

INTRODUCTION

Breast cancer is the most frequent cancer among women and a leading cause of cancer-related fatalities worldwide. Breast cancer caused more than 2.3 million new diagnoses and 685,000 deaths in 2020 alone. Projections suggest that by 2030, the annual global incidence of new breast cancer cases could rise to 2.7 million, resulting in 0.87 million deaths. Breast cancer is expected to become even more prevalent by 2040, with approximately 3 million new cases and 1 million fatalities each year due to growth in the population (Arnold *et al.*, 2022). According to the National Cancer Registry Programme, breast cancer is the leading cause of cancer among Indian women, and more than 50% of Indian women are diagnosed with advanced stages 3 and 4 (Sathishkumar *et al.*, 2022). Population-based cancer registries (PBCRs) serve as a robust framework for examining the prevalence of cancer across various regions and periods in India. Numerous studies on cancer burden, including the latest research conducted by the Indian Council of Medical Research-National Centre of Disease Informatics and Research (ICMR-NCDIR), have indicated that Tamil Nadu, Telangana, and Karnataka, located in southern India, exhibit a high incidence of breast cancer compared to other regions in the country. The low survival rate of breast cancer among Indian women can be attributed to a lack of awareness and poor early screening and diagnosis (Kulothungan *et al.*, 2024). Although breast cancer predominantly affects women, it is important to note that men too are affected with much lower frequency, accounting for 1 in 100 diagnosed cases (Campos *et al.*, 2021).

Breast cancer, like many other types of cancers, has an intricate and multifaceted origin. Tumors of this nature display remarkable diversity, not only between different tumors (intertumoral heterogeneity) but also within the same tumor (intratumoral heterogeneity). This diversity manifests across various levels, including epidemiological and clinical stages, as well as variances in morphology, genomics, transcriptomics, and proteomics (Januškevičienė and Petrikaitė, 2019). This complexity gives rise to unique clinical paths and poses considerable challenges in diagnosing, predicting outcomes,

and determining individual treatment approaches. These complexities underscore the critical need for research and continuous advancements in treatment modalities (Ramón y Cajal *et al.*, 2020).

Breast cancer exhibits in diverse forms, with ductal carcinoma *in situ* being the most commonly diagnosed type, followed by lobular carcinoma *in situ*. Additionally, breast cancer is characterized by its heterogeneity, with distinct molecular subtypes determined by the presence or absence of receptors such as the Estrogen Receptor (ER), Progesterone Receptor (PR), and Human Epidermal Growth Factor Receptor 2 (HER2). The lack of effective treatment plans for all disease subtypes renders it a worldwide health concern. Consequently, it should be considered that treatment should be customized for each patient based on their molecular traits (Satpathi *et al.*, 2023). Breast cancer can have either hereditary or sporadic origin. In hereditary breast cancer cases, the most common mutations are in the DNA damage repair (DDR) genes such as *BRCA1*, *BRCA2*, and *TP53*. On the other hand, sporadic cases account for approximately 85% of all breast cancer cases observed (Wendt and Margolin, 2019).

The risk factors for breast cancer can be classified into two categories: non-modifiable intrinsic factors and modifiable external factors. Intrinsic factors encompass age, sex, race, exposure to endogenous steroid hormones, and genetic susceptibility. These factors are primarily determined by one's genetic makeup and biology (Britt *et al.*, 2020). On the other hand, modifiable external factors are influenced by lifestyle choices and medical interventions. This includes the number of pregnancies, alcohol consumption, smoking, dietary habits, physical inactivity, and prolonged use of pharmacological interventions such as oral hormonal contraceptives or hormone replacement therapy. Individuals have changes in their lifestyle or medical decisions can alter their impact on breast cancer risk (Taha and Eltom, 2018). Only the modifiable external risk factors can be managed from the risk of breast cancer. Epidemiological studies are vital in identifying and understanding the modifiable risk factors. Researchers and multidisciplinary healthcare professionals can leverage this knowledge to develop preventive strategies (Iacoviello *et al.*, 2021). This underscores the importance of primary healthcare approaches and emphasizes the potential for proactive, lifestyle-based prevention to combat this prevalent disease (Osei-Afriyie *et al.*, 2021).

Risk factors are pivotal in driving genetic alterations and the progression of breast cancer. Breast cancer exhibits diverse risk factors and rare variants across different populations worldwide. Rare variants typically refer to genetic alterations that occur at a very low frequency within a population. These variants can involve single nucleotide changes (single nucleotide variants or SNVs), insertions, deletions, or other structural changes in the DNA sequence (Momozawa and Mizukami, 2021). Rare variants are the main reason for disease development in the specific population compared to the common mutation in breast cancer. Addressing the population-specific risk factors, rare genetic variants, and disease progression is challenging (Hu *et al.*, 2021).

Identifying rare gene variants crucial for drug metabolism and response poses a significant challenge. These variants, when identified, can be instrumental in guiding treatment decisions. The inherent heterogeneity of many tumors, combined with the accumulation of rare variants in both inherited (germline) and tumor-specific (somatic) genomes, plays a pivotal role in enhancing tumor malignancy. These intricacies complicates treatment strategies (Lauschke and Ingelman-Sundberg, 2018). Intratumor mutations can lead to the occurrence of rare variants that are resistant to specific treatments, contributing to treatment failure and subsequent relapse after chemotherapy. Identifying the rare genetic alterations helps to develop targeted therapy based on the individual's genetic profile. Thus, it is highly relevant to consider these rare mutations for cancer therapy, which require intense coverage of NGS-based sequencing of tumors (Yadav *et al.*, 2023).

Over the past two decades, there have been remarkable advancements in sequencing genes, exomes, and entire genomes. This technological breakthrough has revolutionized the capacity to thoroughly profile tumors and identify clinically relevant mutations caused by risk factors (Ross *et al.*, 2020). This represents a fundamental principle of 'precision cancer medicine' and has paved the way for innovative treatments tailored to specific breast tumors (Logsdon *et al.*, 2020). Next-generation sequencing (NGS) holds a crucial role in the fields of genomics, particularly in the areas of diagnosis, prognosis, and treatment. The rapid and precise capabilities of next-generation sequencing (NGS) have catalyzed the rise of precision medicine, which tailors treatments to an individual's specific disease-causing molecular changes (Lappalainen *et al.*, 2019).

NGS's development has expanded the scope of gene screening, enabling comprehensive testing of a broader range of genes in a single test. This is particularly beneficial as it can detect alterations even with limited biopsy tissue. As new biomarkers are validated for various tumors, NGS becomes increasingly practical compared to testing only a limited set of alterations. Additionally, NGS identifies well-known common alterations and uncovers the "long tail" of rare mutations, each occurring in less than 1% of patients. These rare mutations can sometimes offer crucial insights into drug sensitivity, enhancing the potential for more personalized and effective treatment approaches (Ruiz-Garcia and Astudillo-de la Vega, 2019).

Notably, WES (Whole exome sequencing) stands out for its significance in medical genetic research, as it can efficiently cover nearly all protein-coding regions within the human genome, encompassing around 85% of disease-causing genetic variants (Staaf *et al.*, 2019). Medical conditions with a genetic basis often result from various DNA alterations. Whole Exome Sequencing (WES) is primarily utilized for identifying Single Nucleotide Variants (SNVs), Single Nucleotide Polymorphisms (SNPs), and small insertions or deletions (indels) within the coding regions of the genome. The utilization of WES has been instrumental in examining the genomic landscape of rare variants in breast cancer (Suwinski *et al.*, 2019).

However, NGS has universally recognized that the variants it identifies, including point mutations and small deletions/duplications, should undergo validation through Sanger Dideoxy Terminator sequencing, which is the gold standard of DNA sequencing (Qin, 2019). Sanger sequencing retains its indispensable status in clinical genomics, providing a robust to confirm the presence of specific rare genetic alterations. It has served diverse purposes for several decades, including delineating the mutational spectrum of tumors and identifying constitutional genetic variants in diagnostic tests. The combined use of NGS and Sanger sequencing ensures comprehensive and precise genetic analysis, contributing significantly to advances in genomics and personalized medicine (Lu *et al.*, 2019).

Pharmacogenomics is an emerging approach to precision medicine that tailors drug selection and administration based on a patient's genetic profile. The genetic

makeup of a patient can significantly influence how they respond to cancer therapies, including chemotherapy, targeted therapy, and immunotherapy. Understanding the genetic variations can help oncologists tailor treatment plans that are more effective and potentially have less side effects for individual patients (Cecchin and Stocco, 2020). Identifying genetic variants through pharmacogenomic testing can provide insights into a drug's pharmacokinetics and pharmacodynamics. This information is precious for selecting the most appropriate cancer treatments and optimizing dosing regimens. Rare variants can affect the metabolism of chemotherapy drugs. Patients with specific variants may metabolize these drugs more slowly or quickly than average, affecting drug efficacy and toxicity. By identifying these variants, clinicians can adjust drug doses to achieve the desired therapeutic effect. Additionally, pharmacogenomics can aid in predicting the risk of adverse drug reactions (ADRs) in cancer patients. Genetic testing enables identifying patients who may benefit most from these therapies by assessing the presence of particular mutations (Ruiz-Garcia and Astudillo-de la Vega, 2019).

In the context of cancer, tumor cells display distinct characteristics or hallmarks that set them apart from normal cells. These hallmarks include uncontrolled proliferation, genomic instability, and the ability to evade apoptosis (Huang *et al.*, 2021). The cancer cells undergo specific modifications in cell signaling pathways, which include PI3K/AKT/mTOR Pathway, p53, MAPK/ERK, Wnt/ β -catenin, NF- κ B (Nuclear Factor-kappa B), notch signaling, TGF- β (Transforming Growth Factor-beta), and hedgehog signaling pathway. These pathways are essential for transmitting signals within cells and coordinating their functions (Sahu and Pattanayak, 2020). When these pathways are altered by risk factors, it leads to the uncontrolled growth and survival of tumor cells. Understanding these molecular mechanisms is crucial in cancer research and treatment. It provides insights into the development of targeted therapies to mitigate these alterations, ultimately contributing to more effective cancer treatments and improved patient outcomes (Glaviano *et al.*, 2023).

The proliferation of breast cancer cells is a crucial consideration in diagnosis and treatment decisions. Within the realm of diagnosis, the mitotic cell count emerges as a crucial biomarker for assessing the proliferation rate and aggressiveness of breast cancer. Traditionally, pathologists have manually scrutinized breast cancer histopathological

images under high-resolution microscopes to identify mitotic cells for assessing proliferation. However, this process is laborious, subjective, and time-consuming, primarily because of the difficulty of distinguishing between mitotic and normal cells (Ibrahim *et al.*, 2022).

Despite the significant advancements in medicine, achieving effective disease diagnosis remains a global challenge. Artificial Intelligence (AI) can potentially revolutionize different aspects of healthcare, particularly in diagnosis (Muhammad and Bria, 2023). AI integration in healthcare provides enormous potential to improve patient care and outcomes. Predictive analytics driven by artificial intelligence (AI) holds promise for improving the precision, efficacy, and financial viability of laboratory testing and disease diagnosis (Uzun Ozsahin *et al.*, 2022). Furthermore, by providing precise, real-time information and optimizing medicine selection, AI can play a critical role in population health management and guideline formulation. Recent studies have demonstrated that correlations between medical imaging features and gene expression pathways can offer insights into lesion genetics. For instance, imaging characteristics of tumor size often correlate positively with gene expression related to cell proliferation and signaling pathways. This integration of imaging data with molecular insights can provide a more comprehensive understanding of disease biology, aiding in more targeted and personalized treatment approaches (Salih *et al.*, 2023).

Artificial intelligence (AI) advancements in the field has paved the way to identify mitotic cells in histopathology images. AI algorithms can analyse large volumes of histopathology images much faster than human observers, enhancing diagnostic accuracy and providing consistent results at low cost (Pan *et al.*, 2021). The accurate detection of mitotic cells using deep learning in histopathology images plays a vital role in identifying the proliferation rate of breast cancer, which, in turn, informs critical decisions regarding the choice of treatment. High mitotic rates often correlate with more aggressive forms of breast cancer, implying a higher likelihood of rapid tumor progression and increased metastatic potential. The ability to accurately assess mitotic activity through advanced image analysis technologies not only contributes to a more personalized approach to treatment but also enhances the overall prognostic value of breast cancer diagnosis. These improvements hold the potential of improving the

accuracy and precision of breast cancer diagnosis, leading to better outcomes for patients (Pantanowitz *et al.*, 2020).

The primary goal of personalized breast cancer medicine is to provide the most effective treatment by categorizing subtypes and tailoring regimens based on patient characteristics and therapy response. Specific therapeutic strategies target individual molecules or mutated genes to identify genetic variations driving selective cell growth in tumors (Sarhangi *et al.*, 2022). Biomarkers, therapeutic targets, and data mining play crucial roles, assisting clinicians in selecting effective regimens to overcome the side effects of therapy and drug resistance. The detection of gene mutations and proliferation has significantly impacted the management of diseases, emphasizing a personalized approach focused on tumor-specific molecular alterations (Jeibouei *et al.*, 2019).

The existing knowledge on breast cancer risk factors, rare genetic alterations, and their role in disease progression is still incomplete in the Tamil Nadu population. So, the research study aimed to evaluate the epidemiological risk factors, genetic variants, and disease progression associated with breast cancer patients, which play a significant role in prevention and diagnosis and pave the way for personalized treatment.

1.1 Hypothesis

The hypotheses for the present study are as follows:

1.1.1 Null hypothesis

- There is no significant association between a specific population's epidemiology profile and breast cancer incidence.
- There are no novel genetic variants associated with breast cancer in the specific population.

1.1.2 Alternate hypothesis

- There is a significant association between the epidemiology profile of a specific population and breast cancer incidence.
- There are novel genetic variants associated with breast cancer in a specific population.

1.2 Objectives

The research design was framed with the following objectives:

- To assess the epidemiological profile of breast cancer-associated risk factors in the Tamil Nadu population - hospital based cohort study.
- To identify rare genetic variants associated with breast cancer by whole exome analysis.
- To validate the identified novel genetic variants using Sanger sequencing.
- Functional and pathway analysis of the rare variants associated with candidate genes.
- To detect mitosis and evaluate breast cancer progression using histopathology images with deep learning methods.

Extensive research literature about the current study was gathered and examined to gain insights into the existing state of research within this particular field of study. A brief review of the literature is presented in the next chapter.