ABSTRACT

Breast cancer is a major global health issue and one of the most common malignancy affecting women globally. Addressing the population-specific risk factors, rare genetic alterations, and disease progression is challenging, and it is crucial for developing targeted therapy and reducing the impact of breast cancer treatment on affected individuals. The comprehensive four-phase study aimed to analyse breast cancer patient's risk factors, genetic alterations, and disease progression. The investigation commenced with an epidemiological survey to identify risk factors prevalent in breast cancer patients residing in Tamil Nadu, followed by whole exome sequencing to explore if there are any rare genetic variants in the patient's primary breast tumor, which may be population-specific. Sanger sequencing was performed to validate the reliability of the identified novel variants. With the advent of artificial intelligence (AI) in all fields, especially the healthcare sector, we tried to use AI to detect breast cancer progression accurately using histopathology images. The convolutional neural network (CNN) was applied to histopathology images to assess the mitotic index for evaluating proliferation in breast cancer patients. The results of an epidemiological study of 517 patients highlighted the significance of demographic characteristics, lifestyle choices, and treatment modalities for breast cancer. Our investigations ascertained that women aged 41-50 were the most affected, having a mean age of 47.40 (SD 11). The increased incidence rate was observed in women with lower levels of education and who were unemployed. Among reproductive factors, menopausal status was found to be a significant determinant, with premenopausal women facing a higher risk ($p<0.001$). Lifestyle factors showed that less physical exercise/activity ($p<0.01$), water intake ($p<0.05$), and sleep duration ($p<0.01$) displayed significant associations with breast cancer incidence. We also identified significant associations between comorbidities ($p<0.01$), tumor location ($p<0.001$), treatment modalities ($p<0.001$), and breast cancer. These findings add valuable insights into the complex web of factors influencing breast cancer incidence and progression within the specific patient population. The second phase involved whole exome sequencing (WES) of 6 primary breast cancer patient samples (5-Female, 1-Male), which covers all the subtypes of breast cancer. We

identified 857 rare genetic variants and found that the common associated disease among five patients is cardiomyopathy. The results indicate that complete cardiac evaluation is necessary before breast cancer therapy. Functional enrichment analysis exhibited that the rare variant candidate genes are mostly involved in biological processes that include signal transduction, G-protein coupled receptor, gene expression regulation, synapse assembly, and the cellular process of the genes occurred in the plasma membrane, cytosol, and cytoplasm, and the molecular function of the genes mainly involved in protein-protein binding. Dysregulation of these functional processes leads to oncogenic properties of cancer cells. The third phase validated the novel genetic variants *MYBPC3*: c.2816G>A and *PTCH1*: c.1889G>A for the first time through Sanger sequencing. Interactome network mapping and KEGG pathway analysis showed that the interacted genes of *MYBPC3* and *PTCH1* are mainly involved in the hedgehog signaling pathway, p53 signaling pathway, circadian rhythm, and cardiomyopathy. Mutation in *MYBPC3* and *PTCH1* leads to a mutation in the interacted genes, which alters the pathways and may cause both breast cancer and cardiomyopathy. Epidemiological factors such as high blood pressure, diabetes and disturbed sleep patterns of cancer patients may be one of the driving forces causing these changes in health conditions in the specific population. In the fourth phase, the convolutional neural network achieved high accuracy (0.966) in detecting mitosis in histopathology images, offering a promising method for identifying breast cancer progression at early stages which is usually missed out or neglected. We are the first to identify mitosis in real-time patient histology images. Integrating epidemiological, genetic, and image analysis approach opens the door for innovative strategies for prevention, treatment, and comprehensive care in the battle against breast cancer. The outcome of the study highlights personalized treatment options based on an individuals' risk factors and genetic alterations thereby offering hope for improved patient outcomes and a better quality of life.