F.No.: COE/ Ph.D./(Notification)/525/2022 **NOTIFICATION DATE: 08-12-2022**

STUDENT NAME: **SAHAR QAZI**
SUPERVISOR NAME: **Dr. KHALID RAZA**
TOPIC: **In Silico Study of POTEE and Analogous Activity of Oncologic Medicaments in Ovarian Cancer**
DEPARTMENT: **COMPUTER SCIENCE**
KEYWORDS: ***Ayurveda, Diagnosis, Epigenetics, High-grade serous ovarian cancer (HGSOC), Person-centric treatment, Prostate, ovary, testis, embryo placenta expressed (POTE)***

Findings

Ovarian cancer is complex –a stealth gynaecological malignancy that develops on the surface of the ovaries in females. Once in its advanced stage (metastatic), it unveils its true signs such as – stomach bloating, menstrual problems, appetite changes, abdominal ache, problems during urination, backache, etc. Scientifically, there are two major factors for development of ovarian cancer – i) Intrinsic factors inclusive of the genetic composition of an individual, age of a person, family history and ii) Extrinsic factors include environmental and social decisions of taking up hormonal replacement therapy (HRT). Traditionally, blood-based biomarkers namely – carbohydrate antigen-125 (CA-125) and human epididymis protein 4 (HE4) are heavily deployed to screen for the disease in females. Ovarian cancer has five main subtypes – high-grade serous ovarian cancer (HGSOC), low-grade serous ovarian cancer (LGSOC), clear cell tumors, mucinous and endometroid. Prostate ovary testis embryo expression (POTE) is a cancer testis antigen (CTA) family that consists of 14 paralogs distributed into three different groups based on phylogenetic evidences and research. Group I include POTEA, while, group II includes POTEB1, POTEB2, POTEB3, POTEC and POTED, group III includes POTEE, POTEF, POTEG, POTEH, POTEI, POTEJ, POTEKP (a pseudo-gene) and POTEM paralogs. Group III POTE paralogs – mainly POTEE, have been studied and identified as the main trigger for causing various cancers such as – non-small cell lung cancer (NSCLC), pancreatic cancer, colon cancer, ovarian cancer etc. To treat ovarian cancer, medical oncologists usually prescribe synthetic medications and/or surgeries. Synthetic medications that have been approved by the World Health Organization (WHO) are – cisplatin, carboplatin, gemcitabine, paclitaxel, etoposide, ifosfamide, mesna, vinblastine, and bleomycin. However, several studies suggest these synthetic medications cause adverse drug reactions like alopecia, nerve problems, weight loss, fatigue, diarrhoea, stomach problems, etc.

The objective of this thesis was to comprehend the role of in POTEE in ovarian cancer and also to identify the best oncologic medicaments to treat the disease. To achieve this, several bioinformatics approaches are employed and tailored for cancer biology research. The research work starts with a) Determining the evolutionary, structural and functional aspects of POTEE gene in order to infer how it is engaged in the progression of ovarian cancer. Identifying potential of POTEE as an epigenetic biomarker using various bioinformatics approaches. It proceeds to b) Mapping the potential herbal drug compounds that pass the pharmacological screenings (ADMET) and their interactions with POTEE to figure out medications are more effective –with wet laboratory validations. And then seeks c) Analysing the behaviour of various regulators (genes, transcriptional factors, miRNA) that are epigenetically altered in the ovarian cancer network. Finally, d) it combines the executed work to give a shape how ovarian cancer treatment can be made “person-centric” given the existing loopholes in computational restrictions in our country – India.

With all the relevant and up-to-date information in hand, we investigated the role of POTEE paralog in ovarian cancer by deploying bioinformatic and systems biology approaches. This was our first objective. We elucidated the structural and functional aspects of POTEE. It is noteworthy to mention

here that currently there is no experimentally crystalized tertiary structure available for POTEE. SwissModel has a reserved short stretch of the actin region of POTEE (ID: Q6S8J3) only. Genomic analyses suggested that POTEE paralog has motifs such as – zinc finger factors, DNA-binding domains, and transcription factors (TFs) present in the human and mouse genome. Additionally, the best match motif identified in POTEE mRNA sequence was a 14nt long motif (CTTCCAGCAGATGT). By deploying deep-neural networking approach, a complete protein tertiary structure was developed which was further validated with the existing AlphaFold POTEE model. The developed structure suggested that POTEE tertiary structure was composed of domains, motifs, and loops. There were many helices and beta strands (residue 248-960) and loops (residues 1-246). To assess the overall stability and macromolecular energy landscape, replica exchange molecular dynamic simulation (REMD) was executed on the developed POTEE structure for 50ns. We found that POTEE model was well simulated, had fewer steric clashes and showcased a good stability and accuracy. Epigenetics-based network analysis was also executed to depict the highly significant network associators of POTEE that could discern why POTEE gets hypomethylated in ovarian cancer or any other cancer for that matter. Eight highly significant epigenetically regulated associators were identified namely – RELA, HMOX2, EZH2, p-10Y-ERBB3-1, WDR1, ERRFI1, PRG2, FMR1 and have been already discerned as epigenetic players in various gynaecological cancers and other diseases.

Currently, synthetic medications are used to treat cancers. However, it is known that they do more harm than good. There are several side effects that are observed in patients undergoing chemotherapies. Therefore, it becomes a necessity to analyse the potential of herbal compounds as suitable medications for ovarian cancer treatment. This was our second objective. Indian Ayurvedic plants have been reported to prevent and treat ovarian cancer and were retrieved using literature mining. A manual database of phytochemicals (n=100) was prepared and they were then subjected for absorption, distribution, metabolism, excretion and toxicity (ADMET) and other essential pharmacological analyses to check for the best druglike and leadlike compounds. Once the best phytochemicals were selected (n=24), they were docked to POTEE paralog. Out of 24, only six phytochemicals namely – cedeodarin, deodarin, hematoxylin, matairesinol, quercetin and taxifolin were selected as they portrayed a good binding affinity with POTEE paralog. The six complexes were then subjected to replica exchange molecular dynamics that allows macromolecular energy landscape evaluation over different temperature range. With the completion of this objective, we concluded that Ayurvedic plants namely- Cedrus deodara, Asparagus racemosus and Cedrus brevifolia and their phytoconstituents such as – cedeodarin, deodarin, matairesinol, hematoxylin, quercetin and taxifolin must be further clinically validated and checked for their efficacy towards causing or promoting programmed cell death (apoptosis) in ovarian cancer.

Our third objective was to validate our in-silico findings in-vitro on SKOV-3 ovarian cancer cell line. We examined the potential of quercetin, taxifolin (dihydroquercetin) and matairesinol from plants – Asparagus racemosus, Cedrus brevifolia and Cedrus deodara respectively. Dihydroquercetin (DHQ) showed and promoted apoptosis in SKOV-3 as observed in cytotoxicity and cell migration assay. At concentration 100µM and 250µM dihydroquercetin was able to cause cell death, proving its strong anti-cancerous activity. To further confirm the efficacy of the proposed drug, the scratch assay was performed viz., an easy and a low-cost assay for anticancer and wound healing prediction of test drugs. This assay has been shown to be very helpful in understanding prospective cancer treatments to prevent metastasis tracking cell migration and proliferation can be used to evaluate the likelihood of metastatic spread. The result obtained through this assay showed that at 24hr the cells treated with concentration of 250µM DHQ illustrated significant inhibition in migration of cancer cells in the scratch or wound region as compared to control and cells treated with a dose of 100 µM DHQ. This result showed that the test drug DHQ creating a hindrance for cancer cells to migrate from one region to the other in a dose dependent manner. Thus, with the literature evidence and our analyses results, we firmly believe that DHQ has the potential to promote apoptosis in ovarian cancer too.

Our final objective was to understand the ovarian cancer network biology and how person-centric treatment platform can be achieved. Reconstruction of ovarian cancer network using identified differentially expressed genes (DEGs) and their network associators was carried out. Gigantic data was collected, pre-processed and optimized from gene expression repositories – The Cancer Genome Atlas Program (TCGA) and Gene Expression Omnibus (GEO) housed in the NCBI. Datasets were retrieved using MeSH terms– {"gene expression of ovarian cancer", "gene expression of ovarian carcinoma", "gene expression of ovarian adenocarcinoma"} in the two repositories. From 25,229 ovarian gene expression datasets, we were able to identify 8 seed genes, out of which 6 hub-genes were found to be playing important role in the progression of ovarian cancer. The 8 seed genes were found to be targeting females of ages 50-69. Their expression at the median age corresponds to the fact usually ovarian cancer gets diagnosed in its advanced stages in females. The reason behind this is simple – lack of specific biomarkers and thus no appropriate treatment strategy for ovarian cancer. An integrative approach was used to understand the importance of each genetic element – genes, miRNAs and transcription factors, etc. Seed genes identified were namely – CLDN3, CLDN4, NFKB1, GSN, MUC16, NANOG, FKBP10 and CD274 and after Kaplan Meier survival estimation, they suggested to have a satisfactory survival median in both low and high expression. These genes are crucial and dominate ovarian cancer proliferation in females. We also executed a gene set enrichment analysis (GSEA) wherein we found that the 8 seed genes were mainly present in the membrane system comprising of the plasma membrane, endomembrane system, and organelle lumen. They are involved with functions such as binding and regulation of different processes and are responsible in regulating different pathways such as –cell adhesion, MAPK signalling pathway, transcriptional mis-regulation in cancer, apoptosis, etc. Reconstruction of the gene regulatory network was done in GeneMania. To simplify the network and to identify crucial miRNAs and small proteins, transcription factors in the network, we applied label propagation algorithm (LPA). We found only 6 hub-genes namely –CLDN4, GSN, CD274, NANOG, FKBP10 and NFKB1 were showing strong associations with miRNAs, proteins and other genes. This sub network was further checked and validated using KEGG and GO databases. The hub genes confirmed to play important role in different cancers, regulating signalling pathways such as – MAPK pathway, TNF pathway, PI3K-Akt pathway etc. To conclude the thesis, we finally penned a summarised crux describing all the important objectives and our findings pointwise. Moreover, we also highlighted the potential challenges and future prospects for commencing personalized platform for diagnosis and treatment of ovarian cancer in India.