ISSN 2310-5674

PREMIER Publishing

The European Journal of Biomedical and Life Sciences

Premier Publishing s.r.o.

2023

1 2 3 4

European Journal of Biomedical and Life Sciences

№ 1 2023

PREMIER Publishing

Vienna
2023

European Journal of Biomedical and Life Sciences

Scientific journal

№ 1 2023

ISSN 2310-5674

Proofreading Kristin Theissen
Cover design Andreas Vogel
Additional design Stephan Friedman
Editorial office Premier Publishing s.r.o.
Praha 8 – Karlín, Lyčkovo nám. 508/7, PSČ 18600
E-mail: pub@ppublishing.org
Homepage: ppublishing.org

European Journal of Biomedical and Life Sciences is an international, German/English/Russian language, peer-reviewed journal. The journal is published in electronic form.

The decisive criterion for accepting a manuscript for publication is scientific quality. All research articles published in this journal have undergone a rigorous peer review. Based on initial screening by the editors, each paper is anonymized and reviewed by at least two anonymous referees. Recommending the articles for publishing, the reviewers confirm that in their opinion the submitted article contains important or new scientific results.

Premier Publishing s.r.o. is not responsible for the stylistic content of the article. The responsibility for the stylistic content lies on an author of an article.

Instructions for authors

Full instructions for manuscript preparation and submission can be found through the Premier Publishing s.r.o. home page at: http://www. ppublishing.org.

Material disclaimer

The opinions expressed in the conference proceedings do not necessarily reflect those of the Premier Publishing s.r.o., the editor, the editorial board, or the organization to which the authors are affiliated.

Premier Publishing s.r.o. is not responsible for the stylistic content of the article. The responsibility for the stylistic content lies on an author of an article.

Included to the open access repositories:

The journal has the GIF impact factor 0.802 for 2021.

Typeset in Berling by Ziegler Buchdruckerei, Linz, Austria.
Printed by Premier Publishing s.r.o., Vienna, Austria on acid-free paper.

Section 1. Life Science

https://doi.org/10.29013/ELBLS-23-1-3-9

Mingxiao Ma,
Dr. Roberto Aguilar,
Gaston Day School

USING HIGH-THROUGHPUT SNP TECHNOLOGIES TO IDENTIFY VARIANCES ASSOCIATED WITH BLADDER CANCER

Abstract. According to the CDC, about 57,000 men and 18,000 women have bladder cancer, and about 12,000 men and 4,700 women die from it each year in the US. These high penetrance traits and the identification of the bladder cancer-predisposing gene *CCL4,* warrant further investigation. In this study, we looked at single nucleotide polymorphisms (SNPs) and their association with bladder cancer. Whole genome sequencing data was obtained from the Sequence Read Archive and then analysis pipelines were constructed to map the sequences against chromosome 17 followed by indexing and variant calling. Through our analytical tools, we were able to detect variants in bladder cancer tissues that were not present in normal sequence samples. These findings help support previous studies that link CCL4 with genetic susceptibility to bladder cancer.

Keywords: bladder, cancer, CCL4, SNPs.

Introduction

Bladder Cancer

Bladder cancer is a common cancer that occurs when the cells within the bladder hyper proliferate and form tumors which can then metastasize to other parts of the body. This type of cancer commonly begins in the innermost lining of the bladder (urothelium or transitional epithelium). There are different types of bladder cancer which include: 1) urothelial carcinoma, considered a transitional cell carcinoma and is the most common one, 2) squamous cell carcinoma, the subtype encountered only 1–2% of the time, 3) adenocarcinoma, less common at 1%, and 4) small cell carcinoma (less than 1%), and 5) sarcoma (less than 1%) (American Cancer Society, 2019).

According to the CDC, about 57,000 men and 18,000 women have bladder cancer, and about 12,000 men and 4,700 women die from the disease each year in the US. According to American Cancer Society data in 2022, there will be about 81,180 new cases of bladder cancer and 17,100 deaths from bladder cancer. Figure 1 shows the number of bladder cancer cases from 1999 to 2019. As you can see, the number increases in a relative linear relation up to the 60,000 which was reached in 2000. These data seem to support the evidence that bladder cancer is one of the most common cancers in the United States.

Stages

Bladder cancer progression is separated into stages. The staging system used most often is called the Tumor Nodes Metastases System which is based on 3 keys: "T"

used to describe the distance from the main (primary) tumor has grown through the bladder wall to nearby tissues, "N" used to describe any cancer cells spread to lymph nodes around the bladder, "M" usually indicates if the cancer cells have reached other sites such as distant organs (American Cancer Society, 2019).

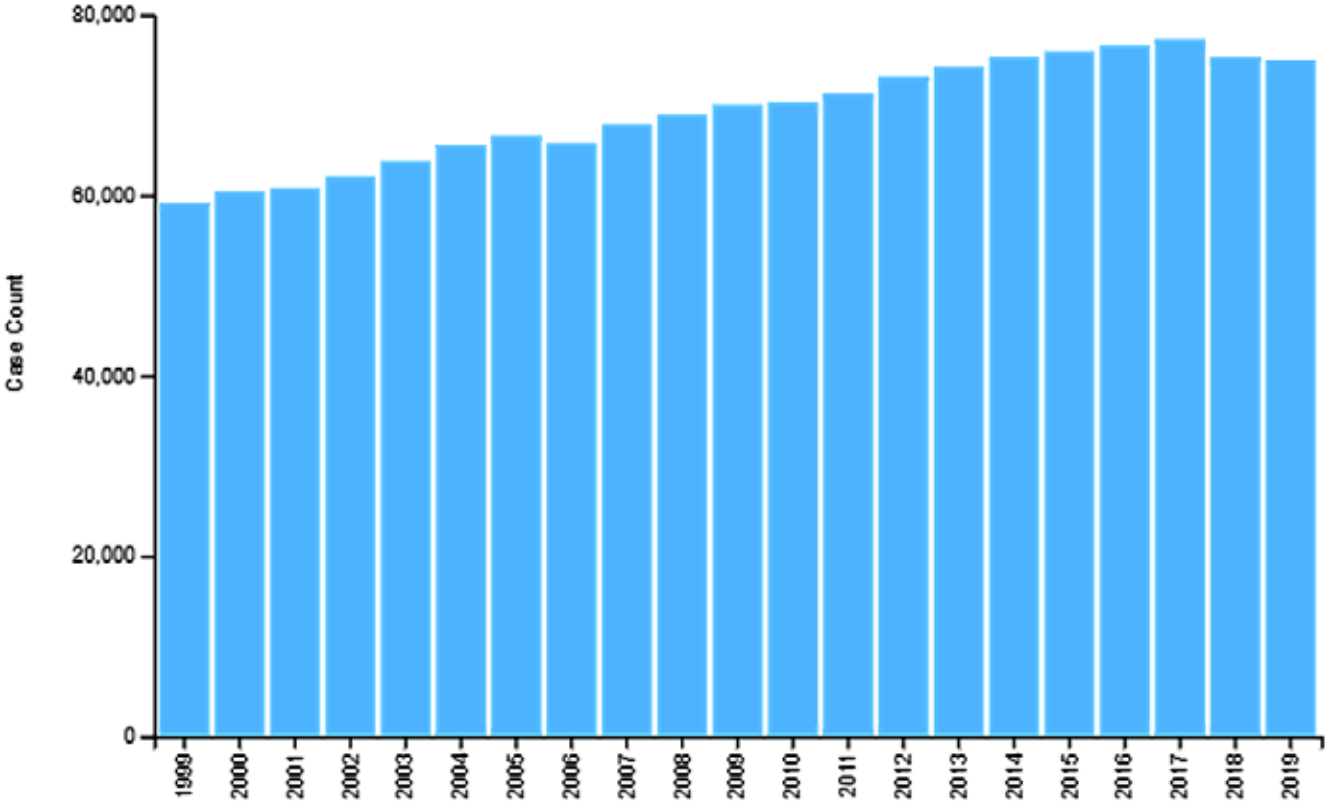


Figure 1: Annual number of new bladder cancer cases from 1999 to 2019 in the United States. This graph shows overall data of new bladder cancer cases from 1999 to 2019. By the year 1999, the case number had already reached 60,000 cases and it kept increasing every year. Obtained from the U. S. Cancer Statistics Working Group. U. S. Cancer Statistics Data Visualizations Tool (2022).

However, there is another numbering system which is not often used that ranges from stage 0 to stage 4. Stage 0 is the earliest stage in which cancer cells are found within the inner layer of the bladder. In stage I, cancer cells can be found growing within the connective tissue beneath the bladder lining. If the cancer cells continue to grow within the connective tissue layer and reach the muscle of the bladder wall, stage II has been reached. Stage III and IV are the two most advanced conditions of bladder cancer, with cancer cells growing through the muscle and into the fat layer, then invading other organs such as the womb, vagina, or the prostate (stage III). In stage IV, the cancer cells can metastasize to nearly all parts of the bladder and then invade other organs (Cancer Research UK, 2018) [5].

Treatment

Treatments of bladder cancer are administered based on the stage it is in. Stage I is most often treated with transurethral resection (TURBT), which is a surgery to remove the tumor from the bladder through the urethra, and intravesical therapy treatment with fulguration 24 hours after surgery (Staff [13]). The side effects of the TURBT include bloody urine, uncomfortable urination, and bladder infection. Intravesical therapy is a major treatment in which the liquid drug is administered into the bladder through a soft catheter and left for up to 2 hours (American Cancer Society [11]). Some common side effects of intravesical therapy include irritation and a burning sensation in the bladder, and bloody urine. The type

of radiation that is often used to combat bladder cancer is named External Beam Radiation Therapy and is used during the early stages of bladder cancer in patients who cannot undergo surgery or chemotherapy due to other complications. This type of radiation therapy can effectively prevent symptoms that are caused by advanced bladder cancer. The possible side effects include: nausea, vomiting, diarrhea, low blood counts, skin changes in the radiation area and bloody urine. Chemotherapy is used to treat bladder cancer where cancer cells are confined to the lining of the bladder and will have a high risk of progression or recurrence to a higher metastasis state. Side effects of chemotherapy include blood in the urine and frequent urination (American Cancer Society [13]).

CCL4

CCL4 (C–C motif chemokine ligand 4) is a gene which is protein-coded located on 17q12, Gene ID6351 (NCBI, 2022). The *CCL4* gene has been found in chimpanzees, rhesus monkeys, dogs, cows, mice, rats, and chickens. The protein encoded by *CCL4* is a mitogen-inducible monokine that is secreted and has chemokinetic and inflammatory functions. Also, It has been known to be one of the major HIV-suppressive factors produced by CD8+ T cells (NCBI [14]). Particularly when it comes to urothelial bladder cancer, one of the most common cancers with a high mortality and recurrence rate, CCL4 plays a role in bladder cancer or its therapy. CCL4 has only been identified as being partially regulated by CC chemokines and being implicated in angiogenesis, tumor development, and metastasis of many malignancies. However, little is known about the function of distinct CC chemokines in bladder cancer.

SNPs

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation among people (National Library of Medicine, [10]). Each SNP denotes a variation in a single nucleotide, or DNA building block. An example of an SNP is the substitution of the nucleotide thymine (T) with the nucleotide cytosine (C) in a specific DNA stretch. SNPs can be used as genetic markers in analysis, and some SNPs located within a gene may directly affect the protein structure or gene expression levels. SNPs are common throughout the genome and can therefore be used in association studies, which postulate that two closely spaced alleles (a gene and a marker) are inherited together, as indications to find disease-causing genes (Shastry [12]). However, SNPs within genes that control telomere maintenance, mitosis, inflammation, and apoptosis, however, have not been thoroughly studied in bladder cancer.In this investigation, we used bioinformatics tools to identify these gene variances within the *CCL4* gene and other regions of chromosome 17. We believe that any variances found in a bladder cancer clinical cohort that are not found in a normal cohort may have an association with the observed phenotypic tendencies of bladder cancer.

Methods

To identify a good candidate gene, we accessed NCBI's Gene Expression Omnibus (GEO) and sifted through microarray data. Once we settled on a published bladder cancer investigation, we took the accession number and entered it into the GEO2R database to identify genes with high expression. Through the careful analysis of each generated graph, the gene *CCL4* was identified as a candidate with high expression in bladder cancer samples but low expression in normal samples. We then accessed the Sequence Read Archive (SRA) database and identified a bladder cancer clinical cohort composed of a total of 105 individuals. Ensembl's FTP database was used to obtain chromosome 17's sequence to be used as a reference. The fastq-dump tool was used to access the SRA database and download single end reads as fastq files which were then quality checked using the Babraham Institute's FastQC program. The resultant HTML file was then evaluated to ascertain whether the bar graphs principally clustered within the green field. Trimmomatic was used to trim any fastq file reads that fell below a Phred score of 40 and then reanalyzed with FastQC. Bowtie 2 was used to map the reads to the reference chromosome 17. The

sequences were then indexed and locally aligned with the Burrows-Wheeler Aligner software package (bwa index and bwa mem). Next, The Sequence Alignment Map (SAM) tools were used to convert SAM output files into compressed Binary Alignment Map (BAM) files and then sort the files according to specific coordinates. BCF tools was used to call and view the detected SNPs, filter, and then convert the files into variant calling format (VCF). Python's Pandas library was used to eliminate any variances found in both the bladder cancer clinical cohort and the normal cohort. The filtered variances were then compared by chromosomal position to NCBI's Entrez Database to identify the variances' SNP accession number. Ensembl's Variant Effect Predictor was then used to identify the genotypic effect of the identified SNPs.

Results

The analysis of chromosome 17 resulted in the identification of hundreds of variants which were common in at least 70% of the individuals in the bladder cancer cohort. After removing those that were also found in the normal population and were not currently characterized as SNPs, only 14 remained. These were then further evaluated using the Variant Effect Predictor which detected the following effects: 1) Intron variant- impacts alternative splicing by interfering with the splice site recognition (Lin [7]). 2) non coding transcript variant- potentially results in the loss of the start or stop codons (Dhamija [5]). 3) nonsense-Mediated mRNA Decay (NMD) transcript variant- targets on the nonsense-mediated mRNA decay function, which under typical circumstances decreases gene expression mistakes by removing mRNA transcripts with early stop codons. 4) regulatory region variant- variant within the non-coding genomic region that has been demonstrated to be associated with different diseases. 5) upstream and downstream gene variants – variants that can affect a gene's regulatory region and may cause damage to the gene. 6) transcription factor (TF) binding site variant- a sequence variant located within a transcription factor binding site which is involved in the regulation of transcription.

As stated earlier, our analysis detected variances within chromosome 17 totaling 13526. After filtering for characterized SNPs that were not found in the normal cohort, we were left with a total of 14 which are listed in Table 1 below. According to VEP results, ten of the 14 SNPs have at least one effect which may be associated with bladder cancer.

Table 1.– Analysis of the bladder cancer cohort identified 14 SNPs with various genetic effects. Table of the gene's location and the SNPs that correspond with and have effects on bladder cancer: Positions on chromosome 17 were entered in the Entrez database to obtain the SNP accession number. Those variants that were not found to be characterized are marked as "None." Variant effects were listed as determined by the Entrez database

Location	SNPs	Effect
1	***2***	***3***
29959924	rs1243604199	Intron variant/Non coding transcript variant/NMD transcript variant/Regulatory region variant
63057821	rs2035858461	**None**
49418233	rs2070798385	**None**
63048721	rs1305782789	Intron variant
63048685	rs1598312390	Intron variant
22521491	rs80284949	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
63034059	rs2034878591	**None**
49421388	rs778329138	**None**

1	*2*	*3*
49421406	rs1438950654	Intron variant/Non coding transcript variant/Upstream gene variant/Downstream gene variant
62998692	rs2033231845	None
22521506	rs80137394	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
22521503	rs77770974	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
12119297	rs1391768914	Intron variant/Non coding transcript variant/NMD transcript variant/Regulatory region variant/Downstream gene variant
22522609	rs201160689	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant

Discussion

Our analysis shows that there is a strong association between the identified 14 SNPs and bladder cancer. Although many variants were identified, not all met the requirements for a common SNP. Still further analysis is needed in order to have a complete picture. Future experiments can involve the analysis of all 23 chromosomes instead of just limiting it to chromosome 17. Further research must also be conducted to correlate these findings with *in vivo* data. Initial DNA can be collected and then purified for sequencing from a clinical cohort. This can be repeated every 10 years as they are simultaneously observed for phenotypic signs of bladder cancer. In order to analyze all 23 chromosomes, at least 30 TB of storage space is required. This coupled with the demand for random access memory places limits on computing power.

Conclusion

Our analyses were able to detect several variances within chromosome 17. After filtering our results for those that were characterized as SNPs on NCBI's Entrez database and those that were not found in normal sequences, we were left with 14 SNPs. Although half were predicted to not have detectable genetic effects, the other half shared genetic consequences ranging from a regulatory region variant to NMD. These SNPs existed in all 105 patients' chromosome 17 and did not exist in normal human mutations. This association presents itself as highly plausible biomarkers that can be used in bladder cancer diagnosis and target treatments.

References:

1. Andrew A. S., Gui J., Sanderson A. C., Mason R. A., Morlock E. V., Schned A. R., Kelsey K. T., Marsit C. J., Moore J. H., & Karagas M. R. (Bladder cancer SNP panel predicts susceptibility and survival. Human genetics. 2009, June). Retrieved August, 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC2763504/#:~: text=Bladder%20cancer%20is%20the%20fourth, assessed%20extensively%20for%20this%20disease
2. Centers for Disease Control and Prevention. Bladder cancer. Centers for Disease Control and Prevention. (2022, July 6). Retrieved August 16, 2022. From: URL: https: //www.cdc.gov/cancer/bladder/index.htm#:~: text=Statistics, women%20die%20from%20the%20disease
3. Database G. C. H.G. (n.d.). CCL4. GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. Retrieved August 16, 2022. From: URL: https: //www.genecards.org/cgi-bin/carddisp.pl?gene=CCL4#:~: text=CCL4%20

Gene%20%2D%20C%2DC%20Motif%20Chemokine%20Ligand%204&text=The%20protein%20encoded%20by%20this, has%20chemokinetic%20and%20inflammatory%20functions
4. Deng N., Zhou H., Fan H., & Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. (2017, November 7). Oncotarget. Retrieved August 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC5746410/
5. Dhamija S., Menon M. B. Non-coding transcript variants of protein-coding genes – what are they good for? RNA Biol.– 15(8). 2018.– P. 1025–1031. Doi: 10.1080/15476286.2018.1511675. Epub 2018. Sep 10. PMID: 30146915; PMCID: PMC6161691.
6. Li Y., Chen X., Li D. et al. Identification of prognostic and therapeutic value of CC chemokines in Urothelial bladder cancer: evidence from comprehensive bioinformatic analysis. BMC Urol – 21, 2021.– 173 p. URL: https: //doi.org/10.1186/s12894-021-00938-w
7. Lin H., Hargreaves K. A., Li R. et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. Genome Biol – 20, 2019.– 254 p. URL: https: //doi.org/10.1186/s13059-019-1847-4
8. Li H., Guo J., Cheng G., Wei Y., Liu S., Qi Y., Wang G., Xiao R., Qi W. and Qiu W. Identification and Validation of SNP-Containing Genes with Prognostic Value in Gastric Cancer via Integrated Bioinformatics Analysis. Front. Oncol.– 11. 2021.– 564296 p. Doi: 10.3389/fonc.2021.564296
9. Mayo Clinic. Bladder cancer. Mayo Clinic. (2022, April 19). Retrieved August 16, 2022. From: URL: https: //www.mayoclinic.org/diseases-conditions/bladder-cancer/diagnosis-treatment/drc-20356109#:~: text=Bladder%20cancer%20treatment%20may%20include, progression%20to%20a%20higher%20stage
10. NHS. (n.d.). Side effects of Chemotherapy. NHS choices. Retrieved August 16, 2022. From: URL: https: //www.nhs.uk/conditions/chemotherapy/side-effects
11. Rojano E., Seoane P., Ranea J. A.G., Perkins J. R. Regulatory variants: from detection to predicting impact. Brief Bioinform. Sep 27,– 20(5). 2019.– P. 1639–1654. Doi: 10.1093/bib/bby039. PMID: 29893792; PMCID: PMC6917219.
12. Shastry B. S. SNPs in disease gene mapping, medicinal drug development and evolution. J Hum Genet – 52, 2007.– P. 871–880. URL: https: //doi.org/10.1007/s10038–007–0200-z
13. Siteman Cancer Center. Effects of treatment for bladder cancer. (2021, July 23). Retrieved August 16, 2022. From: URL: https: //siteman.wustl.edu/treatment/cancer-types/bladder/effects-of-treatment
14. The American Cancer Society medical and editorial content team. (n.d.). What is bladder cancer? American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/about/what-is-bladder-cancer.html
15. The American Cancer Society medical and editorial content team. (2022, January 12). Key statistics for Bladder Cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/about/key-statistics.html
16. The American Cancer Society medical and editorial content team. (n.d.). Bladder cancer staging: Bladder Cancer stages. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/staging.html
17. The American Cancer Society medical and editorial content team. (n.d.). Treatment of bladder cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/by-stage.html#:~: text=Chemotherapy%20followed%20by%20radical%20cystectomy, which%20may%20make%20surgery%20easier.

18. The American Cancer Society medical and editorial content team. (n.d.). Bladder cancer surgery. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/surgery.html
19. The American Cancer Society medical and editorial content team. (n.d.). Intravesical therapy for Bladder Cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/intravesical-therapy.html
20. Transurethral resection of the bladder: Before your surgery. MyHealth.Alberta.ca Government of Alberta Personal Health Portal. (n.d.). Retrieved August 16, 2022. From: URL: https: //myhealth.alberta.ca/Health/aftercareinformation/pages/conditions.aspx?hwid=ug6073#:~: text=Before%20Your%20Surgery-What%20is%20transurethral%20resection%20of%20the%20bladder%3F,)%20or%20malignant%20(cancer).
21. U. S. National Library of Medicine. (n.d.). What are single nucleotide polymorphisms (snps)?: Medlineplus Genetics. MedlinePlus. Retrieved August 16, 2022. From: URL: https: //medlineplus.gov/genetics/understanding/genomicresearch/snp
22. U. S. National Library of Medicine. (n.d.). CCL4 C–C motif chemokine ligand 4 [homo sapiens (human)]. National Center for Biotechnology Information. Retrieved August 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/gene/6351

Section 2. General Biology

https://doi.org/10.29013/ELBLS-23-1-10-18

Jessica Xiong,
High School: Adlai E. Stevenson High School IL, USA
Wei Wang,
Dr. Instructor, Beijing University

HNSCC: DIFFERENTIAL GENE EXPRESSION IN PRIMARY VERSUS RECURRENT TUMORS

Abstract. Head and Neck Squamous Cell Carcinoma, also known as HNSCC, is the sixth most common cancer worldwide. Between now and 2030, new cases are anticipated to increase by 30%, totalling approximately 1.08 million new cases annually. Generally, all tumors that originate in the mucosal epithelium lining of the oral cavity, pharynx, larynx, and sinonasal tract are considered part of HNSCC. Most HNSCCs in the oral cavity and larynx develop due to abusive alcohol and tobacco consumption, whereas development of HNSCCs in the pharynx seems connected to human papillomavirus (HPV). Due to the consumption of carcinogen products, such as the areca nut, which local people tend to chew, HNSCC is most prevalent in South Asia and Australia. It is also very prevalent in the United States and Europe due to higher infection rates of HPV. Additionally, HNSCC is known for its genetic instability, needing multiple genetic transformations to occur, which is what this study will focus on.

Keywords: HNSCC, HPV.

1. Background

HNSCC is a typically localized cancer. Compared to other cancers, it spreads to distant parts of the body more slowly. Development of HNSCC is connected to alcohol abuse, tobacco consumption, and prior positive HPV infections. Originating in the oral cavity (which includes lips, buccal mucosa, hard palate, anterior tongue, floor of mouth and retromolar trigone), the nasopharynx, the oropharynx (which includes palatine tonsils, lingual tonsils, base of tongue, soft palate, uvula and posterior pharyngeal wall), the hypopharynx (which includes the bottom part of the throat, extending from the hyoid bone to the cricoid cartilage), and the larynx, HNSCC usually metastasizes to the lungs or nearby lymph nodes.

Progression of invasive HNSCC usually follows a certain pattern: "epithelial cell hyperplasia, followed by dysplasia (mild, moderate and severe), carcinoma in situ and, ultimately, invasive carcinoma." Since HNSCC is very heterogeneous, cell of origin usually depends on "anatomical location and aetiological agent (carcinogen versus virus)"; however, the most common origin is adult stem cells or progenitor cells, which, after oncogenic transformation, turn into cancer stem cells (CSCs) that have self-renewal and pluripotency properties (1). Although HNSCC CSCs constitute only 1–3% of

cells in primary tumors, there have been a number of molecular biomarkers with prognostic significance. Of these CD44, CD133, and ALDH1 are the most validated. CD44 is a "cell surface receptor for hyaluronic acid and matrix metalloproteinases (MMPs) and is involved in intercellular interactions and cell migration. HNSCC cells with high levels of CD44 are capable of self-renewal, and CD44 levels in HNSCC tumours are associated with metastasis and a poor prognosis. Similarly, increased levels of the membrane-spanning protein CD133 are associated with HNSCC invasiveness and metastasis. ALDH1 is an intracellular enzyme that converts retinol into retinoic acid, plays a part in cellular detoxification and is a marker for both normal stem cells and CSCs. High levels of ALDH1 expression or activity are associated with self-renewal, invasion and metastasis and may have prognostic significance in HNSCC" [2].

To pinpoint the specific cell of origin, it is necessary to look at the development of second primary tumors (SPTs). In HNSCC, SPTs appear at an extremely high rate after the diagnosis of the primary tumor, and they are frequently lethal. The development of SPTs reflects CSCs arising from independent oncogenic transformations by looking at the field cancerization, which "involves the formation of multiple patches of premalignant disease with a higher-than-expected rate of multiple local second primary tumors." This suggests that carcinogens damage large anatomical fields.

Symptoms of HNSCC include persistent sore throat, pain, weakness, or numbness near the head and neck, enlarged lymph nodes, and odd patches or openings in the throat and mouth [3].

Survival rates for HNSCC have improved over the past three decades; the 5-year survival increased from 55% from 1992–1996 to 66% during 2002–2006. If treated with Surgery, Chemotherapy, Radiation Therapy, Immunotherapy, or Targeted Therapy, survival rates are now 56–62% across all five stages. However, after treatment, 15–50% patients develop recurrent HNSCC, which is both difficult to treat and a major cause of morbidity. Recurrent HNSCC is difficult to treat because of the loss of effectiveness due to prior treatments and the infiltrative nature of recurrent diseases in the head and neck area. A study done on the most effective treatment for recurrent HNSCC suggests that aggressive treatments, such as surgery and CCRT, reduces deaths of recurrent HNSCC patients most efficiently; however, there is yet a study to be done on the mutated genes responsible for recurrence.

2. Methods

The tools and databases used in this study are publicly available. NCBI's GEO database was used to search for datasets relevant to the objective studied; those with samples separated through arrays were analyzed with GEO2R, while those with samples separated through high throughput sequencing were analyzed with DESeq2. Source Batch Search was used to annotate gene symbols. Results were copied onto Google Sheets, where samples were filtered into upregulated and downregulated according to p-value and fold change. Then, a venn diagram was drawn to find common genes in multiple studies for more accurate results. These genes were then compiled, searched on GenCard Bank, and separated according to their functions.

2.1 Sample Download and Extraction

NCBI's GEO database, a public international archive storing genomics data submitted by the research community (https://www.ncbi.nlm.nih.gov/geo/), was used. Studies with results relevant to this paper's objective were extracted, and samples were downloaded and separated into two types: array and high throughput sequencing. There were two tumor versus normal samples selected for more accurate results, and one primary versus recurrent sample.

2.2 Array Analysis

GEO2R (https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html) was used for the array datasets. GEO2R is an interactive web tool available to GEO

datasets with array samples, which allows users to compare two or more of these samples to identify genes that are differentially expressed across set experimental conditions. The results are processed and presented by significance as a table of ordered genes, and graphic plots are available to help visualize differentially expressed genes and assess data set quality.

2.3 High Throughput Sequencing Analysis

DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) was used for high throughput sequencing analysis. The R program was installed, and the DESeq2 package was downloaded. Then, using code chunks, results were organized into a table.

2.4 Filtering

The resulting tables from both DESeq2 and GEO2R were put separately into google sheets and filtered according to p-value<0.01 and FC<=0.5 (downregulated), FC>=2 (upregulated).

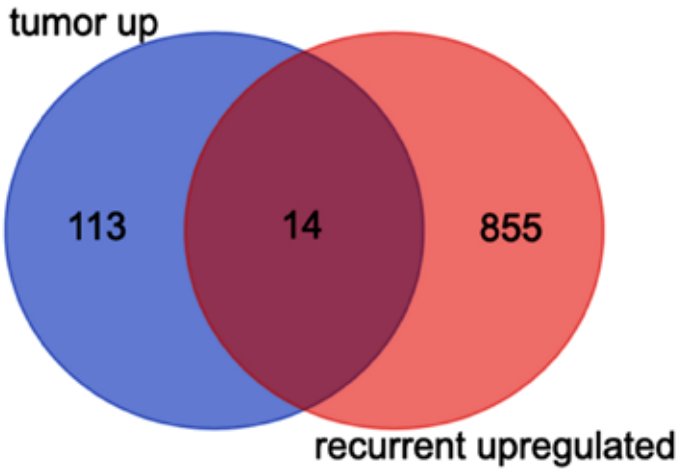
2.5 Commonality Grouping

Resulting gene names were compiled into an online venn diagram tool (https://bioinformatics.psb.ugent.be/webtools/Venn/). The two tumor versus normal upregulated results were inserted, and the common genes were located; this process was repeated for the downregulated genes and the primary versus recurrent genes. The common genes found were separated into upregulated in recurrent tumors and downregulated in recurrent tumors.

Input files:

List names	number of elements	number of unique elements
recurrent upregulated	869	869
tumor up	127	127
Overall number of unique elements		**982**

Image result:



Save Image As PNG Save Image As SVG

Text results:

Save text result

Names	total	elements
recurrent upregulated tumor up	14	MET GJA1 ITGA3 HOXD10 TTYH3 CA2 TENM2 CXCL1 CDSN LAMB3 ITGA6 FMNL2 MMP10 SLC2A1

2.6 Gene Function Research

GeneCards: The Human Gene Database (https://www.genecards.org), an online knowledgebase that automatically integrates gene-centric data from ~150 web sources, was used for the research of functions and locations of genes.

3. Results

The basic information of the genes were compiled into the table below. In total, there are 9 genes that regulate physiological processes, 5 genes that regulate tumor-related functions, 5 genes that regulate inflammation, 4 genes that regulate ion-related functions, 3 genes that regulate cell surface adhesion, 2 genes that regulate immune cells, 2 genes that regulate signaling, 2 that regulate antigens, and 5 whose functions are unrelated to any of the others.

Table 1.

	Full Name	**Main Function**	**Protein/gene family**	**Detailed description**	**Up or Down regulated**
1	***2***	***3***	***4***	***5***	***6***
MET	Mesenchymal Epithelial Transition	Physiological Processes	receptor tyrosine kinase protein family	Regulates proliferation, scattering, morphogenesis; reduces lung fibrosis	Up
GJA1	Gap Junction Protein Alpha 1	Physiological Processes	connexin gene family, encodes protein that's component of gap junctions in the heart	Involved in synchronized heart contraction, embryonic development, bladder capacity, and hearing	Up
ITGA3	Integrin Subunit Alpha 3	Cell surface adhesion	integrin alpha chain protein family	n/a	Up
HOXD10	Homeobox D10	Physiological Processes	Abd-B homeobox protein family	Involved in cell differentiation and limb development; part of developmental regulatory system: provides cells with specific positional identities on anterior-posterior axis	Up
ITYH3	Tweety Family Member 3	Ion Channels	tweety family of proteins	Encoded protein is calcium (2+)- activated large conductance chloride (–) channel; responsible for ion channel transport and transport of inorganic cations/anions and amino acids/oligopeptides	Up
CA2	Carbonic anhydrase 2	Bone reabsorption	isozymes of carbonic anhydrase	Essential for bone resorption and osteoclasts differentiation; regulates fluid secretion into anterior chamber of eye; contributes to intracellular pH regulation in duodenal upper villous epithelium during proton-coupled peptide absorption	Up

1	*2*	*3*	*4*	*5*	*6*
TENM2	Teneurin Transmembrane Protein 2	Physiological Processes, cell surface adhesion, ion channels	tenascin	Enables cell adhesion molecule and signaling receptor binding activity; involved in calcium-mediated signaling using intracellular calcium source; heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules; retrograde trans-synaptic signaling by trans-synaptic protein complex; involved in neural development by regulating proper connectivity within nervous system	Up
CXCL1	C–X-C Motif Chemokine Ligand 1	Inflammation	CXC subfamily of chemokines	Encoded protein is a secreted growth factor that signals through G-protein coupled receptor and CXC receptor 2; plays role in inflammation and as chemoattractant for neutrophils	Up
CDSN	Corneodesmosin	Epidermal	protein found in corneodesmosomes	Epidermal barrier integrity	Up
LAMB3	Laminin Subunit Beta 3	Physiological processes	basement membrane proteins	Mediates attachment, migration, organization of cells into tissues during embryonic development by interacting e other extracellular matrix components	Up
ITGA6	Integrin Subunit Alpha 6)	Cell surface adhesion	integrin alpha chain protein family	Present in oocytes, involved in sperm-egg fusion; plays structural role in hemidesmosome	Up
FMNL2	Formin Like 2	Physiological processes	formin-related protein	Regulates cell morphology and cytoskeleton organization; required in cortical actin filament dynamics	Up
MMP10	Matrix Metallopeptidase 10	Physiological processes	peptidase M10 family of matrix metalloproteinases (MMPs)	Breaks down extracellular matrix in normal physiological processes (embryonic development, reproduction, tissue remodeling, and disease processes like arthritis and metastasis)	Up
SLC2A1	Solute Carrier Family 2 Member 1	Glucose transport	Solute carrier family	Encodes major glucose transporter in mammalian blood-brain barrier; protein mainly found in cell membrane and cell surface, also functions as receptor for HTLV virus I and II	Up
B3GALT5	Beta-1,3-Galactosyltransferase 5	Antigens	membrane-bound glycoproteins	Encoded protein may synthesize type1 Lewis antigens, which are elevated in gastrointestinal and pancreatic cancers	down

1	*2*	*3*	*4*	*5*	*6*
ADH7	Alcohol Dehydrogenase 7	Metabolize substrates	class IV alcohol dehydrogenase 7 mu or sigma subunit	Most active as retinol dehydrogenase, thus may participate in synthesis of retinoic acid (hormone used for cellular differentiation); catalyzes NAD-dependent oxidation of all-trans-retinol, alcohol, and omega-hydroxy fatty acids	down
HPGD	15-Hydroxyprostaglandin Dehydrogenase	Metabolism of prostaglandins, inflammation	short-chain non-metalloenzyme alcohol dehydrogenase protein family	Catalyzes NAD-dependent oxidation of hydroxylated polyunsaturated fatty acids; decreases levels of pro-proliferative prostaglandins such as prostaglandin E2 (whose activity increased in cancer because increase in expression of cyclooxygenase 2); inactivates resolvins E1, D1, D2, which play roles in inflammation	down
SCGB1A1	secretoglobin family 1A member 1	Physiological processes, inflammation	secretoglobin family of small secreted proteins	Anti-inflammation, inhibition of phospholipase A2, sequestering of hydrophobic ligands	down
NUCB2	Nucleobindin-2	Ions, tumor related	calcium binding protein	Calcium level homeostasis, eating regulation in hypothalamus, release of tumor necrosis factor from vascular endothelial cells; non receptor guanine nucleotide exchange factor, binds to and activates guanine nucleotide binding protein (G-protein) alpha subunit GNAI3	down
KRT4	Keratin, type I cytoskeletal 4	Epithelial	keratin gene family	Specifically expressed in differentiated layers of mucosal and esophageal epithelia	down
CXCL12	C–X-C Motif Chemokine Ligand 12	Physiological processes, tumor related, immune cells, ion channels, inflammation	stromal cell-derived alpha chemokine member of intercrone family	Protein functions as ligand for G-protein coupled receptor, chemokine (C–X-C motif) receptor 4; CXCR4 activated to induce rapid and transient rise in level of intracellular calcium ions and chemotaxis. Plays roles in embryogenesis, immune surveillance, inflammation response, tissue homeostasis, tumor growth/metastasis; chemoattractant active on T-lymphocytes and monocytes but not neutrophils, stimulates migration; several critical functions in embryonic development, bone marrow and heart ventricular septum formation, B-cells	down

1	*2*	*3*	*4*	*5*	*6*
DIO2	Iodothyronine Deiodinase 2	Tumor related	iodothyronine deiodinase family	Protein is selenoprotein w non-standard amino acid Sec, which encoded by the UGA codon that signals translation termination	down
CCL21	C–C Motif Chemokine Ligand 21	Immune cells, ions, inflammation	CC cytokines genes	Immunoregulatory and inflammatory processes; encoded protein inhibits hemopoiesis and stimulates chemotaxis; chemotactic in vitro for thymocytes and activated T-cells, not for B cells macrophages or neutrophils; cytokine also plays role in mediating homing of lymphocytes to secondary lymphoid organs	down
GNA14	G Protein Subunit Alpha 14	Signaling	guanine nucleotide binding/ G protein family	Modulators/transducers in various transmembrane signaling systems	down
GCNT3	Glucosaminyl (N-Acetyl) Transferase 3, Mucin Type	Antigens	N-acetylglucosaminyltransferase family	Introduce the blood group I antigen during embryonic development	down
BOC	BOC Cell Adhesion Associated, Oncogene Regulated	Signaling	immunoglobulin/ fibronectin type III repeat family	Cell-surface receptor com-led that mediates cell-cell interactions between muscle precursor cells, promotes myogenic differentiation	down
PLAC8	Placenta Associated 8	Tumor related	cornifelin family	Might enable chromatin binding activity, positive regulation of cold-induced thermogenesis, positive regulation of transcription by RNA polymerase II, acts upstream/within several processes (brown fat cell differentiation, defense response to bacterium, response to cold)	down
GULP1	GULP PTB Domain Containing Engulfment Adaptor 1	Tumor related	nucleocytoplasmic shuttling protein	Protein encoded is adapter protein necessary for engulfment of apoptotic cells by phagocytes; modulates cellular glycosphingolipid and cholesterol transport; may play role in internalization and endosomal trafficking of various LRP1 ligands such as PSAP	down

4. Discussion

According to the resulting table, the most prevalent function of the differentially expressed genes is the regulation of physiological processes. However, as physiological processes is a generally broad topic, it is necessary to take the more detailed description into account. Most notably, there are 5 genes that are used in embryonic development: GJA1, HOXD10, LAMB3, MMP10, CXCL12. The mutation of genes that are responsible for embryonic development causes the increased potential for developing cancer, as the inability of embryonic cells to develop proper structures that regulate important functions may encourage cancer development. It is connected to hereditary cancer and the tendency for certain groups of demographics to develop cancer.

Secondarily, the tumor-related genes. Again, tumor-related is a broad topic, and detailed descriptions are needed; however, the functions of these genes can be easily connected to the reasons behind recurrence and tumor development. As expected, all genes in this section are downregulated – the disappearance of these genes will increase the likelihood of cancer recurring. Thus, this paper will not be discussing their functions in detail.

Third, inflammation. In the inflammation section, there is only one gene that is upregulated: CXCL1. The rest, HPGD, SCGB1A1, CXCL12, and CCL21, are all downregulated. Inflammation is the body's response to tissue damage, which can be caused by physical injury, infection, exposure to toxins, or other types of trauma. Inflammation causes the repairing of damaged tissue and cellular proliferation. If the cause persists or certain control mechanisms fail, inflammation can become chronic. Once this occurs, tissue repair and cell proliferation will often create an environment in which cancers have a tendency to develop [4]. This information supports the notion that the differentially expressed genes in this table are connected to recurrence. If those that usually regulate the inflammatory response shut-down, mutate, and do not occur frequently in the tumor sites, the inflammatory responses will become chronic and cancer will develop again, even if it is removed.

Fourth, ion regulation. Studies show that increases in intracellular calcium may inhibit apoptosis, depending on concentration level, location, and timing [5]. In the results above, all 4 genes, TTYH3, TENM2, NUCB2, and CXCL12, are connected to the calcium ion. Two are upregulated, two are downregulated, respectively. TTYH3 encodes for a calcium channel; if upregulated, it therefore increases the calcium ion channels in the cell membrane, which may cause an increase in the intracellular calcium. Using an intracellular calcium source, TENM2 is involved in calcium-mediated signaling. The upregulation of this gene will therefore cause an influx of calcium into the cell. NUCB2 is responsible for calcium level homeostasis. Downregulation of it will cause a destruction of balance; an increase in intracellular calcium will not be returned to normal. CXCL12 activates CXCR4, which induces a rapid increase of calcium levels inside the cell. All these causes added together increase the possibility of the intracellular level of calcium reaching the point of inhibition of apoptosis. Due to this inhibition, the likelihood of tumor development increases; recurrency can occur.

Fifth, cell surface adhesion. All cell surface adhesion genes are upregulated in recurrent tumors. It is said that cell surface proteins are capable of restricting cell growth through contact inhibition; alterations of these molecules are common in cancer [6]. CAM-DR (Cell adhesion mediated drug resistance) is a significant limitation to the success of cancer therapies, most notably chemotherapy. The explanation to why this occurs can be explained by the FN model, which shows the cellular arrest in the G1 phase of cellular division, which significantly reduces the efficacy of drugs [7]. Thus, the upregulation of cellular adhesion molecules may inhibit the initial development of cancer to a certain extent; however, once cancer develops, it negatively impacts the efficiency of treatment, which explains why it is prevalent in recurrent tumor cells.

5. Conclusion

To summarize, most genetic mutations cause differential gene expressions in genes regulating embryonic development, tumor-regulation, inflammation, intracellular calcium level regulation, and cell surface adhesion molecules. Mutation in these functions are proved to be responsible for cancer development or secondary cancer development. However, the upregulation and downregulation of these genes appears unconnected to HNSCC specifically; instead, they seem more connected to cancers in general. These results are still relevant to the objective of the paper– the understanding of the mechanisms in cancer recurrence is helpful in HNSCC recurrence identification, and potential future treatments. Additionally, since recurrence is connected to survival rates, by looking at the genes in the tumor samples doctors will be able to predict the patients' survival rates after treatment.

Acknowledgments

I would like to thank Dr. Pingzhang Wang for introducing to me the tools used in this study, answering my questions when I was confused, and guiding me to my understanding of this topic.

References:

1. Chang Wu, Yuan T. H. Wu and Wu. Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, and treatment outcomes, NIH 2017.
2. Johnson E., Burtness, C. Leemans, Lui, E. Bauman, and R. Grandis, Head and neck squamous cell carcinoma, NIH 2020.
3. Head and neck squamous cell carcinoma, Medicine Plus 2015.
4. Singh, Baby, Rajguru, Patil, Thakkannavar, Pujari, Inflammation and Cancer, NIH 2019.
5. Fnu, Weber, Alterations of Ion Homeostasis in Cancer Metastasis: Implications for Treatment, NIH 2021.
6. Moh, Shen. The roles of cell adhesion molecules in tumor suppression and cell migration, NIH 2009.
7. Huang, Wang, Tang, Qin, Shen, He, Ju. CAM-DR: Mechanisms, Roles and Clinical Application in Tumors, Frontiers 2021.

Section 3. Preventive Medicine

https://doi.org/10.29013/ELBLS-23-1-19-24

Jiayi Wu,

UNCOVERING THE GENETIC BASIS OF THYROID CANCER: A STUDY OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)

Abstract. Thyroid cancer remains a significant public health concern, with an estimated 43.720 new cases predicted in 2023, affecting both men and women. Our study aimed to explore the genetic underpinnings of thyroid cancer predisposition by analyzing Single Nucleotide Polymorphisms (SNPs). To do so, we leveraged Genome-Wide Sequencing data from the Sequence Read Archive and developed tailored analysis pipelines using Bowtie 2, a popular alignment tool, to map the sequences onto chromosome 7 and perform variant calling. Our analysis revealed 9 SNPs that were present in over 90% of thyroid cancer patients but not in the normal population. These findings hold promise for the development of new strategies for the early detection and prevention of thyroid cancer.

Keywords: thyroid, cancer, SNPs.

Introduction

Cancer

In 2023, thyroid cancer continued to be a significant health challenge in the United States, with The American Cancer Society reporting 43.720 new cases (12.540 in men and 31.180 in women) and 2,120 deaths (970 men and 1,150 women) (American Cancer Society, 2021). This cancer results from the uncontrolled growth of cells in the thyroid gland, and it can be classified into four types: papillary, follicular, medullary, and anaplastic. Among them, papillary carcinoma is the most common, accounting for about 80% of cases, and has a slow-growing nature with the best prognosis among all thyroid cancer types (*Thyroid Cancer* – Patient Version – *NCI*, n.d.). Even though it can spread to nearby lymph nodes, papillary carcinoma rarely causes life-threatening complications. Being aware of the type of thyroid cancer is essential for choosing the most effective treatment and improving the chances of recovery.

Following papillary carcinoma, follicular thyroid cancer is the second most common type, accounting for about 15% of cases. Although it has a good prognosis, follicular cancer is more aggressive than papillary cancer and is more likely to spread to other organs, such as the lungs and bones, even if it does not spread to the nearby lymph nodes. Medullary thyroid cancer, the third most common type, represents 4% of cases and is less differentiated than papillary and follicular cancers. This type of cancer may also spread to the lymph nodes and other organs, and high levels of calcitonin and carcinoembryonic antigen can be indicative of its presence (*Thyroid Cancer* – Patient Version – *NCI*, n.d.). Anaplastic thyroid cancer is the rarest form, making up only 2% of all cases, but it is also the most aggressive type. This cancer grows rapidly, is highly undifferentiated, and can quickly spread to other parts of the body. Understanding the characteristics of each type of thyroid cancer is crucial to determine the appropriate diagnosis, treatment, and

management strategies for patients (*Thyroid Cancer* – Patient Version – *NCI*, n.d.).

Stages

Thyroid cancer is staged based on the extent of metastasis. Stage I is when the cancer is limited to the thyroid gland and has not spread beyond it (Thyroid Cancer – Patient Version – NCI, n.d.). Stage II is characterized by tumor growth and appearance in surrounding tissues (Thyroid Cancer – Patient Version – NCI, n.d.). In Stage III, cancer has spread to nearby lymph nodes within level VI, as shown in Figure 3. Stage IV is the most advanced stage, in which cancer has spread to distant sites outside of the level VI lymph node and may involve gross soft tissue extension (Thyroid Cancer – Patient Version – NCI, n.d.). Please refer to Figures 1–3 for visual representations of the different stages of thyroid cancer.

Table 1: Pathological TNM criteria

	Thyroid carcinoma
STAGE 1	Less than 2 cm in diameter without evidence of disease outside of the thyroid gland.
STAGE 2	Between 2 and 4 cm without evidence of extra thyroidal disease.
STAGE 3	Greater than 4 cm, or level VI nodal metastases or microscopic extra thyroidal invasion regardless of tumor size.
STAGE 4	Any distant metastases, or lymph node involvement outside of level VI, or gross soft tissue extension.

Figure 1. The stages of thyroid cancer, based on the degree of metastasis. Stage 1 represents cancer limited to the thyroid gland, while stage 4 indicates cancer has spread to distant areas. As cancer progresses through each stage, the likelihood of metastasis and the severity of cancer increases (Saeed et al. [15])

Treatment

There are several treatment options available for thyroid cancer. The six standard treatments are thyroid surgery, radiation therapy, chemotherapy, thyroid hormone therapy, targeted drug therapy, and watchful waiting (National Cancer Institute [7]). The type of treatment recommended will depend on various factors, such as the stage and type of thyroid cancer, as well as the patient's age and overall health.

Surgical options for thyroid cancer include thyroidectomy, thyroid lobectomy, and lymph node dissection. Thyroidectomy involves the removal of all or most of the thyroid gland, while thyroid lobectomy involves the removal of only a portion of the gland. In some cases, individuals who have inherited a gene that is likely to cause thyroid cancer may choose to have a thyroidectomy to decrease their chance of developing medullary thyroid cancer (National Cancer Institute [7]). Lymph node dissection involves the removal of lymph nodes in the neck area. However, as with any surgery, there are potential side effects such as temporary or permanent hoarseness or loss of voice, damage to the parathyroid glands, excessive bleeding, formation of a major blood clot in the neck (called a hematoma), and infection (National Cancer Institute [7]).

Radiation therapy is another treatment option and includes the use of radioactive iodine treatment, which is used to kill any remaining thyroid cancer cells after surgery (Mayo Clinic [14]). Chemotherapy may be used in some cases, but it is not as effective for thyroid cancer as it is for other types of cancer. Thyroid hormone therapy is also commonly used to treat thyroid cancer, as it can help to suppress the production of

thyroid-stimulating hormone, which can cause cancer to grow. Targeted drug therapy is a newer treatment option that uses drugs to target specific molecules that play a role in the growth and spread of cancer cells. Finally, watchful waiting may be recommended for patients with small or slow-growing tumors, particularly if the patient is elderly or has other health issues that may make surgery or radiation therapy risky.

Gene

Thyroid cancer commonly involves mutations in the BRAF proto-oncogene and serine/threonine kinase genes, as identified by Xing in 2005. Figure 2 illustrates that these genes encode the protein B-Raf proto-oncogene serine/threonine kinase, which plays a crucial role in the MAP kinase/ERK signaling cascade that regulates cell division, differentiation, and secretion. The gene is located on Chromosome 7,7q34. Mutations in this gene occur in about 44% of papillary thyroid cancer cases, leading to the protein's activation despite signals from other proteins.

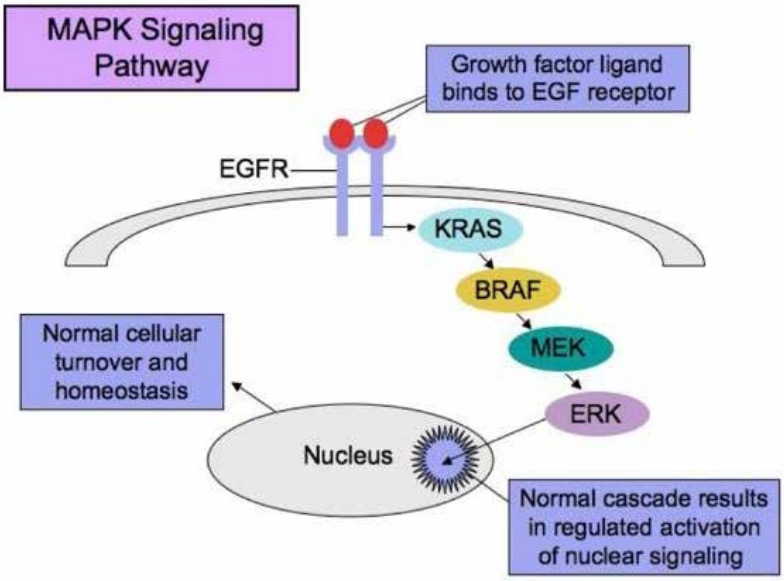


Figure 2. The MAP kinase/ERK signaling pathway. Growth factors bind to their receptors, leading to the activation of the small GTPases H/K and NRAS, which in turn activate the protein kinase B-Raf proto-oncogene serine/threonine kinase (BRAF). B-Raf activates a series of proteins that ultimately activate the extracellular signal-regulated kinase (ERK), leading to the regulation of various cellular processes such as cell division, differentiation, and secretion. Mutations in the BRAF gene, located in chromosome 7q34, are commonly found in papillary thyroid cancer and can lead to constitutive activation of the MAP kinase/ERK pathway (Affiliated Pathologists Medical Group, [20])

SNPs

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in humans, marked by differences in a single nucleotide found in more than 1% of the population (Medline Plus, 2020). They occur once every 300 base pairs of sequence on average, with a minor allele frequency (MAF) greater than 1% (Kruglyak and Nickerson, 2001; Stephens *et al.*). As a biological marker, SNPs can be useful in cancer diagnosis, as they can predict an individual's response to specific drugs, susceptibility to environmental factors, and risk of developing cancer.

Methods

To analyze sequence reads, we first downloaded the human reference genome for chromosome 7 from Ensembl. We then selected SRA sequences (accession number PRJNA887246) based on a library strategy and holistic study design of thyroid cancer

and normal patients. The fastq-dump tool was used to download the sequences, which were then checked for quality using the FastQC tool and trimmed as necessary. Poor quality sequences (with a Phred quality score < 33) were identified through the HTML file visualization and trimmed using Trimmomatic. The sequences were then indexed and locally aligned using Bowtie2, and the output SAM file was converted to a BAM file using SAM tools. The BAM file was sorted by coordinates, and read coverage was calculated for each position. To identify single nucleotide polymorphism (SNP) variants, we used the Binary Variant Call Format (BCF) tools.

Results

Our analysis identified various types of genetic variants, including non-coding variants, non-coding transcript exon variants, intergenic SNPs, synonymous variants, missense variants, and regulatory region variants. Non-coding variants are mainly caused by skipping the first or last exon, which could lead to the loss of start or stop codons, respectively (Dhamija & Menon [7]). Non-coding transcript exon variants can involve changes in non-coding exon sequences within non-coding transcripts. Intergenic SNPs can potentially disrupt regulatory elements (Macintyre et al. [11]). Synonymous variants involve codon substitutions that do not alter the encoded amino acid, while missense variants can result in changes to a single base pair, producing a different amino acid than the one normally produced (Edwards et al. [8]). Some missense mutations may impact the function of the encoded protein. Regulatory region variants located in non-coding genomic regions were also detected, which could have a significant impact on the development of diseases (Rojano et al. [14]). These findings suggest a complex and diverse range of genetic variants in our sample, which may contribute to disease susceptibility and pathogenesis (Adeyemo & Rotimi [1]).

Table 1.– Classification of SNPs found. In the 120 individual sequences analyzed, a total of 9 common SNPs were found. This table displays the chromosome position, and SNP accession number, followed by the percentage of individuals containing the particular variation, and then the genetic consequence: intergenic variants (blue), exon variants (turquoise), synonymous variant (green), and not known variants (pink)

Position	Accession Number	Percentage	Consequence
132251222	rs1799238406	98	n/a
149334844	rs928483266	93	exon variant
89638503	rs553702084	88	intergenic variant
109452135	rs1793466465	88	n/a
109452140	rs1254795666	88	intergenic variant
109452143	rs1019099610	88	intergenic variant
109452176	rs1032989423	88	intergenic variant
109452178	rs953608898	88	intergenic variant
55181370	rs1050171	78	synonymous variant

Discussion

Our study identified 9 different single nucleotide polymorphisms (SNPs) and their respective positions, prevalence, and consequences in a population of 102 individuals. We found that at least 78 percent of each SNP was present in the analyzed population, and we listed all the variants for each SNP. Our findings could have implications for detecting thyroid cancer through specific mutations associated with the disease. As DNA sequencing becomes less expensive, individuals may be more likely to undergo early screening for thyroid cancer. To obtain samples for DNA sequencing, thyroid fine needle aspiration biopsy is a minimally invasive procedure that can be used

to remove a small tissue sample from the thyroid gland (Cha & Koo [6]; Johns Hopkins Medicine [19]).

Moving forward, we recommend conducting further research by obtaining thyroid biopsies from a larger and more diverse population, both with and without thyroid cancer, to determine if these same SNPs are present. This research could lead to improved accuracy in detecting thyroid cancer and ultimately better outcomes for patients.

Conclusion

In summary, our analysis revealed significant differences between the thyroid cancer cohort and the normal group, providing evidence that specific single nucleotide polymorphisms (SNPs) are linked to genetic susceptibility to cancer. Our findings support previous research on this topic and contribute to a better understanding of the molecular mechanisms underlying thyroid cancer. These results may have important implications for the development of new screening and treatment strategies for individuals at high risk of developing thyroid cancer. Further research is needed to validate our findings and expand our knowledge of the complex genetic factors involved in cancer susceptibility.

References:

1. Abdullah M. I., Junit S. M., Ng K. L., Jayapalan J. J., Karikalan B. & Hashim O. H. Papillary Thyroid Cancer: Genetic Alterations and Molecular Biomarker Investigations. International journal of medical sciences,– 16(3). 2019.– P. 450–460. URL: URL: https://doi.org/10.7150/ijms.29935
2. Adeyemo A., & Rotimi C. Genetic variants associated with complex human diseases show wide variation across multiple populations. Public health genomics,– 13(2). 2010.– P. 72–79. URL: https://doi.org/10.1159/000218711
3. APMG melanoma molecular pathways. (2022). Affiliated Pathologists Medical Group. URL: https://www.apmggroup.net/innovation/molecular_testing/melanoma_pathways/melanoma.html
4. BRAF B-RAF Proto-oncogene, serine/threonine kinase [Homo sapiens (human)] – Gene – NCBI. (2023, February 28). National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/gene/673
5. Cabanillas M. E., McFadden D.G., & Durante C. Thyroid cancer. Lancet (London, England),– 388(10061). 2016.– P. 2783–2795.URL: https://doi.org/10.1016/S0140-6736(16)30172-6
6. Cha Y. J., & Koo J. S. Next-generation sequencing in thyroid cancer. Journal of translational medicine,– 14(1). 2016.– 322 p. URL: https://doi.org/10.1186/s12967-016-1074-7
7. Dhamija S., & Menon M. B. Non-coding transcript variants of protein-coding genes – what are they good for? RNA biology,– 15(8). 2018.– P. 1025–1031. URL: https://doi.org/10.1080/15476286.2018.1511675
8. Edwards N. C., Hing Z. A., Perry A., Blaisdell A., Kopelman D. B., Fathke R., Plum W., Newell J., Allen C. E., Shapiro A., Okunji C., Kosti I., Shomron N., Grigoryan V., Przytycka T. M., Sauna Z. E., Salari R., Mandel-Gutfreund Y., Komar A. A., … Kimchi-Sarfaty C. Characterization of coding synonymous and non-synonymous variants in ADAMTS13 using ex vivo and in silico approaches. PloS one,– 7(6). 2012. e38864. URL: https://doi.org/10.1371/journal.pone.0038864
9. Homo sapiens (ID887246) – BioProject – NCBI. (n.d.). National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA887246
10. Key statistics for thyroid cancer. (2023, January 18). American Cancer Society | Information and Resources about Cancer: Breast, Colon, Lung, Prostate, Skin. URL: https://www.cancer.org/cancer/thyroid-cancer/about/key-statistics.html

11. Macintyre G., Jimeno Yepes A., Ong C. S., & Verspoor K. Associating disease-related genetic variants in intergenic regions to the genes they impact. PeerJ, – 2, 2014. e639. URL: https://doi.org/10.7717/peerj.639
12. NCI Dictionary of genetics terms. (n.d.). National Cancer Institute. URL: https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/missense-variant
13. RASD2 RASD family member 2 [Homo sapiens (human)] – Gene – NCBI. (2022, October 26). National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/gene/23551
14. Rojano E., Seoane, P., Ranea J. A. G. & Perkins J. R. Regulatory variants: from detection to predicting impact. Briefings in bioinformatics,– 20(5). 2019.– P. 1639–1654. URL: https://doi.org/10.1093/bib/bby039
15. Saeed M. F., Sakrani N. F., Juma I. M. & Ali A. Medullary thyroid carcinoma: Management and complexities of postoperative follow-up. International Journal of Case Reports and Images,– 8(3). 2017.– 171 p. URL: https://doi.org/10.5348/ijcri-201728-cr-10767
16. Single Nucleotide Polymorphism – an overview | ScienceDirect Topics. (2013). Sciencedirect.com. URL: https://www.sciencedirect.com/topics/medicine-and-dentistry/single-nucleotide-polymorphism
17. Thyroid cancer – Diagnosis and treatment – Mayo Clinic. (2022, May 13). Mayo Clinic – Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/thyroid-cancer/diagnosis-treatment/drc-20354167
18. Thyroid cancer–Patient version. (n.d.). National Cancer Institute. URL: https://www.cancer.gov/types/thyroid#:~: text=There%20are%20four%20main%20types, in%20how%20aggressive%20they%20are
19. Thyroid fine needle aspiration biopsy. (2019, November 19). Johns Hopkins Medicine, based in Baltimore, Maryland. URL: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/thyroid-fine-needle-aspiration-biopsy#:~: text=A%20thyroid%20fine%20needle%20aspiration%20biopsy%20is%20a%20procedure%20that, the%20front%20of%20your%20neck
20. Types of thyroid cancer: Papillary, follicular & other carcinomas. (2022, June 7). Cancer Treatment Centers of America. URL: https://www.cancercenter.com/cancer-types/thyroid-cancer/types
21. What are single nucleotide polymorphisms (SNPs)?: MedlinePlus genetics. (2022, March 12). MedlinePlus – Health Information from the National Library of Medicine. URL: https://medlineplus.gov/genetics/understanding/genomicresearch/snp/#:~: text=Single%20nucleotide%20polymorphisms%2C%20frequently%20called, building%20block%2C%20called%20a%20nucleotide
22. Xing M. BRAF mutation in thyroid cancer. Endocrine-related cancer,– 12(2). 2005.– P. 245–262. URL: https://doi.org/10.1677/erc.1.0978

https://doi.org/10.29013/ELBLS-23-1-25-31

Xiangbo Guo,

BUILDING A PREDICTIVE MODEL OF ADHD AMONG CHILDREN

Abstract. Attention Deficit/Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental disorders of childhood. According to the Centers of Disease Control and Prevention (CDC), the estimated number of children ever diagnosed with ADHD nationalwide is 6.1 million (9.4%). Among those children, 6 in 10 with ADHD had at least one other mental, emotional, or behavioral disorder that may have long-lasting impacts on their development.

In this research, we investigated possible risk factors related to development of ADHD among children and identified the most significant positive and negative factors through logistic regression. We used the 2020 National Survey of Children's Health survey data containing 42.777 complete data samples with features ranging from demographic information to the child's family condition. The response variable is whether a child has ever been diagnosed with ADHD.

After processing the dataset, we built a logistic regression model to predict whether a child will develop ADHD. By investigating the logistic regression coefficients, we found that parents' physical and mental health, the family's financial ability to cover basic living expenses, and whether the parents are divorced are all risk factors. The logistic regression model has achieved an AUROC score of 0.73, with 0.67 true positive rate (TPR) and 0.324 false positive rate (FPR). This predictive model is helpful for healthcare professionals to identify and reduce the risk for the children that are prone to the development of ADHD.

Keywords: ADHD, logistic regression, model, ROC, children, risk.

1. Introduction

Attention Deficit/Hyperactivity Disorder, or ADHD, is usually first diagnosed in childhood and often lasts into adulthood. Children with ADHD may have trouble paying attention, controlling behaviors, and sometimes appear to be reckless. It is normal for children to have trouble in focusing and behaving at one time or another, including failing to give close attention to details, making careless mistakes in schoolwork, at work, or with other activities, and having trouble organizing tasks and activities. In addition, children diagnosed with ADHD also appear to be hyperactive, with typical sympotoms like being overly talkive and easy to be annoyed.

The estimated number of children ever diagnosed with ADHD, according to a national 2016 parent survey, is 6.1 million (9.4%). This number includes: 388,000 children aged 2–5 years; 4 million children aged 6–11 years; 3 million children aged 12––17 years. Boys are more likely to be diagnosed with ADHD than girls (12.9% compared to 5.6%) [1]. According to a national 2016 parental survey, 6 in 10 children with ADHD had at least one other mental, emotional, or behavioral disorder [2].

ADHD is believed to be cuased collectively by multiple factors, including genetic, such as familial inheritance, food additives/diet, lead contamination, cigarette and alcohol exposure, maternal smoking during pregnancy, and low birth weight [3]. The sympotoms of ADHD can be alleviated through ways like mental counseling together with stimulant or nonstimulant medications.

Given the major impact of ADHD on patients' daily life, it is of great importance for healthcare professional to identify children that are at high risk for developing ADHD and help address problems at an

early stage. To fulfill this task, this report discussed the machine learning techniques that can be applied to build predictive models on whether a child will develop ADHD. Specifically, we pre-processed the dataset, built a logistic regression model, and investigated factors most related to the development of ADHD. We also measured the model performance using various validation techniques and analyzed the model coefficients to find the variables that contribute most to our predicted results.

2. Method

2.1 Data

We used 2020 National Survey of Children's Health survey data for this study. The National Survey of Children's Health (NSCH) is conducted by the U. S. Census Bureau for the U. S. Department of Health and Human Services' (HHS) Health Resources and Services Administration's (HRSA) Maternal and Child Health Bureau (MCHB). It is designed to provide national and state-level information about the physical and emotional health and wellbeing of children under the age of 18 in the United States, their families and their communities, as well as information about the prevalence and impact of children with special health care needs. The 2020 NSCH data contains 42.777 complete data samples. We used the following variables as independent variables.

Table 1.– Features used for analysis

Variable	Description	Comments
HHCOUNT	How many people are living or staying at this address?	
A1_BORN	Where were you born?	1: In the U.S., 2: Outside the U.S.
A1_GRADE	What is the highest grade or level of school you have completed?	Higher value indicates higher education
A1_MARITAL	What is your marital status?	
A1_AGE	What is your age?	
A1_PHYSHEALTH	In general, how is your physical health?	Higher value indicates worse physical health
A1_MENTHEALTH	In general, how is your mental or emotional health?	Higher value indicates worse mental health
SC_SEX	What is this child's sex?	1: Male, 2: Female
SC_RACE_R	What is this child's race?	
AGEPOS4	Birth order of this child.	
SC_HISPANIC_R	Is this child of Hispanic, Latino, or Spanish origin?	1: Yes, 2: No
BIRTHWT_L	Low birth weight (< 2500g).	1: Yes, 2: No
ACE1	SINCE THIS CHILD WAS BORN, how often has it been very hard to cover the basics, like food or housing, on your family's income?	Higher value indicates higher frequency
ACE3	Has this child EVER experienced any of the following? Parent or guardian divorced or separated	1: Yes, 2: No
ACE4	has this child EVER experienced any of the following? Parent or guardian died	1: Yes, 2: No

The dependent variable is a binary feature coded as "K2Q31A," which indicates whether the child has ever been diagnosed with Attention Deficit Disorder (ADD) or Attention Deficit/Hyperactivity Disorder (ADHD).

2.2 Exploratory Analysis

A correlation graph is a primitive yet straightforward representation of the cells of a matrix of correlations. The idea is to display the pattern of correlations in terms of their signs and magnitudes by using visual thinning and correlation-based variable ordering. Moreover, the matrix cells can be shaded or colored to show the correlation value. The positive correlations are shown in red, while the negative correlations are shown in blue; the darker the hue, the greater the magnitude of the correlation.

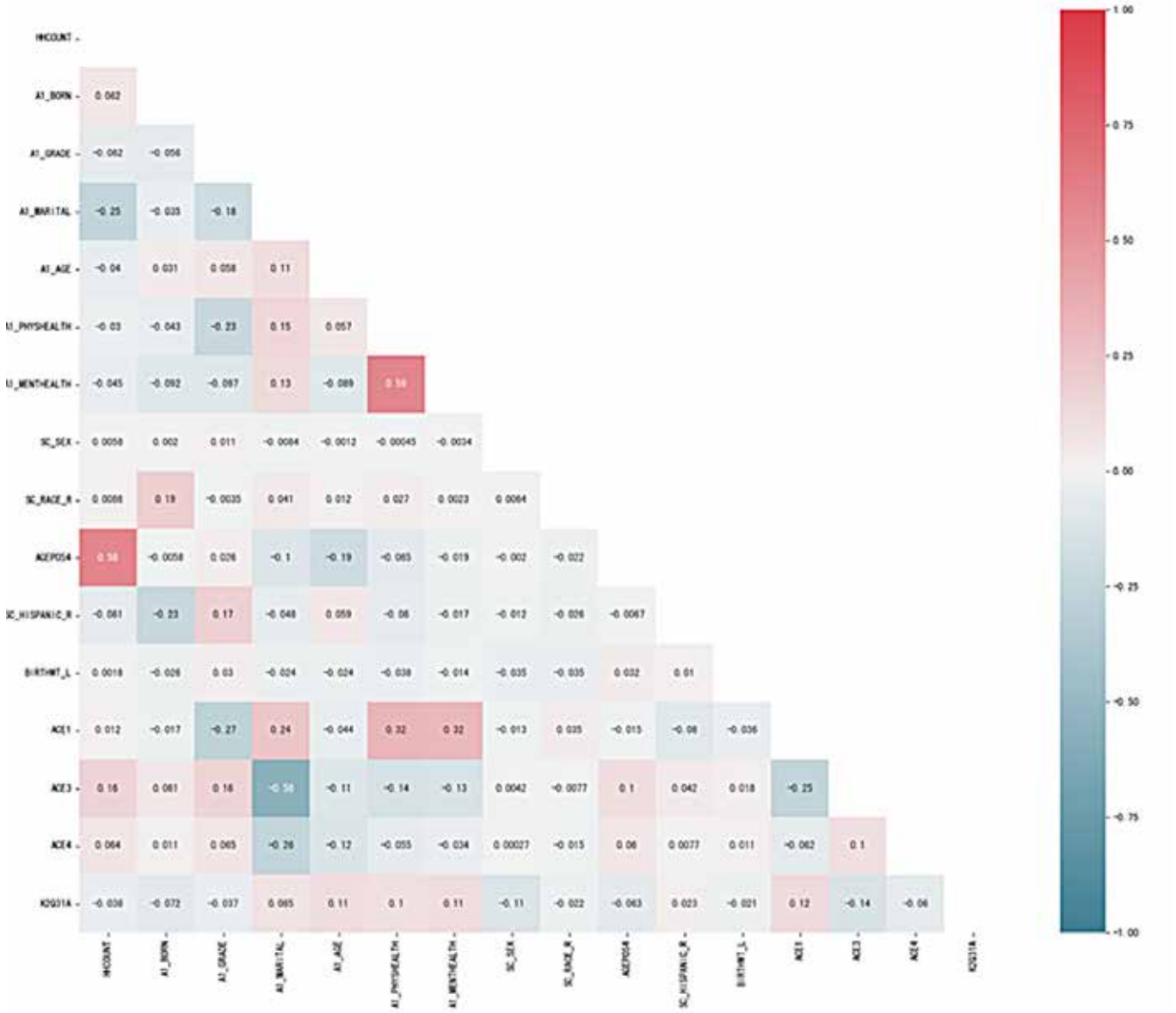


Figure 1. Correlation among variables

The graph above shows that the dependent variable (has ADHD) has the highest positive correlation with ACE1 (hard to cover basic living expenses), while having the highest negative correlation with ACE3 (parents or guardians not divorced).

In addition, the correlation graph also provides valuable information regarding the relationship among features. For example, the correlation between A1_PHYSHEALTH (parents' physical health condition) and A1_MENTHEALTH (parents' mental health condition) is 0.59, indicating that the two variables are significantly positively correlated and generally parents with worse physical health condition may also have worse mental health condition.

2.3 Statistical Method

2.3.1 Pre-processing

The data set is pre-processed in this step to improve both the training speed and accuracy. As the dataset is complete and does not contain any missing values, we did not employ any imputation technique here. In addition, as different features usually have remarkedly different value ranges, we applied the feature standardization technique to transform different features into comparable scales. This measure ensures that different features weigh equally in the training process. For each feature, its mean value and standard deviation are first computed as *avg(x)* and *std(x)*. Then each data point x with respect to that feature is replaced by y_i calculated as:

$$y_i = \frac{x - avg(x)}{std(x)}.$$

Finally, the dataset is partitioned into two datasets for training and test purposes: the training dataset (70%) for model development and the test dataset (30%) for model test and validation.

2.3.2 Logistic Regression

Logistic regression models were used to calculate the predicted risk. Logistic regression is a part of a category of statistical models called generalized linear models, and it allows one to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a combination of these. Typically, the dependent variable is dichotomous, and the independent variables are either categorical or continuous.

The logistic regression model can be expressed with the formula:

$$\ln\left(\frac{y}{1-y}\right) = w_0 + w_1 x_1 + \ldots + w_m x_m$$

In the logistic regression, y is the probability of the sample classified as the positive class, and each feature x_i has its specific weight w_i, where w_0 is the intercept while w_1 through w_m are the coefficients of the independent variables.

Our task is to find a set of parameters $w_0,\ldots,w_m$ such that the loss function between the output y and the actual values u

$$l(y,u) = \| y-u \|_2^2$$

is minimized.

In addition, we applied elastic-net regularization to constrain model complexity and prevent model over-fitting problems with L-1 ratio equaling 0.5.

2.3.3 Model Validation

Consider a two-class prediction problem, where the outcomes are labeled either as positive or negative. There are four possible outcomes from a binary classifier. If the outcome from a prediction is positive and the actual value is also positive, then it is called a true positive (TP); however, if the actual value is negative, then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are negative, and false negative (FN) is when the prediction outcome is negative while the actual value is positive. In this way, the true positive rate (TPR) can be calculated as follows:

$$TPR = \frac{TP}{TP + FN}$$

And the false positive rate (FPR) can be calculated as:

$$FPR = \frac{FP}{TN + FP}$$

A confusion matrix is a table that allows visualization of the performance of an algorithm. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class. An example of the confusion matrix and the meaning of each cell within the table can be found in the graph below. Typically, the confusion

matrix of a good predictive model has high true positive and true negative rates.

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Figure 2. Confusion matrix example

A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [4].

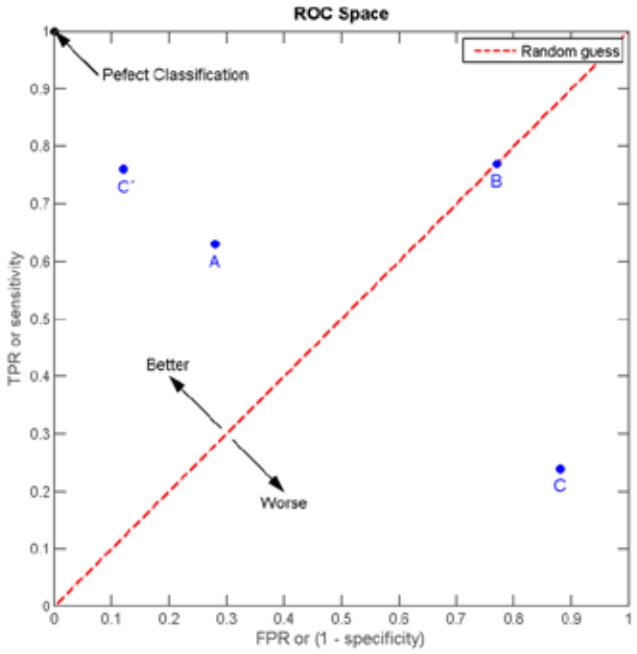


Figure 3. A sample ROC plot

The best possible prediction method would yield a point in the upper left corner of the ROC space. A random guess would give a point along a diagonal line from the left bottom to the top right corners. Points above the diagonal represent better than random classification results, while points below the line represent worse than random results. A sample ROC plot is shown in Figure 2. In general, ROC analysis is one tool to select possibly optimal models and to discard suboptimal ones independently from the class distribution. Sometimes, it might be hard to identify which algorithm performs better by directly looking at ROC curves. Area Under Curve (AUC) overcomes this drawback by finding the area under the ROC curve, making it easier to find the optimal model.

3. Results

3.1 Confusion matrix and ROC curve

Figure 4 shows the confusion matrix of the logistic regression model. The upper left region is true negative, the upper right region is false positive, the lower left region is false negative, and the lower right region is true positive. As shown in Figure 4, the logistic regression model has a relatively high (~67.0%) true positive rate and a relatively low (~32.4%) false positive rate.

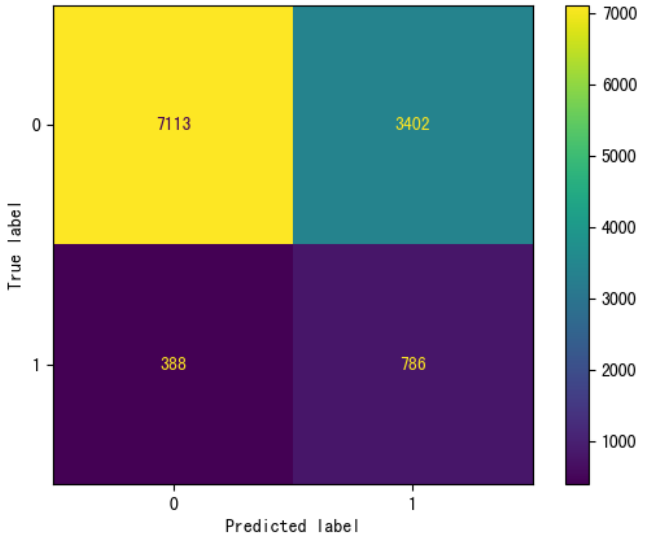


Figure 4. Confusion matrix of the predicted results

Figure 5 displays the ROC curve for the logistic regression model. It can be concluded that the model has results much better than random guessing and the AUROC score is 0.73.

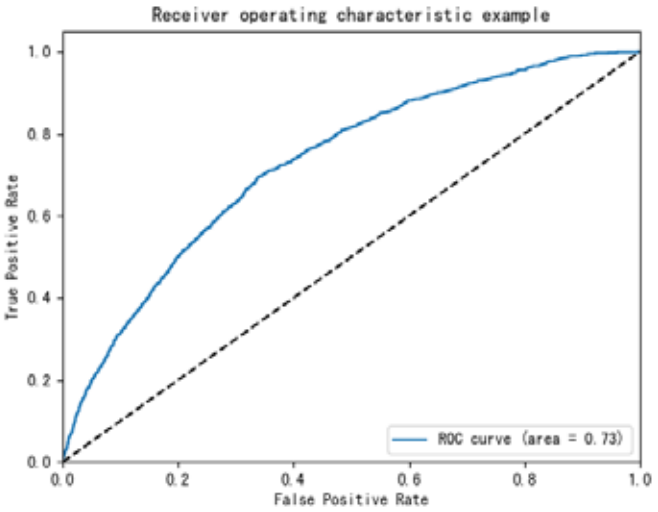


Figure 5. The ROC curve for the logistic regression model

3.2 Feature Importance

Like in linear regression, the coefficients in the logistic regression model also provide valuable information about the direction and magnitude of the impact of each input variable on the dependent variable. In other words, these coefficients can provide the basis for a crude feature importance score. The figure below shows the coefficient of each input variable.

The chart below shows A1_MENTHEALTH (parents' mental health) and ACE1 (hard to cover basic living expenses) are positively related to the development of ADHD, and ACE3 (parents not divorced) are negatively related to the development of ADHD. These results align with our findings from the correlation analysis. In addition, we also found that male children are more likely to develop ADHD and female children are less likely to develop ADHD (SC_SEX). This finding is corresonds with exisitng evidence suggesting that the prevalence of ADHD is greater in males than females [5] and ADHD is more commonly diagnosed in adult males compared with adult females at a ratio of 1.6:1 [6].

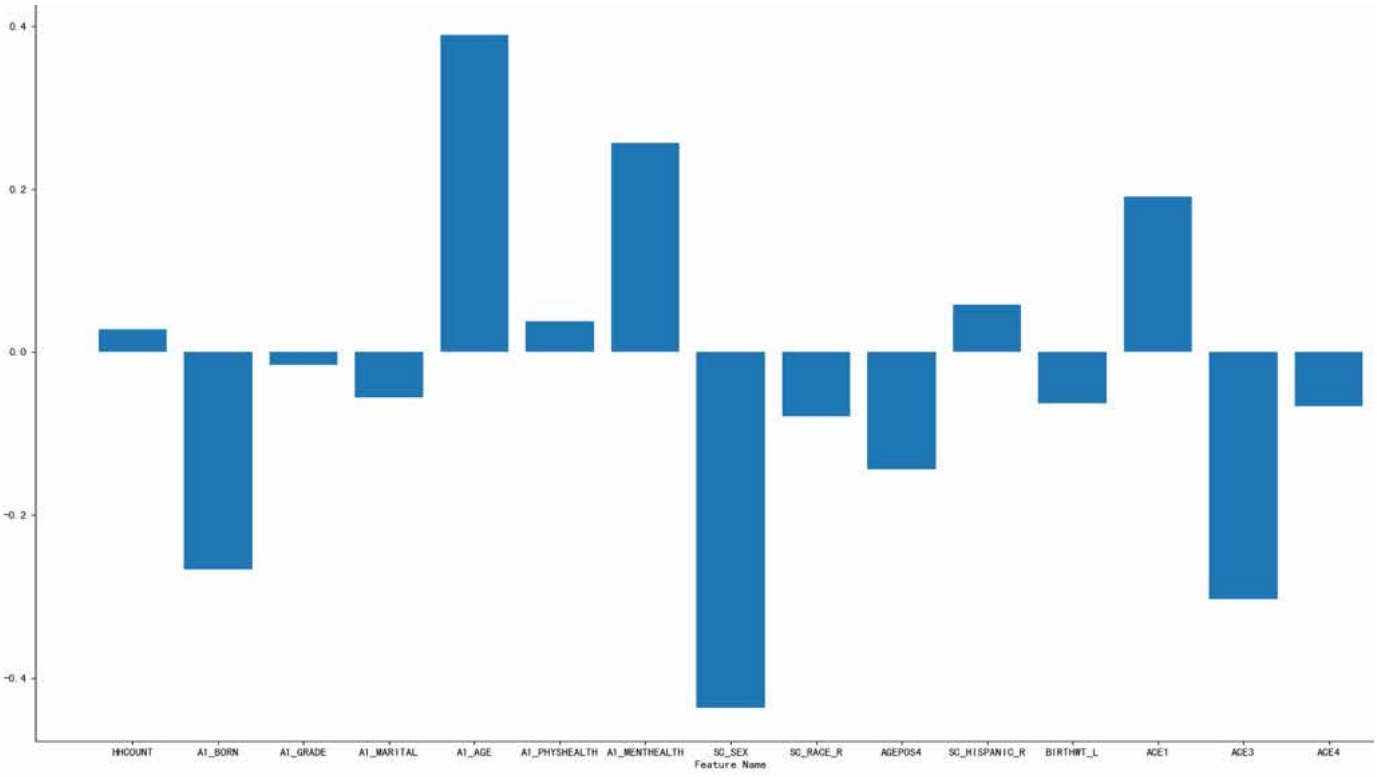


Figure 6. The importance score for each feature

4. Discussion

This study intends to build a predictive model to investigate the factors most related to the development of Attention Deficit/Hyperactivity Disorder (ADHD) among children. Through preliminary analysis, we discovered that parents' physical and mental health, family's financial ability to cover basic living expenses, as well as whether the parents are divorced are all risk factors. A logistic regression model was built, and the AUROC score is 0.73, indicating that the model has achieved relatively good performance in making accurate predictions on whether a child will develop ADHD. The predictive model suggests that A1_MENTHEALTH (parents' mental health) and ACE1 (hard to cover basic living expenses) are top risk factors for ADHD. A possible explanation of the results might be that a parents with worse mental health may have higher violence intention and may even abuse the child. In addition, children grown up in a family that is hard to pay for basic living expenses may be mentally insecure and thus are more prone to mental illnesses such as ADHD. This predictive model is helpful for healthcare professionals to identify children that are at higher risk for ADHD and come up with specific plans to reduce their risk for long-term impacts.

One limitation of this study is that we did not explore how individual independent variables have contributed to the overall predictive performance. Even though we can ascertain how each variable is correlated to our dependent variable through the model, we still have no idea how it influences the final model outcome. Therefore, this direction can be investigated in future studies. In addition, we only employed the vanilla logistic regression classification model in this study. Future studies can apply more complicated machine learning models, such as the artificial neural network, and compare its performance with logistic regression. With a highly accurate classification model, healthcare professionals might provide more customized and better service to children who are likely to have ADHD and find measures to minimize the long-term impacts on their development.

References:

1. NSCH 2003–2011: National Survey of Children's Health, telephone survey data; estimate includes children 4–17 years of age.
2. Danielson M. L., Bitsko R. H., Ghandour R. M., Holbrook J. R., Kogan M. D., Blumberg S. J. Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. Journal of Clinical Child and Adolescent Psychology.– 47:2. 2018– P. 199–212.
3. Banerjee T. D., Middleton F., Faraone S. V. Environmental risk factors for attention-deficit hyperactivity disorder. Acta Paediatr.– Sep; 96(9). 2007.– P. 1269–74. Doi: 10.1111/j.1651-2227.2007.00430.x. PMID: 17718779.
4. Google. Classification: ROC Curve and AUC | Machine Learning Crash Course. Accessed November 25, 2021. URL: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc
5. Nøvik T. S., Hervas A., Ralston S. J., et al. Influence of gender on attention-deficit/hyperactivity disorder in Europe–ADORE. Eur Child Adolesc Psychiatry – 15(Suppl 1). 2006.– I15-I24.
6. Willcutt E. G. The prevalence of DSM–IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics – 9. 2012.– P. 490–499.

Section 4. Physico-Chemical Biology

https://doi.org/10.29013/ELBLS-23-1-32-41

Son Nguyen Ngoc,
Huong Nguyen Thi,
Institute chemistry and materials
17 Hoang Sam, Nghia Do, Cau Giay, Ha Noi, 100000, Viet Nam

CHARACTERIZATION AND ANTIBACTERIAL ACTIVITIES OF ZNO NANOPARTICLES SYNTHESIZED FROM GUAVA LEAF EXTRACTION

Abstract. In recent years, the method of synthesizing materials with limited use of chemicals has attracted the attention of many researchers. Zinc oxide particles, an object with potential in biomedical applications, has also been synthesized by this method. In this study, ZnO NPs were prepared by a very simple green synthesis method using zinc acetate hexahydrate and using guava leaf extract which acted as capping agents without using any additional chemicals. any. Zinc oxide nanoparticles synthesized by this pathway (GL-ZnO) have been described by X-ray, EDX, FTIR, and SEM techniques. The results show that GL-ZnO has high purity, uniform size, and fineness at the nanoscale. The influence of synthesis conditions on the characteristics of ZnO NPs was evaluated. The IR, and EDX results demonstrate the formation of ZnO NPs, while the SEM results show the size of ZnO at the nanoscale. ZnO NPs showed good antibacterial activity against several gram-negative and gram-positive strains such as E. coli, B. subtilis, and S. cerevisiae.

Keywords: ZnO nanoparticles, Guava leaves, antibacterial.

1. Introduction

Zinc oxide nanoparticles (ZnO NPs) are one of the metal oxides attractive to many researchers for their unique properties and applications. It is cheap, abundant, safe, and easy to prepare. ZnO has a broad energy band (3.37 eV) and a high binding energy of 60 meV, whereby they are thermally, electrically, and chemically stable. ZnO NPs also have low toxicity, increased UV radiation absorption, and strong antibacterial properties. As a result, ZnO NPs are widely studied in applications such as biomarkers, biosensors, drug delivery agents, gene delivery, and nanopharmaceuticals. ZnO is also certified by the US Food and Drug Administration (FDA) as a "GRAS" (generally recognized as safe) substance.

Regarding biology, ZnO NPs have high biocompatibility and antibacterial and antifungal properties [1; 2]. Pham et al [3] synthesized ZnO NPs from orange peel extract and evaluated their antibacterial activity. The reaction conditions including pH, and annealing temperature were investigated to give the optimal synthesis conditions. The synthesized nanoparticle has a size of 30 nm, high antibacteri-

al activity, over 99.9% against E. coli, and 89–98% against S. aureus despite the absence of UV light.

G. Madhumitha et al. [4] used phytochemicals present in Pithecellobium dulce bark extract as stabilizing agents to synthesize ZnO NPs. Nanoparticles have been evaluated for their photo degradability with some water contaminants in the fiber industry such as methylene blue (MB). Antifungal activity against some species such as Aspergillus flavus and Aspergillus niger has been evaluated with relatively high results. At 500 ppm and 1.000 ppm, the antibacterial activity against these fungi was 37.81%, 63.57%, and 40.21%, 43.04%, respectively.

In another remarkable publication by Satarudra P. S. et al [5], quantum dot zinc oxide particles were synthesized using Eclipta alba leaf extract as a reducing agent. The optimal synthesis conditions were confirmed to be at pH 7.5 mL of zinc acetate solution (5 mM) together with 7 mL of the reaction extract for 75 min. TEM results confirmed that there are ZnO particles about 5 nm in size in the post-reaction solution. Selective electron scattering (SAED) analysis recorded the crystalline nature of ZnO having a hexagonal wurtzite phase with lattice constants a = b = 0.32 nm and c = 0.52 nm. The bioactivity evaluation results showed that the antibacterial ability of ZnO was significantly enhanced with the size achieved by the nanoparticles.

Research group of Sanaz A. [6] synthesized and evaluated the photodegradation activity and antibacterial activity of ZnO NPs from the leaf extract of Sambucus ebulus. The crystal phase size of the obtained nanoparticles is about 17 nm. The efficiency of photodegradation with MB pollutants was about 80% after 200 minutes of treatment. The ZnO nanoparticles synthesized by this method showed significant antibacterial efficacy compared to other commercial or crude forms of ZnO and they increased with the processing time.

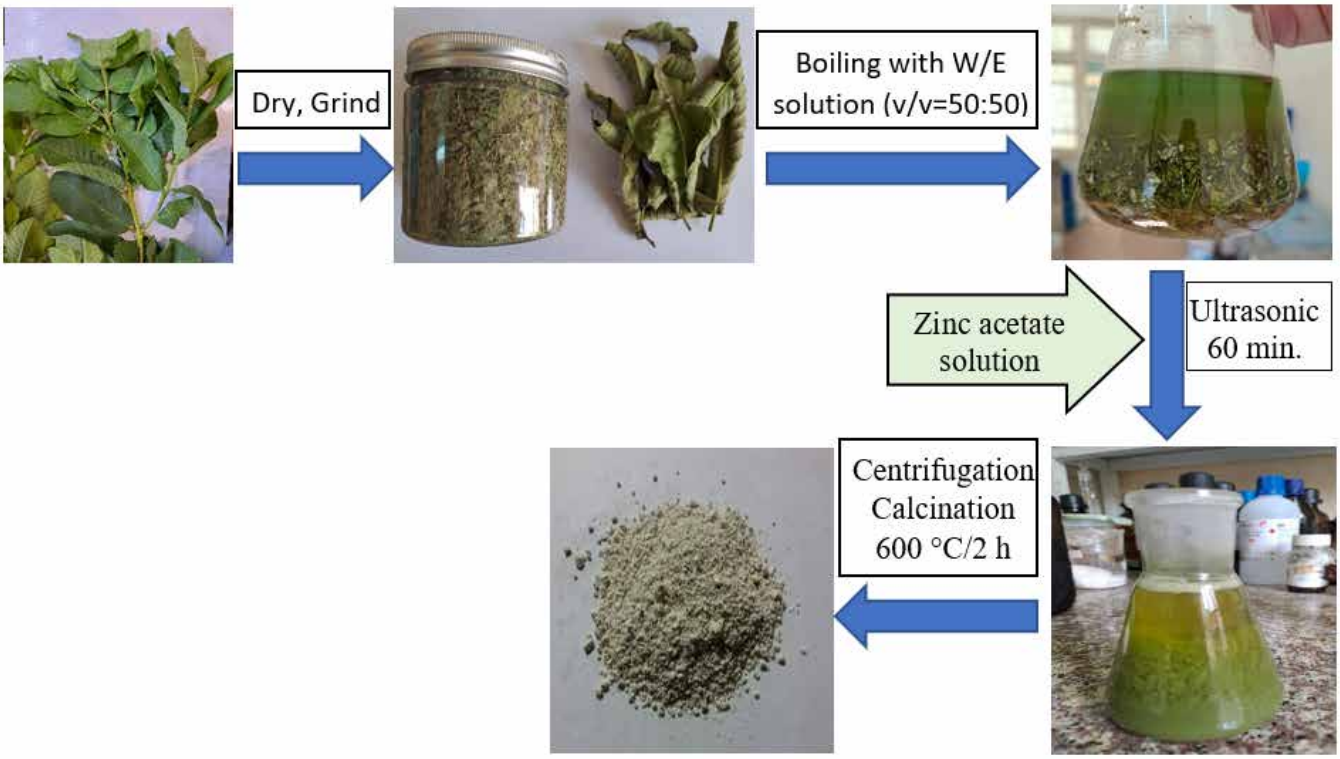


Figure 1. Schematic diagram of ZnO NPs synthesis from guava leaf extract (GL-ZnO)

ZnO NPs are synthesized by chemical, physical, or even biological methods. Precipitation [7; 8], microemulsion [9], sol-gel [10; 11], solvothermal [12; 13], or hydrothermal [14; 15] are examples of chemical processes that consume so much energy or require high temperature and pressure. In the current trend, methods use few or no auxiliary chemicals to synthesize materials to minimize environmental and human impact and reduce costs. Among these, the green synthesis method using plant extracts attracted the most research. ZnO is also the object of many syntheses of this method in orientation for biomedical applications.

Guava is grown in tropical and subtropical countries. In Vietnam, guava leaves are considered a remedy for treating many diseases, such as dysentery, controlling diabetes, losing weight, reducing bad cholesterol, and curing skin allergies. Previous studies have identified guava leaves are rich in phenolic compounds, flavonoids, triterpenoids, tannins, vitamins, essential oils, and sesquiterpene alcohols [16; 17]. As a result, they are the object of attraction for many researchers. Sampath Kumar NS. Et al. [18] took advantage of guava leaf extract's antioxidant and antibacterial activity (GJ) to make jelly without synthetic additives. The results showed that agar from guava leaf extract had high nutritional content with 45.78 g/100 g carbohydrate, 3.0 g/100 g protein, 6.15 mg/100 g vitamin C, 2.90 mg/100 g vitamin B3, and energy 120.6 kcal. The antibacterial activity of guava leaf agar also varied from 11.4 to 13.6 mm against different bacteria. The antioxidant action of GJ towards DPPH free radical was 42.38% and 33.45% for hydroxyl radical. Mass spectrometry confirmed the extract's presence of esculin, quercetin, gallocatechin, 3-sinapoylquinic acid, gallic acid, citric acid, and ellagic acid. These are thought to be the components that contribute to the antibacterial and antioxidant properties of guava leaf extract. In another interesting study, B. Biswas et al. [19] tested the antibacterial activity of guava leaf extract in 4 solvents with increasing polarity, namely hexane, methanol, ethanol, and water, against strains – gram-negative and gram-positive bacteria. The results showed that guava leaf extract using methanol and ethanol solvents had a higher antibacterial effect. Specifically, for the methanol extract, the diameter of the inhibition zone was 8.27 mm and 12.3 mm, and the ethanol extract was 6.11 mm and 11.0 mm, respectively, for B. cereus and S. aureus strains, respectively. These results suggest the potential of guava leaf extract to be used as a natural antibacterial agent.

In this study, we present a green method to synthesize ZnO NPs by using the extraction of guava leaf without adding auxiliary chemicals. The antibacterial activity of the synthesized ZnO NPs was also tested against gram-negative and gram-positive strains including *E. coli*, *B. subtilis,* and *S. cerevisiae*.

2. Materials and methods

2.1. Materials

Guava leaves are harvested at the farm (geographic coordinates: 21.037461, 105.719186) in the summer. Zinc acetate dihydrate ($Zn(CH_3COO)_2 \cdot 2H_2O$) (Merck) as the zinc source; double distilled water and ethanol (Sigma-Aldrich) as the solvents.

2.2. Preparation of Guava leaves extracts

Guava leaves are washed and dried at 32–35°C, 50–57% RH for 2–3 days. Guava leaves were ground and boiled with a mixture of twice distilled water/ethanol (v/v = 50 : 50) at 60 °C for 90 minutes. The guava leaf residue was filtered out; the solution was centrifuged at 5000 rpm for 10 minutes to remove the suspended residue completely. The obtained guava leaf extract was stored at 3–5 degrees Celsius for the following synthesis steps.

2.3 Synthesis of ZnO nanoparticles

50 mL of guava leaf extract was diluted with 50 mL of distilled water. The entire solution was loaded into a single-necked flask in the ultrasonic bath. Dissolve 2 g of zinc acetate in 20 mL of double distilled water. ZnO NPs were synthesized by dripping zinc acetate into guava leaf extract solution,

combined with ultrasound for 60 minutes. The reaction mixture was centrifuged at 10.000 rpm for 5 minutes; the solids were dried at 50 °C for 1 hour before calcining at 600 °C for 2 hours to obtain ZnO NPs. The synthesis process of ZnO NPs is depicted in (Fig. 1).

2.4. Characterization of ZnO NPs

X-ray scattering spectroscopy was used to characterize the crystalline properties of GL-ZnO. The crystal size was calculated using the Scherrer equation (1). The bonds in the synthesized nanoparticle product were studied by FTIR spectroscopy. Their chemical composition was checked by EDX spectroscopy.

$$D = \frac{K \cdot \lambda}{\beta cos\theta} \quad (1)$$

where:

- D is the mean size of the ordered (crystalline) domains, nm;
- K is a dimensionless **shape factor**, with a value close to unity. The shape factor has a typical value of about 0.9;
- λ is the X-ray wavelength, nm;
- β is the line broadening at half the maximum intensity (FWHM), radian;
- θ is the Bragg angle, radian.

2.5. Antibacterial activity test

Antibacterial activity of ZnO was evaluated against gram-positive (*Saccharomyces cerevisiae, Bacillus subtilis* (ATCC9/58) and gram-negative (*E. coli* ATCC 25922) bacteria using the Agar well diffusion method [20]. For microbiological cultivation, the enrichment medium containing meat extract, yeast extract, peptone, glucose, and some mineral salts was utilized. To prepare nutrient agar plates, 37.0 grams of nutrient agar powder was dissolved in 1000 mL of distilled water, and then sterilized in an autoclave at 121 °C /15 lbs pressure for 20 min.

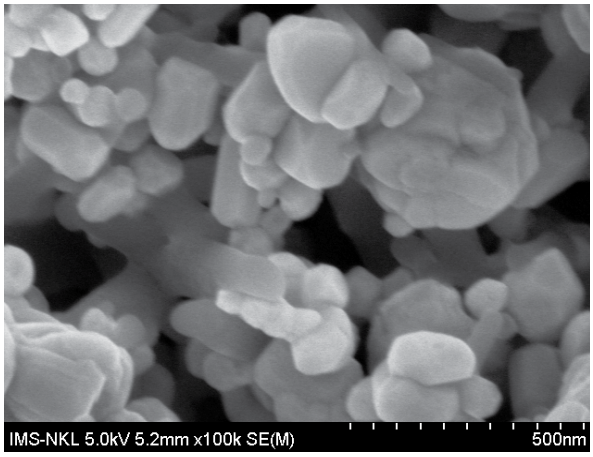
After sterilization treatment, the nutrient agar medium was put into sterile Petri dishes and allowed to solidify. Next, mature broth culture of specific pathogenic bacterial strains in the nutrient broth while distributing all over the surface of agar plates using sterilized L-shaped glass rod. Chitosan and positive-negative control discs were taken into the test for comparison.

The antibacterial activity experiments were conducted with synthesized ZnO NPs and dissolved in 1 mL of DMSO 10% to obtain 200 µg/mL solutions. Under aseptic conditions, 5±1 mm diameter wells were drilled in each Petri dish using the sterile steel cork. Then, 100 µL ZnO NPs were dispersed in 10% DMSO solution and controlled by standard antibiotic Ampicillin (1 mg/ml) as a positive control into the wells. The plates were incubated at 37 °C for 24h before using geometrical Vernier calipers in mm to observe the Zone of inhibition around the wells.

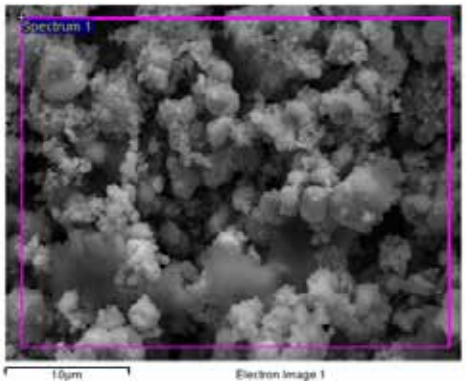
3. Results and Discussion

3.1. Effect of annealing temperature

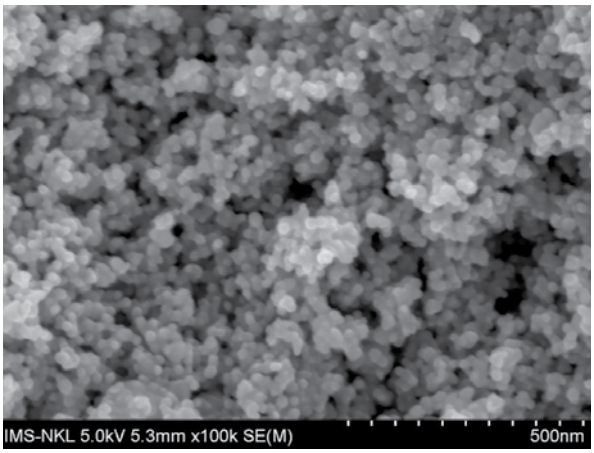
The experimental process shows that the calcination temperature strongly affects the GL-ZnO product. The heating temperatures from 400 °C, 600 °C, and 800 °C were tested, the effects were observed by scanning electron microscope (SEM) images, and the chemical composition was determined by EDX spectroscopy. The results are shown as shown in the figure… shows that with the temperature of 400 °C, the nanoparticle size is quite large, the uniformity is low, and the ratio of O/Zn atoms is approximately 63/37. This shows that the product has not been completely pyrolysis at this temperature, possibly because the organic part has not decomposed completely. With higher temperatures, it is possible to see the nanoparticles with smaller size, fineness, and a higher degree of uniformity. Along with that is the asymptotic O/ZnO ratio of 50/50. At the calcination temperature of 800 °C, the nanoparticle size is 25–30 nm, and the ratio of O/Zn atoms is 51.2/48.8, showing that almost the product after calcination has only ZnO. Thus, this temperature was chosen to synthesize GL-ZnO for further studies.



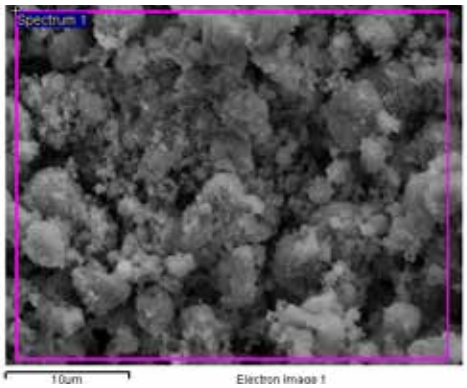
400°C



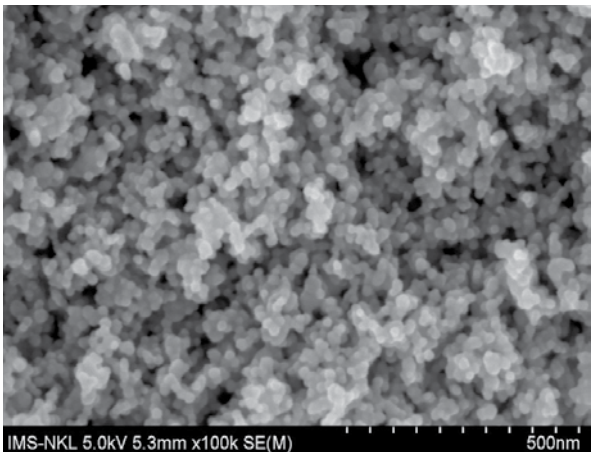
Elements	Weight%	Atomic%
OK	29.01	62.54
ZnK	70.99	37.46
Total	100.00	



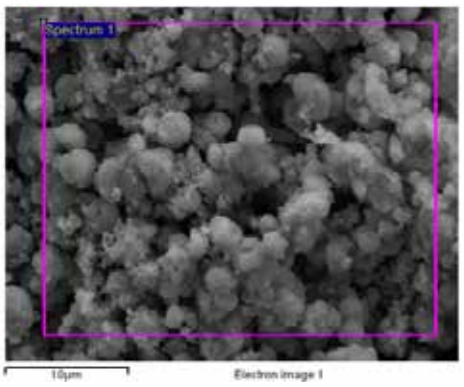
600°C



Elements	Weight%	Atomic%
OK	21.94	53.45
ZnK	78.06	46.55
Total	100.00	



800°C



Elements	Weight%	Atomic%
OK	20.43	51.20
Zn K	79.57	48.80
Total	100.00	

Figure 2. SEM (left) and EDX (right) results of GL-ZnO at different calcining temperatures

3.2 Characterization of- ZnO

The XRD pattern of the as-prepared ZnO NPs (Figure 3) showed peaks scattering angle 2 thetas at 31.79°, 34.44°, 36.28°, 47.56°, 56.59°, 62.88°, 66.39°, 67.96°, 69.15°, corresponding to (100), (002), (101), (102), (110), (103), (200), (112), (201) lattice planes.

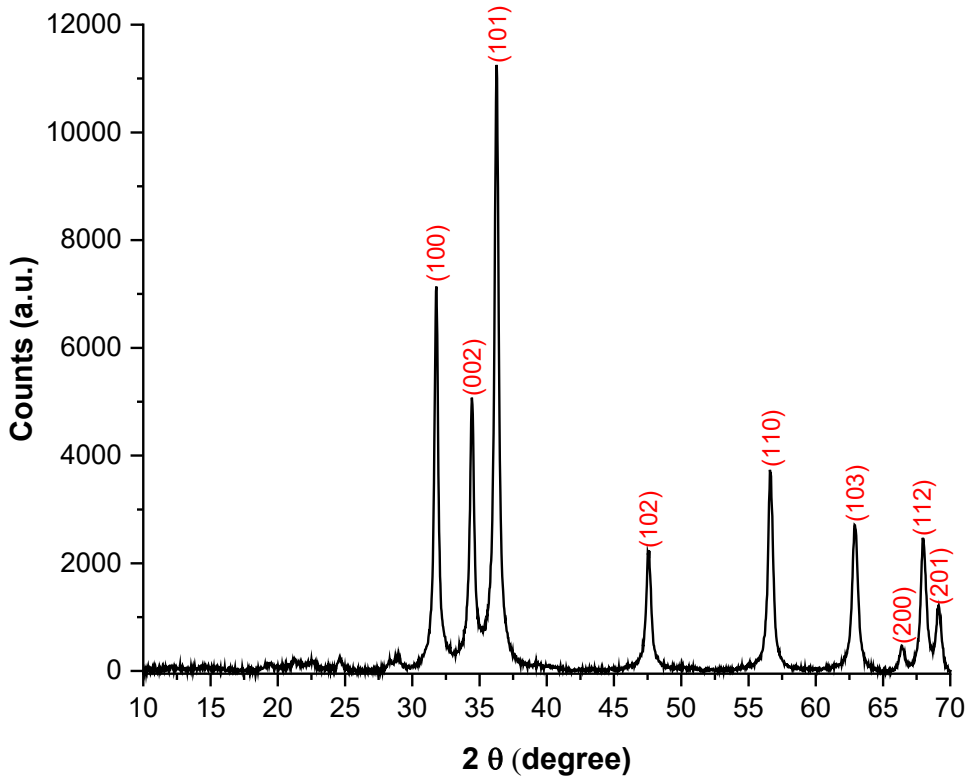


Figure 3. XRD pattern of GL-ZnO

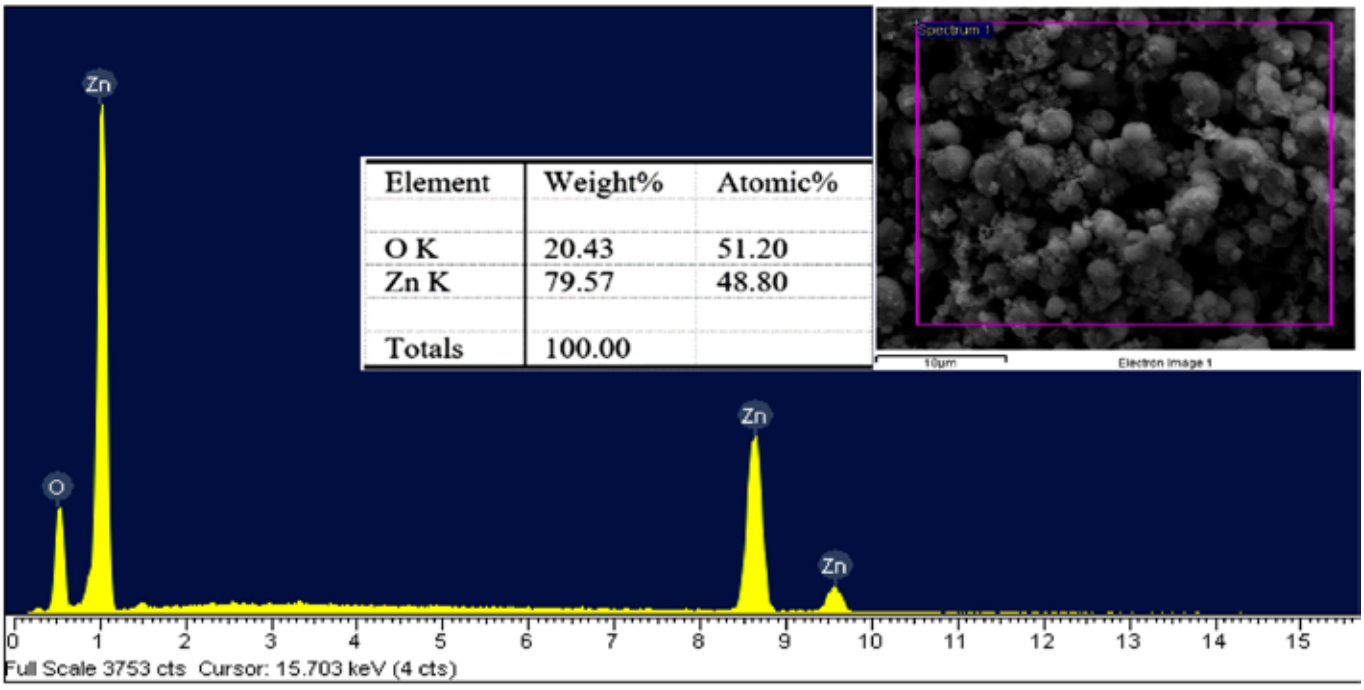


Figure 4. EDX spectrum of GL-ZnO

This result is entirely fitted with the standard XRD pattern of ZnO NPs (JCPDS:36–1451). In addition, the XRD pattern with sharp, intensity peaks, and almost no peaks other than those of ZnO NPs proves that the prepared nanoparticles have a well-crystallized structure and high purity. The EDX results (Figure 4) show only two elements, oxygen, and zinc, with approximately equal atomic ratios. This agrees with the theory and further confirms the purity of the ZnO NPs synthesized by this work. The crystallite size of the nanoparticles calculated by the Scherrer equation is more than 18 nm. As mentioned above, the atomic ratio of these two elements, reaching 51.2/48.8, is approxi-mately the ideal ratio of 1/1. This result shows that the purity level of the GL-ZnO product is very high.

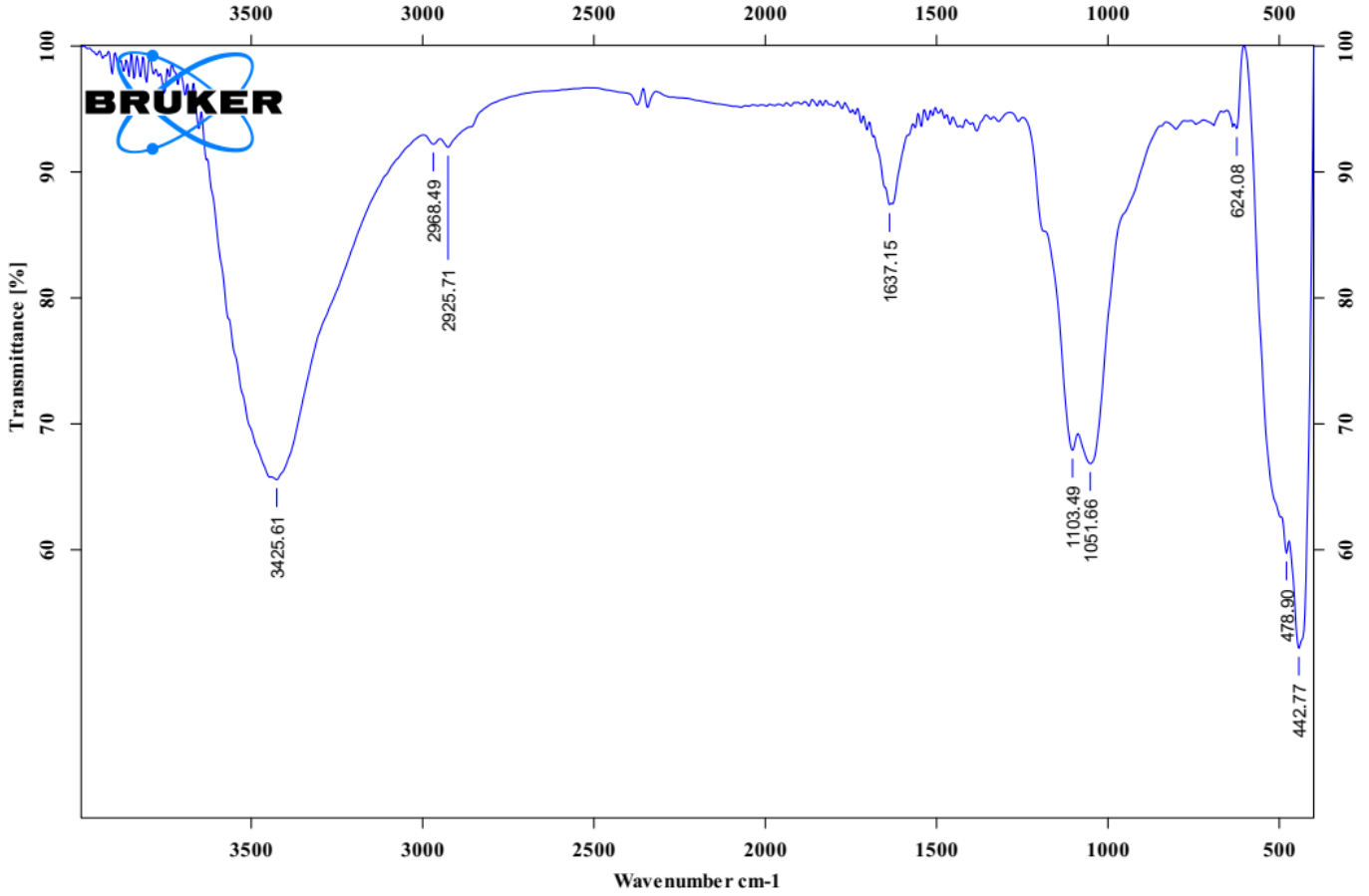


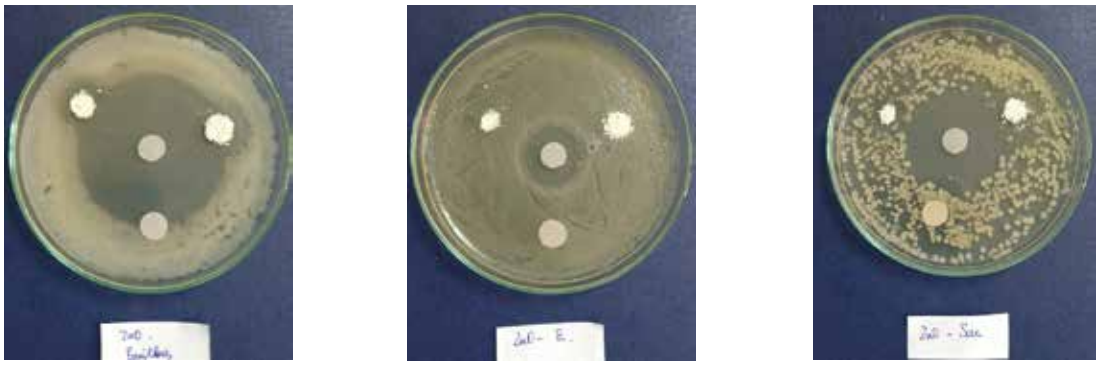
Figure 5. FTIR spectrum of GL-ZnO

The band around 450 cm^{-1} is assigned to the tensile vibration of the Zn-O bond. In addition, in the spectrum, there are signs of the presence of phenolic compounds in the GL extract with strong absorption peaks at wave numbers 3425 cm^{-1} assigned to the valence vibration of the O-H bond, 1051 cm^{-1} and 1103 cm^{-1} represent the valence vibrations of the C-O bond and 1640 cm^{-1} represent the valence band of the C=C and C=O bonds. The weaker absorption band in the region 2925–2968 cm^{-1} is also observed showing the presence of CH_2, and CH_3 groups. In the FTIR spectrum of ZnO, peaks are also observed at this position. It is probably because the calcination time of ZnO is not long enough, and the organic components that act as the capping agent have not been completely pyrolysis.

3.3 Antibacterial activity of GL-ZnO

Figure 6 (a, b, c) depicts testing materials sprinkled on Petri dishes infected with testing microorganisms. The results suggested that ZnO NPs achieved the high-

est capability to inhibit *Bacillus subtilis*. Additionally, clear zones (non-bacterial zones) formed at the sites where ZnO NPs were sprinkled over *Bacillus subtilis*-inoculated Petri dishes.



a b c

Figure 6. Antibacterial activities of ZnO NPs against bacterial Control samples: Bacillus (a), E.coli (b), S.cerevisiae (c)

Table 1.– Antibacterial activity of ZnO NPs on selected bacterial strains

Samples	**Concentration**	**Zone of inhibition in mm**		
		E. coli **(Mean±SE)**	***Bacillus*** **(Mean±SE)**	***Saccharomyces*** **(Mean±SE)**
ZnO NPs	15 mg	–	6 ± 0.14	2 ± 0.15
ZnO NPs	30 mg	–	7 ± 0.17	4 ± 0.16

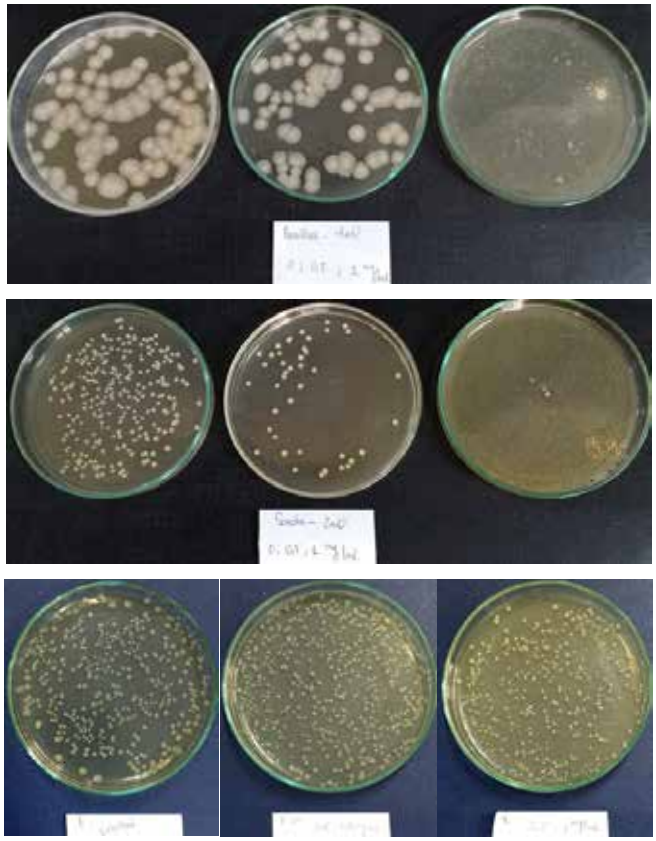


Figure 7. Images showing the antibacterial activities of ZnO NPs with different concentrations

The testing of the capacity of materials to inhibit large microorganisms when exposed to suspensions of the testing organisms as illustrated in Figure 7. Without the requirement for testing materials, the suspensions were evenly dispersed on the controlled plates, and the proliferation of microorganisms on the dish surfaces was also observed. Yet, when *Bacillus subtilis* and *S.cerevisiae* samples solution was exposed to ZnO NPs, the cellular density decreased significantly. Hence, ZnO NPs were able to suppress *Bacillus subtilis* and *S.cerevisiae.*

Figure 7 shows images of the Petri dishes, which is the MIC test results of the ZnONPs. The MIC values against *E. coli, B. subtilis, S. cerevisiae* of ZnONPs were 1.0, 0.5, and 0.5 mg/ml, respectively.

ZnO nanoparticles interact with the membrane of bacterial cells and bind to the mesosome. This mesosome's DNA replication, cell division, and cellular respiration capabilities were diverted, resulting in a larger bacterial cell membrane surface area. On the surface of ZnO NPs, the heavy metal ions Zn^{2+} interact with the microbial cell membranes. The nanoparticles then react quickly with the harmful bacteria to destroy their membrane integrity and cells, resulting in the death of the pathogens [21; 22].

4. Conclusion

This study proposed a straightforward method to synthesize ZnO NPs with almost no use of alkalis as the precipitating agent. The XRD and EDX results have verified that the synthesized ZnO NPs are smaller than 30 nm and high purity. Antibacterial tests with strains of gram-negative bacteria (Bacillus), and fungi (Saccharomyces) show that as-prepared ZnO NPs have a reasonably good killing effect on bacteria with minimum inhibitory concentrations (MIC) are 1.0, 0.5 and 0.5 mg/ml, respectively.

References:

1. Ilves M., et al. Topically applied ZnO nanoparticles suppress allergen induced skin inflammation but induce vigorous IgE production in the atopic dermatitis mouse model. Part Fibre Toxicol,– 11. 2014.– 38 p.
2. Lembo G., et al., Zinc oxide: A new formulation specifically for sensitive skin. Annali Italiani di Dermatologia Allergologica Clinica e Sperimentale,– 60. 2006.– P. 68–72.
3. Thi T. U.D., et al., Green synthesis of ZnO nanoparticles using orange fruit peel extract for antibacterial activities. RSC Advances,– 10(40). 2020.– P. 23899–23907.
4. Madhumitha G., et al., Green synthesis, characterization and antifungal and photocatalytic activity of Pithecellobium dulce peel-mediated ZnO nanoparticles. Journal of Physics and Chemistry of Solids,– 127. 2019.– P. 43–51.
5. Singh A. K., et al., Green synthesis, characterization and antimicrobial activity of zinc oxide quantum dots using Eclipta alba. Materials Chemistry and Physics,– 203. 2018.– P. 40–48.
6. Alamdari S., et al., Preparation and Characterization of Zinc Oxide Nanoparticles Using Leaf Extract of Sambucus ebulus. Applied Science,– 10(10). 2020.– 3620 p.
7. Mohan Kumar K., et al., Synthesis and characterisation of flower shaped zinc oxide nanostructures and its antimicrobial activity. Spectrochim Acta A Mol Biomol Spectrosc,– 104. 2013.– P. 171–4.
8. Raoufi D. Synthesis and microstructural properties of ZnO nanoparticles prepared by precipitation method. Renewable Energy,– 50. 2013.– P. 932–937.
9. Vorobyova S. A., Lesnikovich A. I. and Mushinskii V. V. Interphase synthesis and characterization of zinc oxide. Materials Letters,– 58(6). 2004.– P. 863–866.
10. Shakti N. and Structural G. P.S, and Optical Properties of Sol-gel Prepared ZnO Thin Film. Applied Physics Research, 2010.– 2 p.

11. Wu Q.-H., ZnO nanostructures prepared using a vapour transport method. Journal of Experimental Nanoscience,– 10(3). 2015.– P. 161–166.
12. Segovia M., et al., Zinc Oxide Nanostructures by Solvothermal Synthesis. Molecular Crystals and Liquid Crystals,– 555(1). 2012. P. 40–50.
13. Chen S.-J., et al., Preparation and characterization of nanocrystalline zinc oxide by a novel solvothermal oxidation route. Journal of Crystal Growth,– 252(1). 2003.– P. 184–189.
14. Chen D., Jiao X. and Cheng G. Hydrothermal synthesis of zinc oxide powders with different morphologies. Solid State Communications,– 113(6). 1999.– P. 363–366.
15. Polsongkram D., et al. Effect of synthesis conditions on the growth of ZnO nanorods via hydrothermal method. Physica B: Condensed Matter,– 403(19). 2008.– P. 3713–3717.
16. Altemimi A., et al. Phytochemicals: Extraction, Isolation, and Identification of Bioactive Compounds from Plant Extracts. Plants (Basel), 2017.– 6(4).
17. Shaheena S., et al., Extraction of bioactive compounds from Psidium guajava and their application in dentistry. AMB Express,– 9(1). 2019.– 208 p.
18. Sampath Kumar N. S., et al. Extraction of bioactive compounds from Psidium guajava leaves and its utilization in preparation of jellies. AMB Express,– 11(1). 2021.– 36 p.
19. Biswas B., et al., Antimicrobial Activities of Leaf Extracts of Guava Psidium guajava L. on Two Gram-Negative and Gram-Positive Bacteria. International Journal of Microbiology, 2013.– 746165 p.
20. Magaña S. M., et al. Antibacterial activity of montmorillonites modified with silver. Journal of Molecular Catalysis A: Chemical,– 281(1). 2008.– P. 192–199.
21. Kannan K., et al., Structural studies of bio-mediated NiO nanoparticles for photocatalytic and antibacterial activities. Inorganic Chemistry Communications,– 113. 2020.– 107755 p.
22. Vinotha V., et al. Synthesis of ZnO nanoparticles using insulin-rich leaf extract: Anti-diabetic, antibiofilm and anti-oxidant properties. J Photochem Photobiol B,– 197. 2019.– 111541 p.

Contents