanoma progression by activating tumor-infiltrating CD8⁺ T-cells through the interferon regulatory factor 5 signaling pathway

Hongyin Sun[1,2,3†], Kui Deng[4†], Xingchen Zhou[1,2†], Dongsheng Cao[5], Yan Cheng[6], Xiang Chen[1,2*], and Mingzhu Yin[1,2*]

[1]Department of Dermatology, Hunan Engineering Research Center of Skin Health and Disease, Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, China

[2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[3]Department of Pharmacy, Shantou University Medical College, Shantou University, Shantou, China

[4]Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, 18 Shilongshan Rd, Cloud Town, Hangzhou, China

[5]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, Hunan, China

[6]Department of Pharmacy, The Second Xiangya Hospital, Central South University, Changsha, China

Abstract

Cutaneous malignant melanoma is the most lethal skin cancer. The advent of immunotherapy has revolutionized the status of clinical therapies of melanoma, which brought new hope to these patients. However, only a small proportion of patients are responders. Therefore, the identification of novel prognostic and immune-related biomarkers is crucial to guide the development of melanoma clinical treatments. Herein, RNA-seq data of the cutaneous melanoma from public database were used for identifying prognostic gene signatures, and we found that high expression of lymphocyte cytosolic protein 2 (LCP2) in melanoma patient was significantly associated with better prognosis for melanoma. Kyoto Encyclopedia of Genes and Genomes and gene ontology analyses demonstrated that the differentially expressed genes are significantly involved in lysosome, B-cell receptor signaling pathways, Fc epsilon RI signaling pathway, and T-cell receptor signaling pathway, indicating that these signaling pathways play important roles in melanoma. LCP2 expression was positively correlated with CD8⁺ T-cell and the overall survival of melanoma patients, and this positive correlation was directly confirmed by fluorescence-activated cell sorting experiment. The *in vivo* experiment showed that LCP2 knockdown significantly promoted the melanoma progression and decreased interferon regulatory factor 5 (IRF5) expression. In conclusion, we identified that LCP2 is a possible prognostic gene signature for progression-free survival of melanoma patients and regulates melanoma progression by activating tumor-infiltrating CD8⁺ T-cells through the IRF5 signaling pathway, indicating that LCP2 could serve as a prognostic biomarker and therapeutic target in immunotherapy.

[†]These authors contributed equally to this work.

***Corresponding authors:**
Xiang Chen
(chenxiangck@126.com)
Mingzhu Yin
(yinmingzhu@csu.edu.cn)

Citation: Sun H, Deng K, Zhou X, *et al*., 2023, A novel gene prognostic signature lymphocyte cytosolic protein 2 regulates melanoma progression by activating tumor-infiltrating CD8⁺ T-cells through the interferon regulatory factor 5 signaling pathway. *Tumor Discov*, 2(1): 318.
https://doi.org/10.36922/td.318

Received: December 29, 2022
Accepted: February 27, 2023
Published Online: March 24, 2023

Keywords: Cutaneous melanoma; Lymphocyte cytosolic protein 2; Prognostic gene signature; Immunotherapies; Interferon regulatory factor 5

1. Introduction

Melanoma is known to be the most lethal type of skin cancer, with an estimated over 160,000 new cases diagnosed each year[1,2]. Although melanoma constitutes <5% of all skin cancers, 80% of skin cancer-related deaths are caused by melanoma. Despite great advances in biological research and therapeutic approaches in the past years, the morbidity and mortality rates have still increased in the past half century[2]. Due to high immunogenicity of melanoma, the recent breakthroughs achieved in immunotherapies have revolutionized the way we treat melanoma, and this has led to a dramatic improvement in the prognosis of the patients[3,4]. However, only a small proportion of patients respond to these immunotherapies, which means that the mechanism of interaction between immune system and melanoma invasion has not yet been fully elucidated. Therefore, it is necessary to identify transcriptome signatures within tumor immune microenvironment that affects tumor progression and patient survival.

Several studies have successfully identified transcriptome signatures that could predict the survival of melanoma patients[5,6]. These signatures could help discern melanoma patients with high risk in advance and evaluate response to certain therapeutic interventions. Few studies have explored the relationship between the immune gene expression and the prognosis for melanoma. Zhao *et al.* identified a 25-gene signature from immune cells-related genes though selecting the genes with the highest correlation with the ImmuneScore that reflects leukocyte infiltration, and calculated leukocyte infiltration score to predict the survival of melanoma patients[7]. However, not all 25 genes identified in this study may be associated with the development and progression of melanoma.

Lymphocyte cytosolic protein 2 (LCP2), one of the SLP-76 family of adapters, is primarily expressed in hematopoietic cells[8]. In addition, LCP2 is also reported to be expressed in platelets, macrophages, neutrophils, mast cells, developing B-cells, and natural killer cells[8,9]. However, the current data showed that LCP2 is differentially expressed in many different types of tumors and is correlated with prognosis in gastric cancer, breast cancer[10], lung adenocarcinoma,[11] and metastatic melanoma[12] patients. Furthermore, it has been reported that LCP2 is involved in T-cell-receptor signaling[13] and is essential for normal T-cell development and activation[14]. LCP2 can activate the promoter of interleukin (IL)-2 through promoting the activation of T-cell, and IL-2 secretion subsequently results in the release of cytokines that lead to apoptosis of target cells[15,16]. Moreover, Siggs *et al.* found that the mice with LCP2 mutation secreted excessive amounts of proinflammatory cytokines, IgE, and autoantibodies, suggesting that slight decrease in LCP2 expression is enough to trigger immune dysregulation[17]. In a nutshell, LCP2 plays a significant important role in the immune system. However, the specific role of LCP2 in the tumor microenvironment is still obscure.

As melanoma exhibits high immunogenicity, we hypothesized that identifying immune-related genes that are associated with the occurrence and prognosis of melanoma could help delineate the melanoma-specific mechanisms of immunoregulation, which may shed new light on immunotherapy research. In the present study, the RNA-seq data of cutaneous melanoma from the public database were used for identifying prognostic gene signatures, and we found that LCP2 was highly expressed in the patients with melanoma. The expression of LCP2 was positively correlated with $CD8^+$ T-cell and the overall survival (OS) of melanoma patients, and this positive correlation was directly confirmed by flow cytometry analysis. Moreover, we further explored the potential mechanism of LCP2 in tumor microenvironment of melanoma *in vivo*. Knockdown of LCP2 significantly promoted the progression of melanoma and decreased the IRF5 expression. Based on our results, we identified that LCP2 is a possible prognostic gene signature for progressive-free survival of patients with melanoma, which repressed the growth of melanoma through positively regulating the IRF5 signaling pathway followed by activation of $CD8^+$ T-cells and promotion of antitumor response.

2. Materials and methods

Major immune pathways ($N = 29$) were retrieved from the Reactome pathway database (https://reactome.org/)[18], which contains three main immune pathways including innate immune system, adaptive immune system, and cytokine signaling in immune system, and 2070 genes in total.

2.1. Data sets used

We used two expression profile data sets to identify differentially expressed genes (DEGs) between melanomas and controls. These data sets are GSE3189 (including 45 melanoma samples, 18 benign skin nevus samples, and seven normal skin samples) and GSE31879 (including ten melanoma samples and four normal melanocyte samples). GSE15605 (including 58 melanoma samples and 16 normal skin samples) was used to validate the identified DEGs. The RNA-seq data of cutaneous melanoma from the Cancer Genome Atlas (TCGA) were downloaded from the UCSC Cancer Brower (https://xenabrowser.net/), which contained 471 melanoma samples. The TCGA RNA-seq data were utilized to further identify prognostic genes that could predict the survival of melanoma, and to construct prognostic models. In addition, GSE65904, which contains 214 melanoma samples, was used to validate the identified prognostic gene signatures. Variables with missing values >20% of all samples were excluded from this study, and the remaining missing values were imputed by nearest neighbor averaging method using the R package *impute*[19]. The characteristics of the five data sets used in our study are provided in Table S1.

2.2. Identification and validation of DEGs

GSE3189 and GSE31879 data sets were used to select DEGs. DEGs between melanoma and benign skin nevus, melanoma and normal skin tissue, melanoma, and normal melanocyte were screened by Wilcoxon rank-sum test, separately. The false discovery rate (FDR) method was used to correct type I error occurring when conducting multiple comparison[20]. Variables with q value generated by FDR <0.01 were selected, and the DEGs that were shared between melanoma and three controls (nevus, normal skin, and melanocyte) were screened as the robust DEGs. The robust DEGs were then validated in GSE15605, and DEGs that were also statistically significant (q value < 0.01) in GSE15605 were used in following analyses. The identified DEGs were further validated in the Gene Expression Profiling Interactive Analysis (GEPIA) online database (http://gepia.cancer-pku.cn/)[21], which contains a large number of normal samples that were derived from the Genotype-Tissue Expression (GTEx) projects in addition to TCGA melanoma samples.

2.3. Pathway and functional enrichment analyses of DEGs

To gain more insights into the functions and biological processes of the DEGs, we utilized the R package *clusterProfiler* for Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) enrichment analyses[22]. Significant pathways and GO items were selected with FDR $q < 0.05$.

2.4. Identification and validation of prognostic gene signatures

DEGs that could be used to predict the OS of patients with melanoma were selected using the univariate Cox proportional hazards models in TCGA data set. Genes with $P < 0.05$ were then fitted in the multivariate LASSO Cox regression model with the R package *glmnet*[23]. The optimal λ value was determined by ten-fold cross validation that gives a minimum cross-validated error. Coefficients of genes without significant effect on OS were shrunk to zero. The correlations among significant genes were assessed using Spearman's correlation method. Then, the prognostic immune score (PIS) for each patient was constructed, which is the linear combination of the expression of genes and the corresponding coefficients estimated in LASSO model. The patients in TCGA data set were divided into high-risk group and low-risk group based on the median of PIS. The time of OS and disease-free survival (DFS) between the two groups were compared using Kaplan–Meier analyses and log-rank tests, separately. The identified prognostic gene signature was validated in an independent cohort GSE65904. PIS was calculated according to the expression of each identified gene and its associated coefficient obtained from the TCGA data set. Subsequently, the high-risk group and low-risk group were formed according to the median of PIS, and the time of OS and distant metastasis-free survival (DMFS) between the two groups were compared by Kaplan–Meier analyses and log-rank tests, separately. Since TCGA data set has relatively more complete clinical information than other data sets, we investigated whether prognostic value of PIS was independent of clinical baseline information using multivariate Cox proportional hazards models in TCGA data set. $P < 0.05$ was set as a cutoff for significant difference. All data were statistically analyzed with R software (version 3.4.4).

2.5. Exploring the associations between identified prognostic genes and tumor-associated leukocyte (TAL) subsets

Leukocyte compositions were inferred from the bulk tumor transcriptomes using CIBERSORT, which could estimate relative proportions of TAL from expression profiles of bulk tumors and outperform other TAL decomposition methods[24]. After applying CIBERSORT to TCGA data set, 22 distinct TAL subsets were obtained, and the correlation between each identified prognostic gene and the specific TAL was calculated using Spearman's rank correlation. Genes that had the largest absolute correlation coefficients with certain TAL subset among all TAL subsets were considered to have strong association with that TAL subset. In addition, melanoma subset was defined as the average expression of known melanoma markers (MLANA, S100A1, PAN2, SOX10, TYR, S100B, and MITF), and the association between each identified prognostic gene and melanoma subset was explored using Spearman's rank correlation.

2.6. Cell culture, stimulation, and cell transfection

A375 and B16F10 cells were purchased from ATCC. They were cultured in DMEM (01-052-1A, Biological Industries, Israel) cell culture medium with 10% fetal bovine serum (04-001-1ACS, Biological Industries, Israel). In cell culture experiments, transfection of LCP2 shRNA (GenePharma Biotechnology Company, Shanghai, China) was conducted based on the operation manual. Non-targeting shRNA was used as a control. shRNAs were transfected into A375 and B16F10 cells with Lipofectaminected i (L3000015, Invitrogen) according to corresponding instructions. For obtaining stable LCP2 knockdown cells, pMD2.G, psPAX2, and shRNA were cotransfected into HEK-293T cells with TurboFect (R0532, Thermo Fisher Scientific) based on the corresponding operation manual. RT-PCR was used for efficiency verification of transfection.

2.7. Animal studies

Female BALB/c nude mice (4-6 weeks old) and female C57BL/6 mice (6-8 weeks old) were used for A375 experiments and B16F10 experiments, respectively. 1ect^{6} A375 or B16F10 cells (0.1 mL) were subcutaneously implanted to female nude or C57BL/6 mice. To determine tumor growth, a caliper was used to record the tumor sizes every 5 days. All animal studies were evaluated and approved by the Institutional Animal Care and Use Committee of Xiangya Hospital, Central South University (No. 2022020685).

2.8. RT-qPCR

Tripure (Bio Teke, RP1002) and a reverse transcription kit (HiScript Q RT, Vazyme, R223-01) were used to extract the total RNA and reverse-transcribe RNA into cDNA, respectively. The PCR system was prepared according to the SYBR qPCR Mix kit (Bimake, B21703). The relatively expression levels of LCP2 and IRF5 were detected by normalization to GAPDH using ΔΔCt method. The primer sequences used for the RT-qPCR are listed in Table 1.

2.9. Flow cytometry

2% paraformaldehyde solution in phosphate-buffered saline (PBS) was used for cell fixation, then stained with fluorescent-labeled antibodies for cell-surface markers for 1 h at 4°C in the dark, and subsequently washed and resuspended for flow cytometry analysis. The antibodies used in this part are as follows: anti-mouse CD4 FITC (100406), anti-mouse CD45 APC (103112), anti-mouse CD4 PE (100407), anti-mouse/human Granzyme B PE (372208), AF647 anti-human/mouse Granzyme B antibody (515405), and anti-mouse CD8 APC (100766). The FACSCalibur (BD Biosciences) and FlowJo software (Version 10.4) were used for flow cytometry analysis and data analysis, respectively.

Table 1. Primer sequences used for the RT-qPCR

Gene	Primer sequence (5' to 3')
LCP2	
Forward	AGAATGTCCCGTTTCGCTCAG
Reverse	TGCTCCTTCTCTCTTCGTTCTT
IRF5	
Forward	AGAGACAGGGAAGTACACTGAAG
Reverse	TGGAAGTCACGGCTTTTGTTAAG
GAPDH	
Forward	CTCTGCTCCTCCTGTTCGAC
Reverse	GCCCAATACGACCAAATCC

RT-qPCR: Real time-quantitative polymerase chain reaction

2.10. Identification of differentially expressed genes associated with LCP2

To further explore the mechanism of LCP2 regulating progression of melanoma, we analyzed the DEGs between LCP2 knockdown group and control group. Ribo-Zero rRNA removal kit was used to remove rRNA from 1 µg of total RNA. TruSeq RNA sample prep kit (Illumina) was used to generate sequencing libraries, and then sequenced as 151-bp paired-end reads using an Illumina HiSeq X Ten platform, with $P < 0.05$ and $|\text{foldchange}| > 1$ as cutoff value.

2.11. Statistical analysis

All data from the experiments are expressed as mean ± SEM. Statistical analyses of these data were performed using GraphPad Prism 9 (GraphPad Software, La Jolla, CA) with an unpaired Student's *t*-test or Tukey's multiple comparison. $P < 0.05$ was set as a cutoff for significant difference.

3. Results

3.1. Identification of DEGs

Among the 2070 immune-centric genes, there were 1582 immune-centric genes in GSE3189 data set and 1842 in GSE31879 data set. Based on the criteria of FDR q value < 0.01, the number of DEGs between melanoma and nevus, melanoma and normal skin, and melanoma and melanocyte were 720, 801, and 335, respectively, in GSE3189 and GSE31879 data sets. The intersection of these three comparisons was 105, which were differently expressed in all three comparisons (Figure 1A). Among 105 genes, 47 were confirmed as DEGs in GSE15605 (Figure 1B). GO functional enrichment analysis showed that the changes in 47 genes were mainly related to the biological processes of neutrophil and autophagy (Figure 1C). KEGG pathway enrichment analysis demonstrated that these 47 DEGs were significantly involved in lysosome, B-cell receptor signaling pathways, T-cell receptor signaling pathway, Fc epsilon RI signaling pathway, and so on (Figure 1D).

3.2. Identification of prognostic gene signature

Twelve out of 47 DEGs were statistically significant in predicting the OS in univariate Cox proportional hazards models in TCGA data set (Figure 2). Among them, *BLNK*, *DAPP1* and *IL18* were downregulated, and other genes were upregulated in melanoma samples compared to normal samples. These genes were then fitted in the LASSO model, and ten of them that were predictive of OS were selected (Figure 3A). These genes include *BLNK*, *CD81*, *CLEC4E*, *CPPED1*, *IL18*, *ISG20*, *LCP2*, *MGRN1*, *RAB5C,* and *TRIM32*. Our results revealed that elevated expression of CD81, IL18, MGRN1, RAB5C, and TRIM32 were associated with worse OS for melanoma, and elevated expression of BLNK, CLEC4E, CPPED1, ISG20, and LCP2 were associated with better prognosis for melanoma. In addition, there were strong correlations among BLNK, IL18, CLEC4E, and LCP2 (Figure 3B). To investigate whether the prognostic value of the identified ten-gene signature was independent of clinical characteristics, we employed univariate and multivariate Cox regression models to consider the covariates of age, sex, sample type, tumor stage, Breslow thickness, and ulceration in TCGA data set. Since sex was not statistically significant in univariate Cox regression models for predicting both OS and DFS, it was excluded when fitting multivariate Cox regression models. Results showed that PIS was statistically significant for predicting OS and DFS after adjusting for effects of several clinical information (Figure 3C and D & Tables S2 and S3).

3.3. Construction and validation of PIS

The PIS was calculated based on the linear combination of selected ten genes and corresponding coefficients, and patients with melanoma in TCGA data set were divided into high-risk group (N = 235) and low-risk group (N = 236). The OS of high-risk group was significantly worse than that of low-risk group (log-rank P = 3e-10, HR: 2.343, 95% CI:

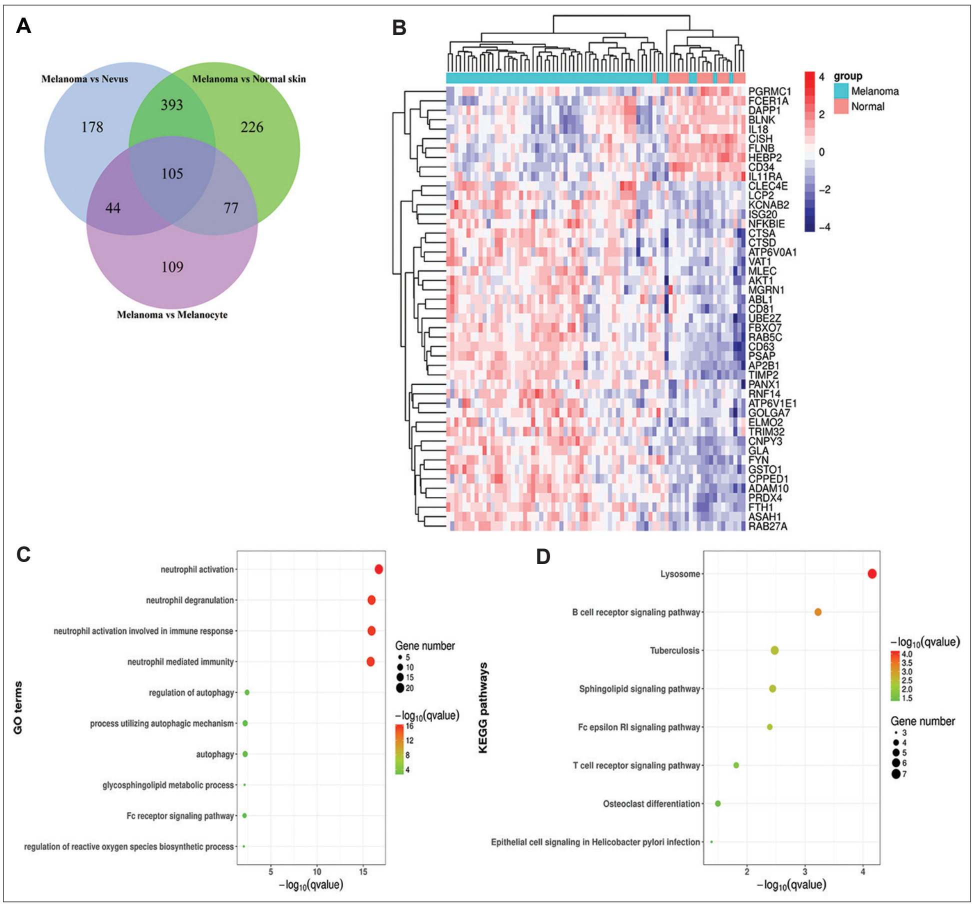


Figure 1. Identification of DEGs. (A) Overlapping DEGs shared between melanoma and benign skin nevus, melanoma and normal skin tissue, and melanoma and normal melanocyte. (B) Heatmap of DEGs. (C) GO functional enrichment analysis (top 10 results). (D) KEGG pathway enrichment analysis (top 10 results).
DEGs: Differentially expressed genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene ontology.

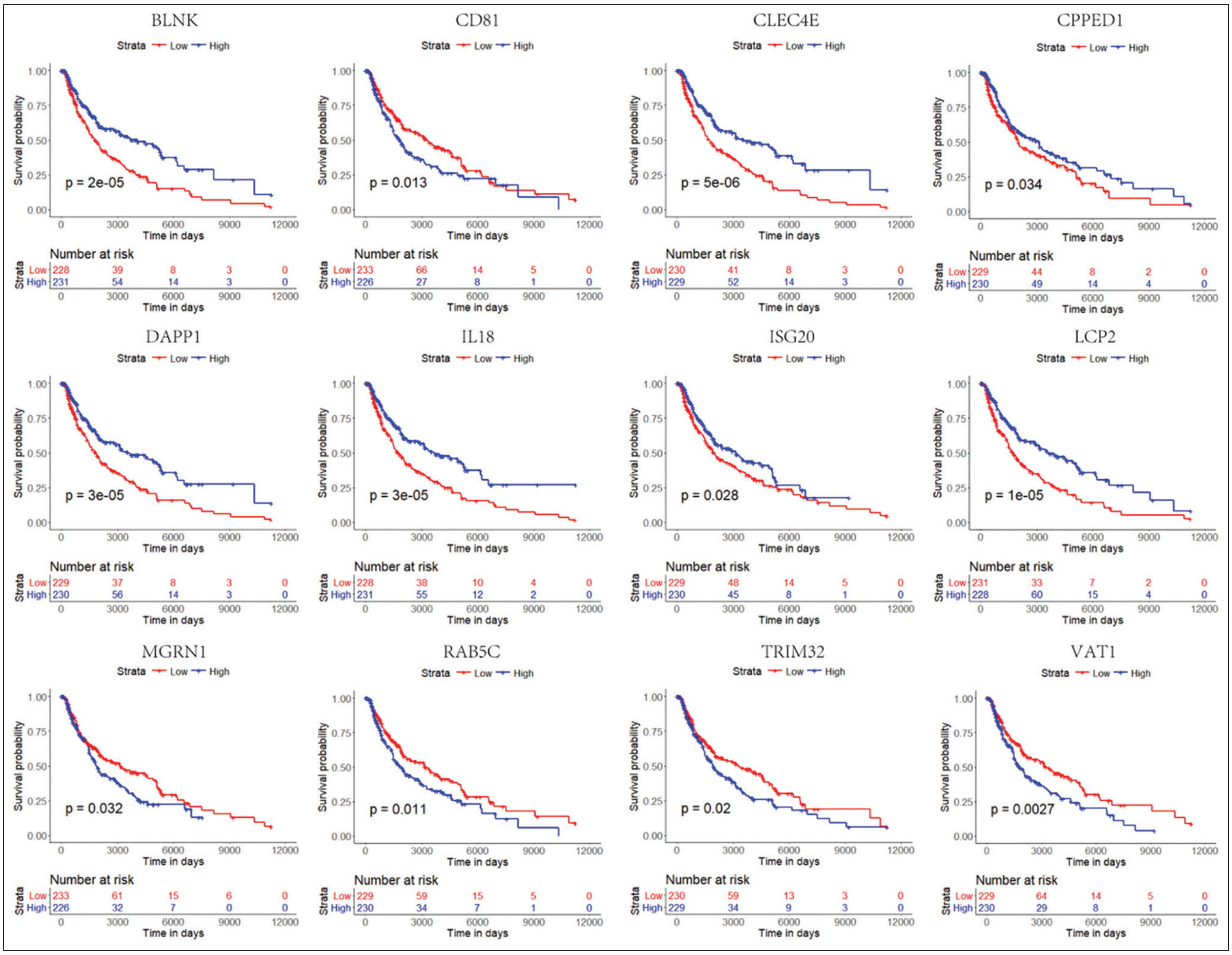


Figure 2. Univariate survival analysis for DEGs in TCGA data set.
DEGs: Differentially expressed genes; TCGA: The Cancer Genome Atlas.

1.785 – 3.076, Figure 4A), with the median survival time of 1628 and 4601 days, respectively. Meanwhile, our results showed that the DFS of high-risk group was significantly worse than that of low-risk group (log-rank $P = 8e\text{-}07$, HR: 2.895, 95% CI: 1.867 – 4.489, Figure 4B), with the median survival time of 2195 and 6299 days, respectively. To explore the robustness of ten-gene signature for predicting the survival of melanoma, we externally validated the prognostic ability of ten-gene signature in an independent GSE65904 data set. Results showed that the OS of high-risk patients in GSE65904 was significantly worse than that of low-risk patients (log-rank $P = 0.0036$, HR: 1.786, 95% CI: 1.202 – 2.653 days, Figure 4C), with the median survival time of 629 and 1404, respectively. In addition, patients in high-risk group had significantly worse DMFS than those in low-risk group (log-rank $P = 0.0082$, HR: 1.786, 95% CI: 1.155 – 2.762, Figure 4D), with the median survival time of 379 and 1401 days, respectively.

3.4. Associations between identified prognostic genes and TAL subsets

The associations between the selected ten prognostic genes and the specific TAL were calculated using Spearman's rank correlation. Genes that had the largest absolute correlation coefficients with certain TAL subset among all TAL subsets were considered to have strong association with that TAL subset. Results showed that LCP2, ISG20, and CLEC4E expression were significantly positively correlated with $CD8^+$ T-cells. CD81 and MGRN1 expression were significantly positively correlated with Treg cells, while BLNK and IL-18 expression were significantly negatively correlated with M0 macrophages. In addition, RAB5C and TRIM32 expression had strong association with B-cells and natural killer cells, respectively (Figure 5). Given the important role of LCP2 in the immune system, we focused our efforts on investigating LCP2 in the subsequent studies.

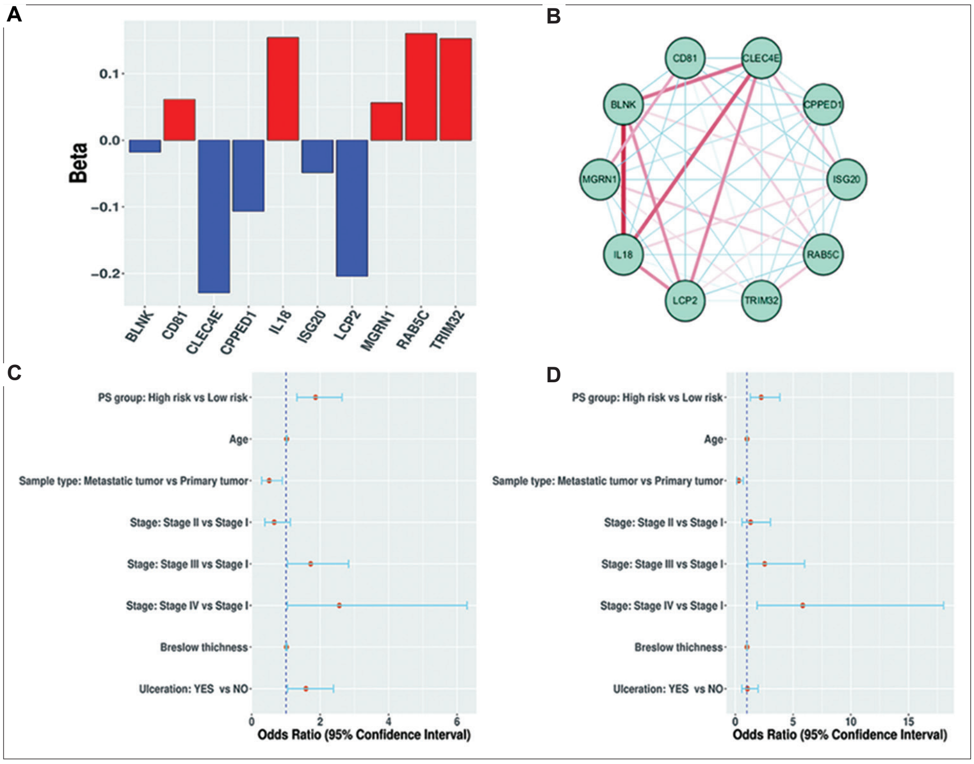


Figure 3. Identification of prognostic gene signature. (A) The identified prognostic gene signature in LASSO model. (B) Correlations among identified prognostic genes. (C) Multivariate analysis for the association between PIS and OS after adjusting for covariates. (D) Multivariate analysis for the association between PIS and DFS after adjusting for covariates.

PIS: Prognostic immune score; OS: Overall survival.

3.5. LCP2 inhibited the growth of melanoma possibly through the positive regulation of IRF5 signaling pathway to activate CD8$^+$ T-cells

To determine the function of LCP2 in melanoma, we established the LCP2 knockdown mouse tumor model. At first, we validated the knockdown efficacy of LCP2 in B16F10 cell through RT-qPCR (Figure 6A). It was obvious to see that knockdown of LCP2 significantly promoted the tumor growth (Figure 6B–D) in wild-type mice but had no effect on tumor sizes in nude mice, indicating the negative correlation between LCP2 and melanoma growth, as well as the importance of LCP2 in the tumor immune microenvironment (Figure 6E). Furthermore, flow cytometry analysis of shLCP2 melanoma cells showed no significantly changes of CD4$^+$ and CD8$^+$ T-cells, but remarkable reduction of Granzyme B-positive cells, which implied that LCP2 loss decreased the activity of CD8$^+$ T-cells and led to the promotion of tumor growth (Figure 6F–J). Moreover, to further explore the underlying mechanism of LCP2 regulating the activity of CD8$^+$ T-cells, DEGs between shLCP2 group and control group were analyzed. In total, 846 DEGs were found with $P < 0.05$ and $|\text{foldchange}| > 1$, among which 407 DEGs were upregulated and 439 DEGs were downregulated. The volcano plot of DEGs showed that the IRF5 was significantly downregulated in the shLCP2 group (Figure 6K). All the DEGs in this part are listed in **Table S4**. Similar result was also noted in RT-qPCR experiment (Figure 6L), suggesting that LCP2 may positively regulate the IRF5 signaling pathway to activate CD8$^+$ T-cells and then inhibit the growth of melanoma (Figure 6M).

4. Discussion

Melanoma is known to be the most lethal type of skin cancer. The advent of immunotherapy has revolutionized the status of clinical therapies of melanoma, which brought new hope to these patients. However, only a small proportion of patients are responders. Therefore, novel prognostic and immune-related biomarkers are urgently needed to guide the development of melanoma treatments. In the present study, RNA-seq data of cutaneous melanoma from the public database were used for identifying prognostic gene signatures. KEGG pathway enrichment analysis showed that 47 DEGs were significantly involved in lysosome, B-cell receptor signaling pathways, T-cell receptor signaling pathway, and Fc epsilon RI signaling pathway, indicating that these signaling pathways play an important role in melanoma.

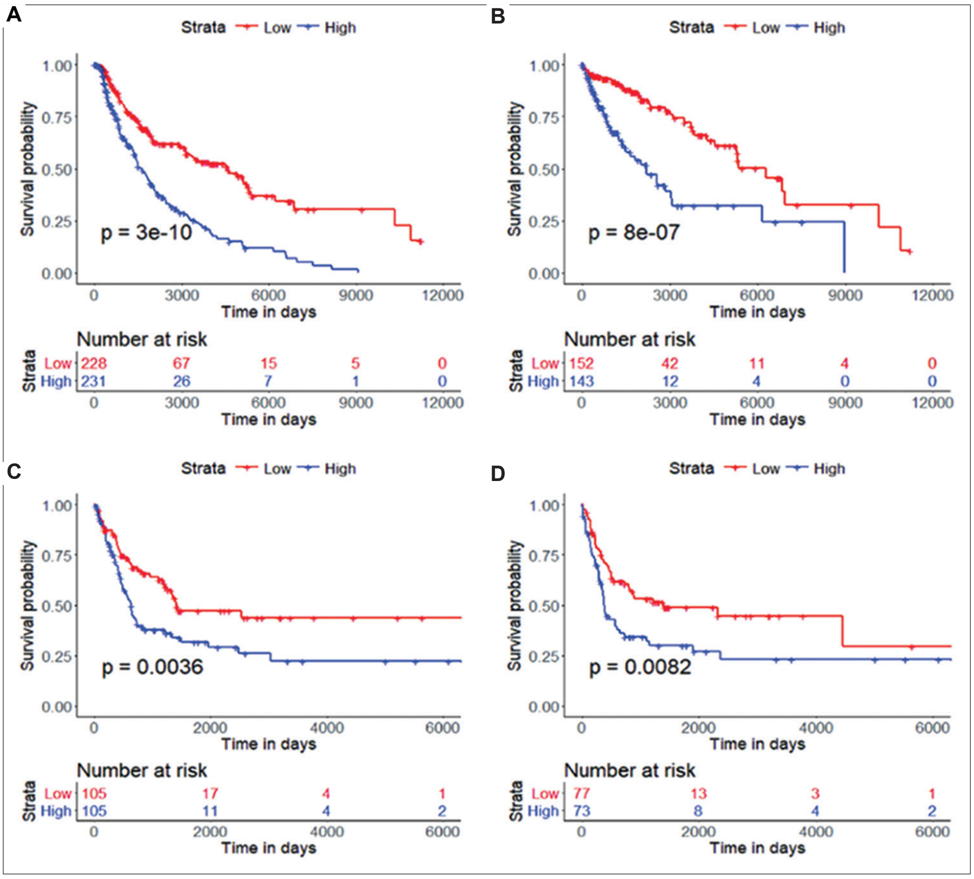


Figure 4. Survival curves for high- and low-risk patients based on PIS. (A) Survival curves for DMFS in GSE65904 data set. (B) Survival curves for OS in GSE65904 data set. (C) Survival curves for DFS in TCGA data set. (D) Survival curves for OS in TCGA data set.
PIS: Prognostic immune score; OS: Overall survival; DMFS: Distant metastasis-free survival; DFS: Disease-free survival; TCGA: The Cancer Genome Atlas.

Furthermore, ten (*BLNK*, *CD81*, *CLEC4E*, *CPPED1*, *IL18*, *ISG20*, *LCP2*, *MGRN1*, *RAB5C*, and *TRIM32*) out of the 47 DEGs were statistically significant in predicting the OS in univariate Cox proportional hazards models and LASSO model in TCGA data set. Our results revealed that elevated expression of CD81, IL18, MGRN1, RAB5C, and TRIM32 were associated with worse OS in melanoma patients, and elevated expression of BLNK, CLEC4E, CPPED1, ISG20, and LCP2 were associated with better prognosis in melanoma patients. The PIS was calculated based on the linear combination of the expression of selected ten genes, and the PIS was statistically significant in predicting OS and DFS in melanoma patients after adjusting for several clinical covariates.

At present, numerous studies have reported the important roles of various tumor-infiltrating immune cells in melanoma[25-28]. To investigate the association between the selected ten prognostic genes and tumor-infiltrating immune cells, we applied CIBERSORT to TCGA data set and obtained 22 distinct TAL subsets, and used Spearman's rank correlation to analyze the correlation between each identified prognostic gene and the specific TAL. We found that these identified prognostic genes were mainly correlated with CD8$^+$ T cells, Treg cells, B cells, M0 macrophages, and natural killer cells. It is worth noting that the expression level of LCP2 was positively associated with CD8$^+$ T-cells.

A few studies have shown that LCP2 was differentially expressed in many different types of tumors and was correlated with prognosis in numerous cancers[10-12]. Moreover, LCP2 was reported essential for normal T-cell development and activation[14]. In our present study, the bioinformatics analyses revealed that high LCP2 expression in melanoma patients was statistically significant for predicting the OS in univariate Cox proportional hazards models in TCGA data set and was associated with better prognosis in melanoma patients. In addition, we found that LCP2 expression was significantly positively correlated with CD8$^+$ T-cells, indicating that LCP2 could be a potential immunotherapy target. Moreover, given the important role of LCP2 in the immune system as reported in numerous published studies, we hence focused our efforts on investigating LCP2 in the subsequent studies. The *in vivo*

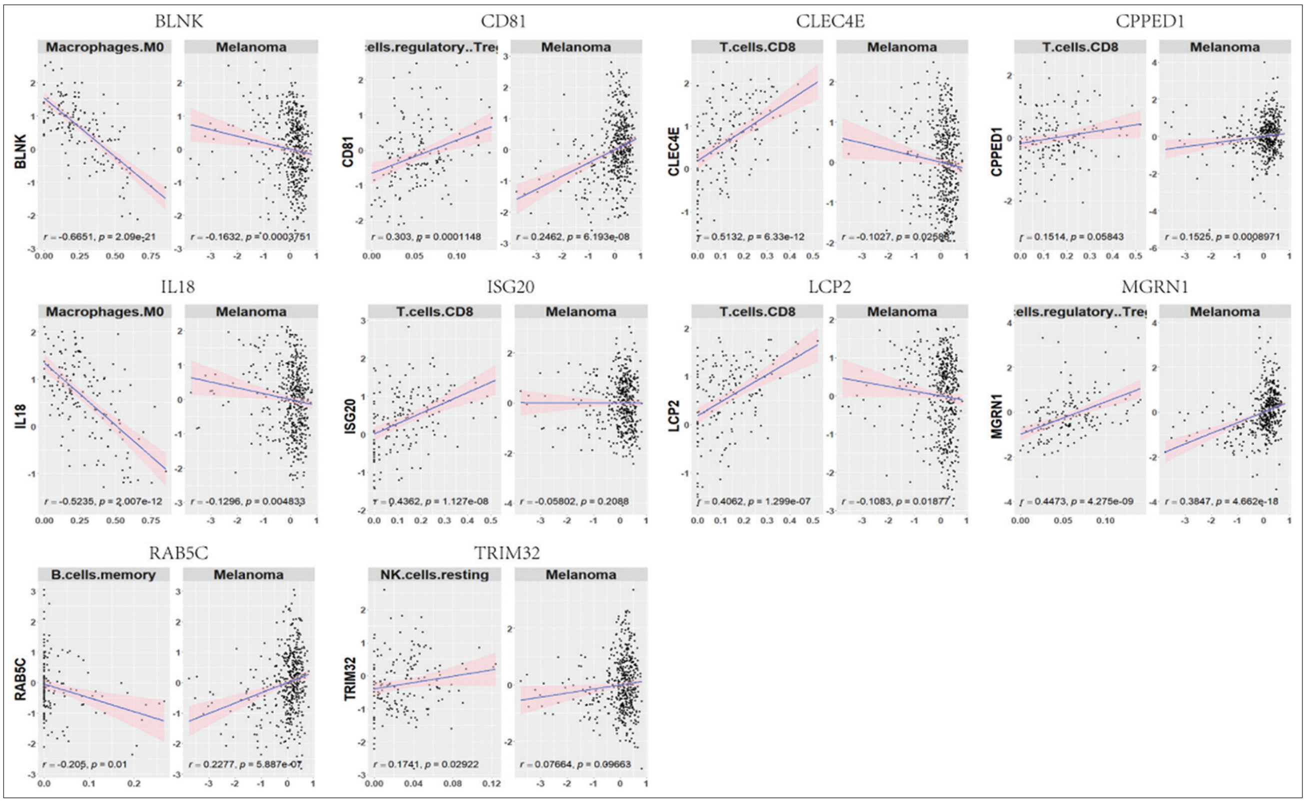


Figure 5. Association between identified prognostic genes and TAL subsets and melanoma.
TAL: Tumor-associated leukocyte.

experiment showed that knockdown of LCP2 significantly promoted the progression of melanoma in wild-type mice but had no effect on tumor sizes in nude mice, indicating the negative correlation between LCP2 and melanoma growth, as well as the importance of LCP2 in the tumor immune microenvironment. The result of flow cytometry analysis directly confirmed the positive correlation between LCP2 expression and CD8$^+$ T-cell activity. However, the pathway by which LCP2 regulates T-cell activity is unknown.

To explore the mechanism through which LCP2 regulates T-cell activity, DEGs of shLCP2 group versus control group were analyzed in this study, and we found that IRF5 expression was significantly downregulated. Interferon regulatory factors (IRFs) are transcriptional mediators of type 1 interferon signaling pathway induced by pathogens and viruses, which are also involved in immune response, apoptosis, cell growth regulation, and oncogenesis[29-31]. Some members of the IRFs have been shown to be involved in regulating development and function of the immune cells[32-34]. IRF5, as a member of IRFs, has been reported to play a key role in immune regulation of autoimmune diseases and inflammatory responses, which therefore attracted much attention in recent years. In addition, it has been revealed that IRF5 is involved in the activation of inflammatory cytokines and chemokines that are important for recruitment of T-cells[35,36]. Moreover, the intratumoral IRF5 deletion in breast cancer cells dysregulated secretion of many chemokines and cytokines, leading to the inaccurate and untimely transport of immune cells to the tumor sites[37]. Takaoka *et al.* found that Irf5 knockout mice were unable to secrete proinflammatory cytokines (such as tumor necrosis factor α [TNF-α], IL-6, and IL-12), and exhibited resistance toward endotoxic shock induced by lipopolysaccharide, indicating that IRF5 was generally involved in the downstream of TLR-MyD88 signaling pathway and induced the gene expression of proinflammatory cytokines[38]. Moreover, studies have shown that most of the chemokines induced by virus in cells showed lymphocyte chemotactic activity in the case of overexpression IRF5[35], which therefore were significant for T-lymphocyte recruitment, indicating a possible key role of IRF5 in lymphocyte trafficking. Therefore, we speculated the key bridge linking of IRF5 between LCP2 and CD8$^+$T-cell. In our study, the expression of IRF5 was remarkably decreased in LCP2 knockdown B16F10 cells. Taking together, we speculated that high expression of LCP2 in melanoma upregulated the IRF5 expression and IRF5 then mediated the secretion of proinflammatory

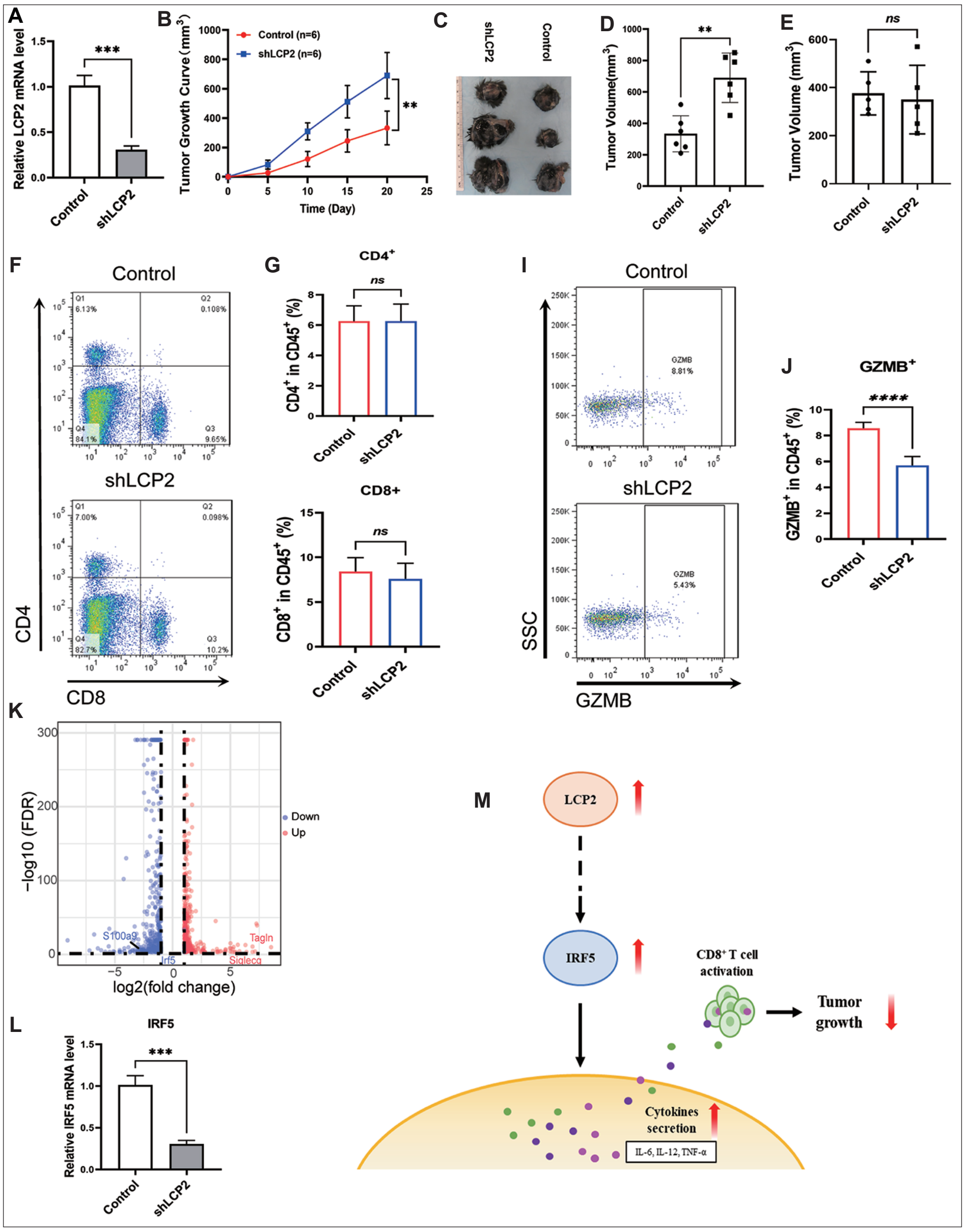


Figure 6. Analysis of underlying mechanism of LCP2 knockdown suppressing the growth of melanoma. (A) RT-qPCR analyses of LCP2 in control and LCP2 knockdown (shLCP2) B16F10 cells (n = 3). (B) B16F10 cells and shLCP2 B16F10 cells (10^6) were injected subcutaneously into C57BL/6 mice. Tumor volume was measured every 5 days and calculated using this formula: (length×width^2)/2. (C) Representative tumors were resected and photographed at day 20. Tumor volume in C57BL/6 mice (D) and in BALB/c nude mice € were recorded at day 20 (*n* = 6 in C57BL/6 mice group, *n* = 5 in BALB/c nude mice group). Flow cytometry analyses (F) and quantitative analyses of CD4$^+$ (G) and CD8$^+$ cells (H), and flow cytometry (I) and quantitative analysis of Granzyme B-positive (GZMB$^+$) cells (J) in the melanoma tumor from immunocompetent wild-type BALB/c mice with subcutaneous injection of shLCP2 and control B16F10 cells. (K) Volcano plot of DEGs shared between control group and shLCP2 group (significantly upregulated genes are denoted by red nodes; significantly downregulated genes are denoted by blue nodes). (L) RT-qPCR analyses of IRF5 in control and shLCP2 B16F10 cells (*n* = 3). (M) A schematic diagram showing the hypothetical mechanism of high LCP2 expression positively regulating IRF5 signaling pathway to activate tumor-infiltrating CD8$^+$ T cells and then inhibit the growth of melanoma. Tukey's multiple comparison or Student's *t*-test was used to calculate *p* values for results shown in this figure.

ns: Not significan; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; LCP2: Lymphocyte cytosolic protein 2; RT-qPCR: Real time-quantitative polymerase chain reaction; DEGs: Differentially expressed genes; IRF5: Interferon regulatory factor 5.

cytokines and chemokines to activate CD8$^+$ T lymphocytes, which finally suppressed the growth of melanoma.

5. Conclusions

LCP2 is a novel prognostic biomarker of melanoma, which could positively regulate the IRF5 signaling pathway to activate tumor-infiltrating CD8$^+$ T-cells and then inhibit the growth of melanoma. Nevertheless, it is important to point out that there are some limitations in our study. First, the potential association between LCP2 and IRF5 is unclear, which needs to be systematically and deeply explored *in vitro* and *in vivo*. Second, more public data on melanoma patients who received immunotherapy should be included for bioinformatics analysis in this study to investigate prognostic significance of LCP2 in these patients.

Acknowledgments

None.

Funding

This study was supported by grants from General Program, the National Natural Science Foundation of China (No. 81874138, 82073020 and 81903222).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Mingzhu Yin
Data curation: Hongyin Sun, Kui Deng
Investigation: Hongyin Sun, Kui Deng
Validation: Xingchen Zhou, Dongsheng Cao, Yan Cheng
Writing – original draft: Hongyin Sun, Kui Deng, Mingzhu Yin
Writing – review & editing: Xiang Chen, Mingzhu Yin

Ethics approval and consent to participate

All the procedures of animal study were evaluated and approved by the Institutional Animal Care and Use Committee of Xiangya Hospital, Central South University (No. 2022020685).

Consent for publication

Not applicable.

Availability of data

The data used to support the findings of this study are available from the corresponding author on reasonable request.

References

1. Torre LA, Bray F, Siegel RL, *et al.*, 2015, Global cancer statistics, 2012. *CA Cancer J Clin*, 65(2): 87–108.

   https://doi.org/10.3322/caac.21262

2. Schadendorf D, Fisher DE, Garbe C, *et al.*, 2015, Melanoma. *Nat Rev Dis Primers*, 1: 15003.

   https://doi.org/10.1038/nrdp.2015.3

3. Luke JJ, Flaherty KT, Ribas A, *et al.*, 2017, Targeted agents and immunotherapies: Optimizing outcomes in melanoma. *Nat Rev Clin Oncol*, 14(8): 463.

   https://doi.org/10.1038/nrclinonc.2017.43

4. Redman JM, Gibney GT, Atkins MB, 2016, Advances in immunotherapy for melanoma. *BMC Med*, 14(1): 20.

   https://doi.org/10.1186/s12916-016-0571-0

5. D'Arcangelo D, Giampietri C, Muscio M, *et al.*, 2018, WIPI1, BAG1, and PEX3 autophagy-related genes are relevant melanoma markers. *Oxidat Med Cell Longev*, 2018: 1471682.

   https://doi.org/10.1155/2018/1471682

6. Gerami P, Cook RW, Wilkinson J, *et al.*, 2015, Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. *Clin Cancer Res*, 21(1): 175–183.

   https://doi.org/10.1158/1078-0432.CCR-13-3316

7. Zhao Y, Schaafsma E, Gorlov IP, *et al.*, 2018, A leukocyte infiltration score defined by a gene signature predicts melanoma patient prognosis. *Mol Cancer Res*, 17(1): 109filtra.

   https://doi.org/10.1158/1541-7786.MCR-18-0173

8. Clements JL, Ross-Barta SE, Tygrett LT, *et al.*, 1998, SLP-76 expression is restricted to hemopoietic cells of monocyte, granulocyte, and T lymphocyte lineage and is regulated during T cell maturation and activation. *J Immunol*, 161: 3880–3889.

9. Jumaa H, Wollscheid B, Mitterer M, *et al.*, 1999, Abnormal development and function of B lymphocytes in mice deficient for the signalling adaptor protein SLP-65. *Immunity*, 11: 547–554.

   https://doi.org/10.1016/s1074-7613(00)80130-2

10. Zhang J, Wang L, Xu X, *et al.*, 2020, Transcriptome-based network analysis unveils eight immune-related genes as molecular signatures in the immunomodulatory subtype of triple negative breast cancer. *Front Oncol*, 10: 1787.

    https://doi.org/10.3389/fonc.2020.01787

11. Huo Y, Zhang K, Han S, *et al.*, 2021, Lymphocyte cytosolic protein 2 is a novel prognostic marker in lung adenocarcinoma. *J Int Med Res*, 49(11): 3000605211059681.

    https://doi.org/10.1177/03000605211059681

12. Wang Z, Peng M, 2021, A novel prognostic biomarker LCP2

correlates with metastatic melanoma-infiltrating CD8 T cells. *Sci Rep*, 11(1): 9164.

https://doi.org/10.1038/s41598-021-88676-9

13. Koretzky GA, Abtahian F, Silverman MA, 2006, SLP76 and SLP65: Complex regulation of signalling in lymphocytes and beyond. *Nat Rev Immunol*, 6(1): 67–78.

    https://doi.org/10.1038/nri1750

14. Navas VH, Cuche C, Alcover A, *et al.*, 2017, Serine phosphorylation of SLP76 is dispensable for T cell development but modulates helper T cell function. *PLoS One*, 12(1): e0170396.

    https://doi.org/10.1371/journal.pone.0170396

15. Sommers CL, Menon RK, Grinberg A, *et al.*, 2001, Knock-in mutation of the distal four tyrosines of linker for activation of T cells blocks murine T cell development. *J Exp Med*, 194(2): 135–142.

    https://doi.org/10.1084/jem.194.2.135

16. Maltzman JS, Kovoor L, Clements JL, *et al.*, 2005, Conditional deletion reveals a cell-autonomous requirement of SLP-76 for thymocyte selection. *J Exp Med*, 202(7): 893–900.

    https://doi.org/10.1084/jem.20051128

17. Siggs OM, Miosge LA, Daley SR, *et al.*, 2015, Quantitative reduction of the TCR adapter protein SLP-76 unbalances immunity and immune regulation. *J Immunol*, 194(6): 2587–2595.

    https://doi.org/10.4049/jimmunol.1400326

18. Fabregat A, Jupe S, Matthews L, *et al.*, 2017, The reactome pathway knowledgebase. *Nucleic Acids Res*, 46(D1): D649–D655.

    https://doi.org/10.1093/nar/gkx1132

19. Hastie T, Tibshirani R, Narasimhan B, *et al.*, 2016, Impute: Imputation for microarray data. *Oral History Rev*, 2016.

20. Benjamini Y, Hochberg Y, 1995, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*, 57(1): 289–300.

    https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

21. Tang Z, Li C, Kang B, *et al.*, 2017, GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*, 45(W1): W98–W102.

    https://doi.org/10.1093/nar/gkx247

22. Yu G, Wang LG, Han Y, *et al.*, 2012, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, 16(5): 284–287.

    https://doi.org/10.1089/omi.2011.0118

23. Friedman J, Hastie T, Tibshirani R, 2009, Glmnet: Lasso and Elastic-net Regularized Generalized Linear Models. R Package Version, 1(4).

24. Newman AM, Liu CL, Green MR, *et al.*, 2015, Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 12(5): 453–457.

    https://doi.org/10.1038/nmeth.3337

25. Balatoni T, Mohos A, Papp E, *et al.*, 2018, Tumor-infiltrating immune cells as potential biomarkers predicting response to treatment and survival in patients with metastatic melanoma receiving ipilimumab therapy. *Cancer Immunol Immunother*, 67(1): 141–151.

    https://doi.org/10.1007/s00262-017-2072-1

26. Liu D, 2019, CAR-T "the living drugs", immune checkpoint inhibitors, and precision medicine: A new era of cancer therapy. *J Hematol Oncol*, 12(1): 113.

    https://doi.org/10.1186/s13045-019-0819-1

27. Dabrosin N, Juul KS, Georgsen JB, *et al.*, 2019, Innate immune cell infiltration in melanoma metastases affects survival and is associated with BRAFV600E mutation status. *Melanoma Res*, 29(1): 30–37.

    https://doi.org/10.1097/CMR.0000000000000515

28. Fu X, Yin M, 2022, Monocytes in tumor: The perspectives of single-cell analysis. *Tumor Discov*, 1(1): 4.

    https://doi.org/10.36922/td.v1i1.4

29. Honda K, Taniguchi T, 2006, IRFs: Master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol*, 6(9): 644–658.

    https://doi.org/10.1038/nri1900

30. Honda K, Takaoka A, Taniguchi T, 2006, Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. *Immunity*, 25(3): 349–360.

    https://doi.org/10.1016/j.immuni.2006.08.009

31. Tamura T, Yanai H, Savitsky D, *et al.*, 2008, The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol*, 26: 535–584.

    https://doi.org/10.1146/annurev.immunol.26.021607.090400

32. Lohoff M, Mak TW, 2005, Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol*, 5(2): 125–135.

    https://doi.org/10.1038/nri1552

33. Battistini A, 2009, Interferon regulatory factors in hematopoietic cell differentiation and immune regulation. *J Interferon Cytokine Res*, 29(12): 765–780.

    https://doi.org/10.1089/jir.2009.0030

34. Zhao GN, Jiang DS, Li H, 2015, Interferon regulatory factors: At the crossroads of immunity, metabolism, and disease. *Biochim Biophys Acta Mol Basis Dis*, 1852(2): 365–378.

    https://doi.org/10.1016/j.bbadis.2014.04.030

35. Barnes BJ, Kellum MJ, Field AE, *et al.*, 2002, Multiple

regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes. *Mol Cell Biol*, 22(16): 5721–5740.

https://doi.org/1128/MCB.22.16.5721-5740.2002

36. Kaur A, Lee LH, Chow SC, *et al*., 2018, IRF5-mediated immune responses and its implications in immunological disorders. *Int Rev Immunol*, 37(5): 229–248.

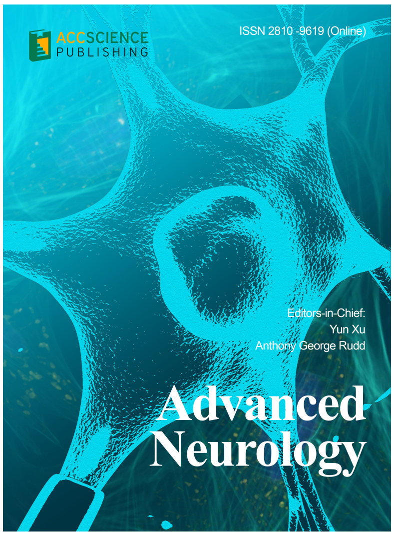
https://doi.org/10.1080/08830185.2018.146962

37. Pimenta EM, De S, Weiss R, *et al*., 2015, IRF5 is a novel regulator of CXCL13 expression in breast cancer that regulates CXCR5(+) B-and T-cell trafficking to tumor conditioned media. *Immunol Cell Biol*, 93(5): 486–499.

https://doi.org/10.1038/icb.2014.110

38. Takaoka A, Yanai H, Kondo S, *et al*., 2005, Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature*, 434(7030): 243–249.

https://doi.org/10.1038/nature03308


ACCSCIENCE PUBLISHING
ISSN 2810-9619 (Online)
Editors-in-Chief:
Yun Xu
Anthony George Rudd
Advanced Neurology


Global Translational Medicine
Editor-in-Chief
Prof. Lemin Zheng
ACCSCIENCE PUBLISHING

https://accscience.com/journal/TD

ACCSCIENCE
PUBLISHING

Contact
www.accscience.com
8 Burn Road, #15-03 Trivex, Singapore 369977
Email: editorial@accscience.com
Phone: +65 8182 1586