4

RESULTS AND DISCUSSION

Globally, cervical cancer continues to rank among the most common cancers to affect women. Screening for cervical cancer started many years ago to reduce mortality and complications by identifying the onset of the disease early. However, some regions still face challenges in achieving comprehensive screening coverage (Kaur *et al*., 2023). Challenges to HPV vaccine adoption in India include cost, competing health priorities, limited data on HPV-related cancer, and cultural barriers, including concerns about vaccine safety and awareness gaps (Shetty *et al*., 2019). Increased public health education on HPV and cervical cancer can drive higher vaccination rates and screening participation. However, cultural and misinformation barriers remain in LMICs, reducing access to these services (Shrestha et al., 2018). Research supports that targeted interventions can bridge these gaps, and hospital-based studies are essential for gathering sociodemographic and biological risk data to guide effective policies and strategies.

Advances in next-generation sequencing (NGS), especially whole exome sequencing (WES), have transformed cervical cancer research by identifying genetic variants contributing to the disease's onset and progression. WES, which focuses on the exome—the 1% of the genome encoding proteins and where most disease-causing mutations occur—enables the discovery of mutations that increase cervical cancer risk and affect its progression and treatment response (López *et al*., 2019). WES maps the genetic alterations in cervical cancer and enables personalized therapies, including targeted and immunotherapies, which use genetic biomarkers to improve treatment precision and minimize side effects (Wang *et al*., 2023). Findings from WES are confirmed with Sanger sequencing to reduce false positives and verify mutation pathogenicity, a step crucial for clinical applications (Gargis*et al*., 2012). By merging genetic research with public health efforts, more culturally sensitive and effective prevention and treatment approaches can be developed (Lehmann *et al*., 2019).

Hence, in the present investigation, we would like to evaluate the knowledge and create awareness about cervical cancer and its preventive measures. Concurrently, we would like to explore cervical cancer patient's sociodemographic profiles and biological risk factors, followed by identifying the most prevalent histological type of cervical cancer in the study population. Further, our study aims to identify novel variants of the most prevalent histological cervical cancer type.

Phase 1

The study's first phase was designed to gather data regarding the level of awareness and measures available to screen for cervical cancer. We initiated our study during the onset of the first COVID pandemic; we decided to do an online response collection, especially from college-going female groups. The current study assessed knowledge about HPV infection causes of cervical cancer, its signs and symptoms and its risk factors among women in South India. The study sought to identify the sociodemographic elements affecting women's awareness of HPV and cervical cancer risk factors. The significant findings were (i) the participants were unaware of HPV infection, cervical cancer risk factors, and its symptoms, and (ii) the majority of the participants said they had never done cervical cancer screening. The responses collected are discussed below.

4.1. Sociodemographic characteristics of enrolled individuals

Sociodemographic factors serve as critical predictors of cervical cancer screening, influencing both knowledge of the disease and its prevention as well as an individual's willingness to participate in screening programs. In our study, we specifically examined several key variables, including age, employment status, marital status, income, education level, and place of residence (urban or rural), all of which have been shown to impact attitudes toward cervical cancer screening (Salehiniya *et al*., 2021).

Our research included 2,145 eligible women through an online questionnaire, and 2,100 of these participants completed all questions, resulting in an impressive response rate of 94.8%. This high engagement level confirms the importance of the topic, and the willingness to narrate personal experiences and

knowledge associated with cervical cancer is very crucial. The age of the study participants ranges from 18 to 56 years, and this captures a large, diverse group of perspectives ranging from one life stage to the next.

The sociodemographic profile of the respondents included different aspects such as age, literacy status, marital status, economic status, and professional background. It is important to understand these aspects as they may elucidate the pattern and trend of knowledge and behavior towards cervical cancer screening. For example, education has been a major determinant for a better awareness of health issues and the need for preventive measures, such as regular screenings.

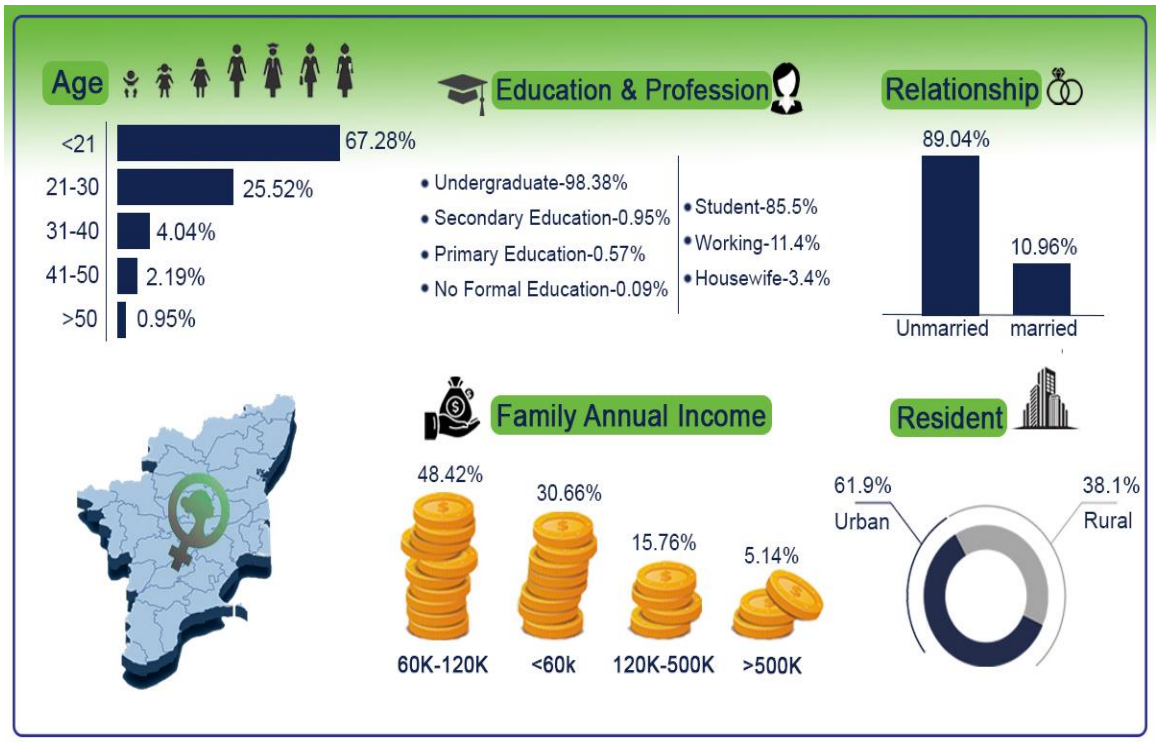
Further, the analysis of employment status of the participants gives an insight into the economic conditions that may influence their access to health care services. Marital status may also influence health-seeking behavior because married or partnered individuals may have different support systems for engaging in preventive health practices than single individuals. Additionally, income levels significantly determine healthcare accessibility, with higher-income individuals typically having better access to screening services. Figure 2 illustrates these sociodemographic variables, visually representing the distribution of characteristics among participants.

In our study, participants were categorized into two distinct groups based on their place of residence: rural and urban. Of the 2,100 participants, 800 hailed from rural areas, while 1,300 were from urban regions. This division is crucial for analyzing how geographic location may influence knowledge and behaviour regarding cervical cancer screening.

The average age of participants from rural areas was 30 years, compared to 35 years for urban residents. This difference in average age may reflect varying life experiences and perspectives regarding health practices, potentially affecting attitudes toward cervical cancer screening and awareness. Notably, a significant proportion of young women participated in this study, with 584 (73%) individuals from rural areas and 829 (63.76%) from urban regions under the age of 20. This makes it necessary to focus on the education of the younger population about

cervical cancer and the prevention of HPV, as they are most likely to benefit from learning about these issues early on.

Figure 2. Participant profile: socioeconomic and demographic factors



Most of the participants from both groups reported being single, with 89.6% of the rural participants and 88.69% of the urban participants reporting their marital status as single. This commonality may suggest that educational efforts can be effectively designed to resonate with younger, unmarried women who may be less aware of cervical cancer risks and prevention strategies.

The study found that nearly all participants had a college level of education; this includes 98.1% of rural respondents and 98.5% of urban respondents in meeting this criterion. This high educational background can relate to higher awareness and understanding of cervical cancer and its prevention and it may indicate that educational status will positively influence health literacy and motivation to join screening programs.

Economic status was measured based on annual income, and close to half of the respondents within the rural and urban categories made between 60,000 and 120,000 per year, with 45% from the rural areas and 50.5% of urban regions falling

in this income category. This can be interpreted that most of the respondents have an economic base of stability, which may impact access to health services, such as cervical cancer screening.

Researchers have conducted several studies to measure the awareness and knowledge of human papillomavirus infection and cervical cancer among women in rural as well as urban India. A notable study published by Reichheld *et al*. in 2020 explicitly focused on rural women who had only received elementary school education. The findings revealed that this demographic possessed a limited understanding of cervical cancer and its implications, highlighting the urgent need for targeted educational interventions in such populations. The low rates of HPV vaccination in India can be attributed to a complex interplay of factors, including socioeconomic challenges, social stigmas, and cultural or spiritual beliefs. These barriers significantly hinder women's access to preventive healthcare services. Our recent study further highlights this issue, particularly among educated women. Despite their higher educational attainment, a fundamental lack of awareness regarding cervical cancer persists. This finding is particularly concerning, as it suggests that education alone may not be sufficient to ensure knowledge about important health issues.

Interestingly, the rural and urban communities surveyed in our study found the results statistically significant, with a p-value of 0.05 (Table 1). This suggests that the lack of awareness about cervical cancer screening is a pressing issue across different educational backgrounds and geographical locations. To address this knowledge gap, it is imperative to develop targeted awareness campaigns and educational programs that not only inform women about cervical cancer and HPV but also address the sociocultural barriers that inhibit access to vaccination and screening services.

Table 1. Sociodemographic profile of women from both rural and urban community

Variable	Rural		Urban	
	n=800	%	n=1300	%
Age (years)				
≤20	584	73	829	63.769
21–30	173	21.6	363	27.92
31–40	27	3.375	58	4.5
41–50	12	1.5	34	2.6
≥51	4	0.5	16	1.2
Education				
No formal education	1	0.125	1	0.07
Primary	4	0.5	8	0.6
Secondary	10	1.25	10	0.7
College	785	98.1	1281	98.5
Marital status				
Married	83	10.37	147	11.3
Unmarried	717	89.6	1153	88.69
Family annual income (in Rupees)				
Below 60,000	315	39.3	329	25.3
60,000 to 1,20,000	360	45	657	50.5
1,20,000 to 5,00,000	98	12.25	233	17.9
Above 5,00,000	27	3.37	81	6.2
Profession				
Student	688	86	1108	85.2
Housewife	31	3.87	41	3.1
Employed	81	10.12	151	11.6

4.1.1. Knowledge about HPV infection, and cervical cancer

A chronic infection with the Human Papillomavirus (HPV) leads to cervical cancer. Our study population's knowledge about human papillomavirus among females in rural and urban communities was quite good (74.5% and 70.38%, respectively). The results of the current study showed that participants' knowledge about cervical cancer influenced how they evaluated services for cervical cancer screening. In this regard, the question "Are you aware of cancer cervix?" 63.1% from rural and 71.3% from urban areas responded that they knew about It. In rural areas, 35.75% of persons were unfamiliar with cervical cancer, whereas 18.8% were unfamiliar in urban areas (Table 2). It is a big concern because there is only modest awareness of cervical cancer in both rural and urban areas, which may hinder early detection, diagnosis, and treatment of the condition.

Table 2. Knowledge about cervical cancer and HPV

Particulars	Rural	%	Urban	%	P value
Have you heard about cervical cancer?					0.0001*
I know	505	63.1	927	71.3	
I don’t know	295	36.87	373	28.69	
Do you know about HPV?					0.0454*
I know	596	74.5	915	70.38	
I don’t know	204	25.5	385	29.6	

According to a prior study done in North India, 57.1% of people in urban settlements and 49.7% of people in rural ones knew about cervical cancer (Kadian et al., 2021). Similarly, it has been documented worldwide that women with different demographic setups and other attributes are aware of cervical cancer. For instance, our results align with a study conducted in Cambodia (Touch and Oh, 2018) and Central Ethiopia (Gebisa *et al*., 2022), where 74% of study women had heard about cervical cancer. Our finding revealed better awareness than the study reported in Bangladesh, where 45.2% (Alam *et al*., 2022); in Pakistan, where

51.3% (Riaz *et al*., 2020); in Congo, where 55% (Ali-Risasi *et al*., 2015) and in Zambia (Nyambe *et al*., 2019) of them were having enough knowledge about cervical cancer. In a study conducted in Gabon, 91.6% (Assoumou *et al*., 2015), in Tanzania 83.1% (Mabelele *et al*., 2018), and in Bangladesh where 89% from urban and 93.4% from rural (Islam *et al*., 2018), study participants knew about cervical cancer which is lower observation compared with our study, this gap may be caused by the different sample sizes, educational interventional tactics, and poor awareness creation levels in the study areas.

4.1.2. Source of information on cervical cancer

In rural areas, 17.5% of the respondents reported that they gained knowledge about cervical cancer from education; however, 26.6% of urban respondents had the same opinion. Apart from that, 16.87% of the rural women and 22.6% of the urban women acquired information about cervical cancer from the media. Transferring information about cervical cancer from healthcare providers, friends, and family members was very less in rural and urban regions (Table 3). The importance of education resources and media can be seen here because education is essential to raise awareness of cervical cancer in women. Now, with easy accessibility to media like radio and television, information can be relayed to a wider range at an affordable cost (Tadesse *et al*., 2022).

Primary cervical cancer prevention is far more feasible than secondary or tertiary cervical cancer treatment. Understanding the complexities of cervical cancer, identifying risk factors, and creating and implementing appropriate treatments are all cost-effective techniques (Muluneh *et al*., 2019; Taye *et al*., 2021). It is not shocking that literacy and place of residence are related to CC knowledge. This demonstrates that even educated persons have limited understanding and awareness of cervical cancer and immunization practices.

Table 3. Source of information regarding cervical cancer

Particulars	Rural	%	Urban	%
Source of information about cervical cancer				
Television	88	11	129	9.9
Newspaper/magazines	30	3.75	67	5.1
Medical practitioners	43	5.37	95	7.3
Friends/neighbours/relatives	78	9.75	124	9.5
Internet	135	16.87	294	22.6
Education	140	17.5	346	26.6
I don’t know	286	35.75	245	18.8

4.1.3. Knowledge about cervical cancer prevention strategies

The saying "An ounce of prevention is worth a pound of cure," credited to Benjamin Franklin in 1735, is better fitted in medicine. Prevention has made a significant role in oncology, as cancer is one of the major present challenges in public health. Knowledge about cancer, prevention, and early detection are fundamental to effective screening tests (Wilailak *et al*., 2021). The factors causing limited participation in cervical cancer screening include lack of information/ awareness, low perceived risk, late recognition of initial signs and symptoms, a social stigma attached to cancer, fear of the disease, financial problems, family responsibilities, and embarrassment (Taneja *et al*., 2021). Early detection using cytological screening, such as a Pap smear, can lower the incidence and death rate of invasive cervical cancer.

As shown in Table 4, both community group members (26.75% and 19.76%, respectively) had relatively limited knowledge about HPV vaccines and their advantages. Women who lived in urban and rural areas had low HPV vaccination rates (14.7% and 11%, respectively). More knowledge of the Pap test is required by rural and urban populations (18.6% and 27.23%, respectively). More than half

of the population, 81.37% and 72.76%, respectively, required appropriate information regarding the Pap test. Regarding women's attitudes, only 33.23% of women in urban areas were willing to undergo cytological screening, while 37.8% of women in rural areas were willing. More than half of the women from rural and urban areas said they were not prepared to participate in cytological screening, with 62.1% from rural areas and 66.76% from urban areas declining. The participants in our study were reluctant to volunteer for the first test.

Table 4. Knowledge about cervical cancer preventive measures

Variable	Rural	%	Urban	%	P value
Knowledge about Pap test					0.00001*
I know	149	18.6	354	27.23	
I don’t know	651	81.37	946	72.76	
Knowledge about HPV vaccine					0.0002*
I know	88	26.75	257	19.76	
I don’t know	712	73.25	1043	80.23	
Receiving HPV vaccine					0.0144*
Yes	88	11	192	14.7	
No	712	89	1108	85.23	
Do you think that you should undergo cytological examination					0.034*
Yes	303	37.8	432	33.23	
No	497	62.1	868	66.76	

Our findings showed better responses from participants than Hossaina's 34.8% (Aweke *et al*., 2017) and Wolaita's 45.5% (Tekle *et al*., 2020) participants who had positive opinions about cervical cancer screening. A similar observation to our study was made in Pondicherry, where 3% and 45% of individuals were

aware of the HPV vaccine and cervical cancer, respectively (Saha *et al*., 2023; Jain *et al*., 2016). The amount of information available regarding the significance of cervical cancer screening, which has a significant impact on respondents' health-seeking behaviours, might be one of the factors causing the inequalities.

The reasons for not participating in cytological screening, such as Pap smears or HPV tests, highlight key socio-cultural and psychological barriers that impede the early detection of cervical cancer, which is crucial for improving patient outcomes. These reasons are consistent across different geographic regions, but distinct cultural, educational, and healthcare dynamics shape them. Here is a scientific discussion of these factors:

According to studies conducted in Nigeria, Uganda, and worldwide, one of the most common causes is the low level of awareness related to cervical cancer and its screening measures (Ghebre *et al*., 2017). This is a learning point in public health education, indicating the need for further targeted awareness campaigns. Most women do not understand the fact that early cervical cancer rarely has few or no symptoms, creating a barrier for women in getting early preventive screenings.

These results imply that information efforts should remind cervical cancer target groups of relevant risk factors and emphasize the notion that early detection is crucial from a public health point of view. Being symptom-free does not imply being disease-free. Information interventions could rectify misconceptions but also improve preventive health-seeking behaviour with specific focus on the resource-poor setting of women.

In studies conducted in Yemen and Uganda, fear of the screening procedure and general distrust of medical procedures are commonly reported. This may result from ignorance about the procedure or a past adverse experience with the healthcare system. Women feel the test is painful or intrusive and fear taking the test. The other fear that might be present is the fear of being diagnosed with cancer, associating it with stigma or poor outcomes. Healthcare professionals should make information about screening clear and accessible to women, reassuring them that the procedure is safe and the benefits of early detection are

important. A patient-centered approach that promotes trust between patients and providers, coupled with visual aids or videos explaining the procedure, could help alleviate some of these fears (Abdul-Aziz, 2012).

In the Ugandan study, a prevalent perception among participants was the belief that they were not at risk for cervical cancer. This belief may stem from a lack of knowledge about the risk factors associated with cervical cancer, including human papillomavirus (HPV) infection, which is sexually transmitted and the primary cause of cervical cancer. In many cases, women may not consider themselves at risk if they have limited sexual partners or if they perceive cervical cancer as a disease affecting only certain groups. Educational programs must address the universal risk posed by HPV infection, particularly given its prevalence in low-resource countries. Promoting HPV vaccination as a preventive measure can also help reduce the perceived risk associated with screening and encourage proactive health-seeking behaviour (Usman *et al*., 2023).

The barriers identified in these studies highlight the importance of culturally tailored interventions in cervical cancer prevention programs. Healthcare systems must be equipped to address not only the logistical and resource challenges in low- and middle-income countries but also the social and psychological barriers to screening. Community-based interventions that engage local leaders incorporate cultural norms, and peer education models may be particularly effective in overcoming resistance to screening. Moreover, task-shifting strategies, where trained community health workers perform screening or education, could bridge the gap in regions with limited access to healthcare professionals. Efforts to normalize discussions around cervical cancer, increase access to female healthcare providers, and ensure the affordability and accessibility of screening services will be crucial in reducing the global burden of cervical cancer.

4.1.4. Knowledge about clinical spectrum of cervical cancer

In our study, the rural population identified the most commonly experienced symptoms among cervical cancer patients as painful and irregular periods (16.87%), pain or bleeding during or after sexual activity (7%), unusual or smelly vaginal discharge (4.87%), and pelvic pain (4.6%). In contrast, urban women recognized painful and irregular menstruation (20.5%), pain or bleeding during or after sexual activity (9.4%), and unusual or smelly vaginal discharge (6.1%) as typical signs of cervical cancer. Notably, over half of the women in both groups—61.75% from rural areas and 53% from urban areas—were not aware of the correct symptoms related to the progression of cervical cancer, as illustrated in Table 5.

Table 5: Evaluation of cervical cancer symptoms in women from rural and urban areas

Symptoms	Rural	%	Urban	%
Itching in genital area	25	3.1	39	3
Pain in pelvis	37	4.6	75	5.7
Irregular and painful menstrual period	135	16.87	267	20.5
Pain / Bleeding during or after sexual activity	56	7	123	9.4
Discomforting Urination	14	1.75	21	1.6
Unusual / smelly discharge from vagina	39	4.87	80	6.1
I don't know	494	61.75	695	53

Lack of symptoms or 'no reason' was the primary reason for not undergoing screening for cervical cancer. The burden of this disease can be decreased with proper awareness of cervical cancer and its symptoms. Women who are unaware of the signs and symptoms of cervical cancer are less likely to seek medical attention, which could result in cervical cancer not being found in its early stages (Qayum *et al*., 2021). Women who are less serious about symptoms are the main reason for being unaware of the severity of the illness and are reluctant to talk

about it (Kadian *et al*., 2021). According to our findings, respondents from both communities were less knowledgeable about the signs and symptoms of cervical cancer. In our survey, many women believed that symptoms such as pelvic pain, vaginal discharge, and an odorous vagina were normal. Consequently, as noted by Sudha *et al*. (2022), they did not perceive these symptoms as a cause for concern.

Our finding compares with what had been reported in similar studies worldwide. For instance, in the South African study, when asked whether women with vaginal discharge and bleeding should be screened for cervical cancer, most participants (58.8%) answered "No," and 13.7% said "Don't know" in regards to the signs and symptoms of cervical cancer (Tiiti *et al*., 2022). Out of 667, 22% and 31.6% of students indicated vaginal blood and vaginal foul-smelling discharge during sexual intercourse as symptoms of cervical cancer (Tadesse *et al*., 2022). However, many studies reported that most of the women were unaware of the link between these clinical spectrums and cervical cancer. In this study population, irregular and painful menstrual cycles, pain or bleeding during or after sexual activity, strange or unpleasant vaginal discharge, and pain in the pelvis were the most reported symptoms. Our results indicated that additional efforts were necessary to educate women about the signs of cervical cancer to help reduce morbidity and mortality.

4.1.5. Knowledge about risk factors

The awareness of risk factors associated with cervical cancer was evaluated among women in both communities. In the rural community, the most frequently recognized risk factors were HPV infection (22.87%) and having multiple sexual partners (7.25%). In the urban area, HPV infection was also the most common (24.38%), followed by multiple sexual partners, a weakened immune system (4.69%), smoking (4.46%), and long-term use of oral contraceptives (3.3%). Factors such as being overweight, early pregnancy, marrying at a young age, and multiple pregnancies were identified as the least recognized risk factors for cervical cancer among women in both rural and urban settings. Additionally, approximately 48.87% of women from rural areas and 43.46% from urban areas were unaware of these risk factors (Table 6).

Table 6. Knowledge about risk factors associated with cervical cancer in rural and urban females

Risk factors	Rural	%	Urban	%
Being overweight	5	0.625	18	1.3
Early pregnancy	13	1.62	25	1.9
Early start of sexual activity	16	2	40	3.07
Genetic factor	21	2.62	31	2.38
Having weakened immune system	35	4.37	61	4.69
HPV infection	183	22.87	317	24.38
Long term use of oral contraceptives	18	2.25	43	3.3
Marriage at early age	10	1.25	27	2.07
Multiple sexual partners	58	7.25	94	7.2
Repeated pregnancy	13	1.62	21	1.6
Smoking	37	4.6	58	4.46
I don't know	391	48.87	565	43.46

According to the findings of this study, HPV infection, having many sexual partners, smoking, having a weakened immune system, and hereditary variables are all recognized as potential risk factors for cervical cancer in the rural population. On the other hand, urban areas were thought to have a higher incidence of cancer due to HPV infection, multiple sexual partners, a weakened immune system, smoking, long-term oral contraceptive usage, early onset of sexual activity, and early pregnancy. Kadian *et al*. (2021) discovered that most women in North India were unaware that HPV infection was a substantial risk factor for cancer advancement. In Tamil Nadu, 60% of women have never heard of cervical cancer (Tamilarasi *et al*., 2018).

In Central Ethiopia, 49% of participants were unaware of the cervical cancer risk factors (Gebisa *et al*., 2022). According to the Alam *et al*., 2022 study, women who lived in urban regions were more aware of the risk factors than women who lived in rural areas. According the Bangladeshi women, having multiple sexual partners, beginning sexual activity at a young age, and using oral contraceptives for an extended period of time were the most important risk factors. The least recognized risk factors in rural and urban locations are not getting routine Pap screening test and having HPV. Many researches have shown the relationship of having multiple sexual partners with the risk of developing cervical cancer. These findings were that people had the same amount of knowledge about cervical cancer risk factors.

The results of phase I showed that nearly three-fourths of our study participants are students with higher educational background and unaware of HPV infection, cervical cancer, and HPV vaccines. Even though most respondents of the survey answered "yes" to whether they knew of cervical cancer, they were uninformed of the correct symptoms or risk factors. Our study showed that knowing cervical cancer does not mean that a person knows more about the disease and its preventive measures. This outcome was aimed at finding out more about cervical cancer. This survey may benefit the participant to understand the risk factors and preventive measures associated with a cervical cancer.

Further, the subsequent phase of the study elaborated information about the lifestyle, clinical profile, and treatment outcomes of cervical cancer patients. This fundamental information about a patient's clinical profile can aid in predicting personalized medicine as a therapy option. A patient's clinical profile includes a patient's medical history, lifestyle, and disease characteristics, such as the histology type of cervical cancer. Understanding these factors helps doctors tailor treatments, such as targeted therapies or personalized screening schedules, to a patient's unique needs.

Phase II

4.2 Epidemiological profiling of cervical cancer patients

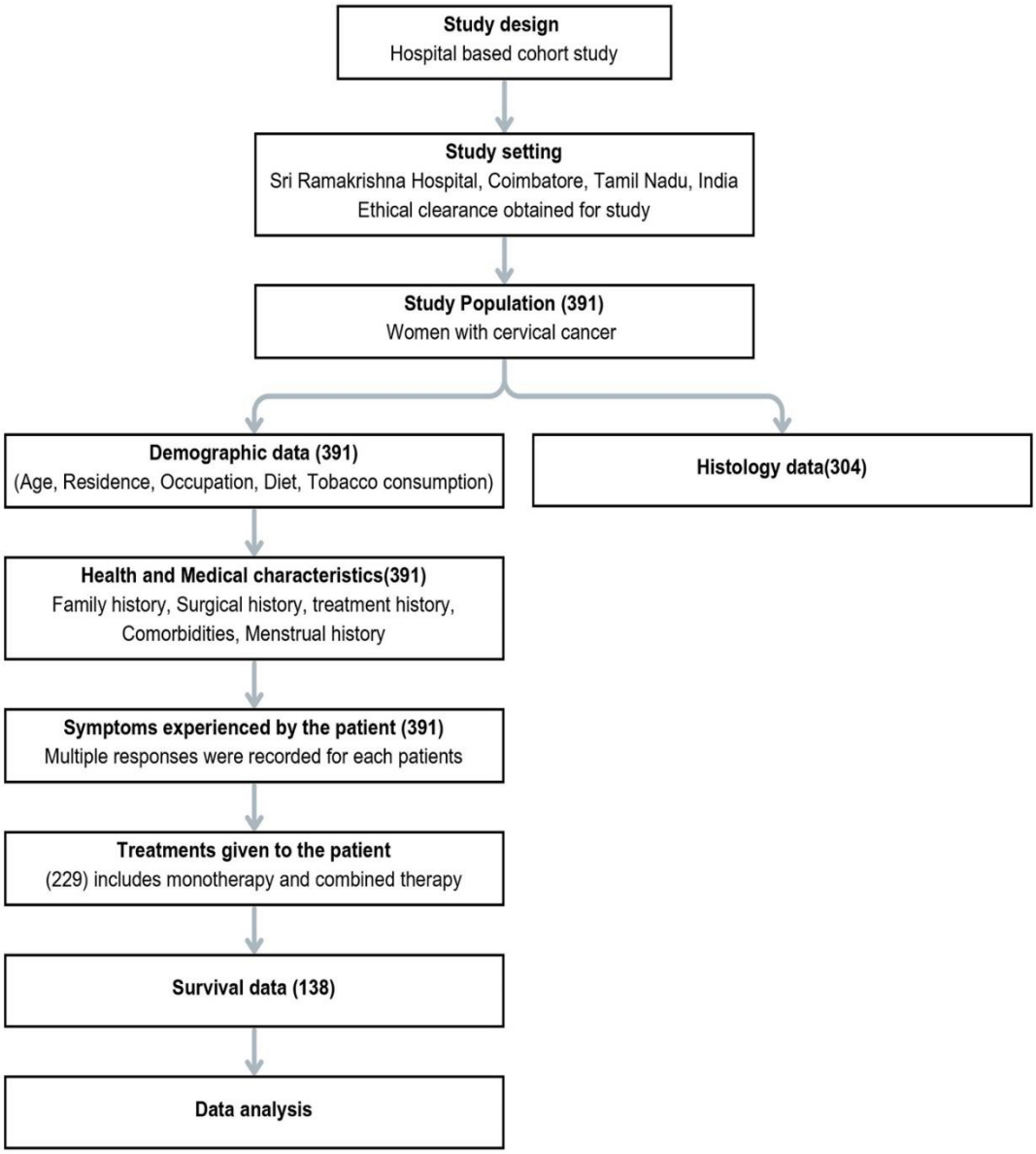
In the next phase, we conducted a hospital-based cohort study at Sri Ramakrishna Hospital in Coimbatore, Tamil Nadu, India, which aimed to explore various demographic, health, medical, and treatment-related factors in women diagnosed with cervical cancer. Ethical clearance was obtained prior to the study. This phase is designed to understand the sociodemographic profile of the participants, the symptoms they experienced, and the treatment options they opted for. The results were statistically analysed based on responses collected through a structured questionnaire.

As shown in Figure 3, a total of 391 women with cervical cancer were enrolled in the study. All participants provided responses documenting their sociodemographic characteristics, including age, occupation, diet, weight loss, and loss of appetite. The study also recorded various clinical signs and symptoms, each experiencing different or multiple symptoms. Hence, multiple responses for clinical signs and symptoms were documented for all the patients. Histological data were available for 304 participants, shedding light on the cellular characteristics of the disease and its progression.

Regarding treatment options, 229 participants received treatment, with 44 patients undergoing monotherapy and 185 receiving combined therapy. Survival data was available for 138 of these 229 treated patients, providing insights into the outcomes of different treatment regimens.

This phase of the study offers a comprehensive analysis of treatment outcomes, patient demographics, symptom patterns, and the histological nature of cervical cancer. These findings will contribute to a deeper understanding of the factors influencing treatment efficacy and survival in women diagnosed with cervical cancer in similar settings.

Figure 3. Comprehensive study of cervical cancer: patient data, symptoms, treatment, and outcomes



4.2.1. Demographic profile of cervical cancer survivors

During the study's 2019–2022 time-frame, 391 patients visited the Sri Ramakrishna Hospital in Coimbatore. The socio-behavioural characteristics of cervical cancer are summarised in Table 7. The socio-demographic and lifestyle characteristics of the patient population revealed several key trends. Most patients were middle-aged to elderly, with the highest proportion in the 50-59 age group (31.45%), followed by the 40-49 years group (27.87%). Patients aged 60-69 accounted for 21.73%, and those aged 70-79 made up 11.25%, while a small percentage were over 80 (1.27%). Younger patients below 40 years were relatively rare, with only 0.51% under 30 years and 5.88% between 30-39 years. The median age of patients was found to be 55.69 years. The majority of patients lived in rural areas (74.42%), while urban residents represented a smaller proportion (25.58%). Regarding occupation, the majority were homemakers, who make up more than half of the population (52.69%). The next largest group are daily wage workers (29.16%), followed by smaller groups in occupations such as agriculture (6.91%), health workers (4.86%), and textile mills (2.56%). A small percentage of them were employed as teachers (1.02%) or in other sectors (2.81%). In terms of diet, a significant majority of patients followed a mixed diet (92.84%), while only 7.16% adhered to a vegetarian diet. This dietary preference could be relevant for understanding health outcomes within this population. Furthermore, a notable proportion of patients (20.72%) have a history of tobacco consumption, indicating a substantial level of exposure to this risk factor. The remaining 79.28% do not have a history of tobacco use.

These findings highlight the need for targeted cervical cancer screening and prevention in rural areas, focusing on middle-aged to elderly homemakers and daily wage workers. Dietary interventions could aim to promote balanced eating habits, while tobacco cessation programs may not be as critical for this group.

Table 7. Socio-behavioural characteristics of cervical cancer patients

Variables	Number of patients (%)	%
Socio demographic characteristics		
Age (Years)	(N=391)	
Below 30	2	0.51
30-39	23	5.88
40-49	109	27.87
50-59	123	31.45
60-69	85	21.73
70 – 79	44	11.25
Above 80	5	1.27
Residents	(N=391)	
Urban	100	25.58
Rural	291	74.42
Occupation	(N = 391)	
Daily wages	114	29.16
Agriculture	27	6.91
Textiles	10	2.56
Health workers	19	4.86
others	11	2.81
House wife	206	52.69
Teacher	4	1.02
Lifestyle and Behavioural Variables		
Diet	N = 391	
Mixed	363	92.84
Veg	28	7.16
History of tobacco consumption	N = 391	
Yes	81	20.72
No	310	79.28

In our study, the median age of cervical cancer patients was 55 years, which is consistent with findings from various studies in India. Research suggests that cervical cancer is most common among women aged 35 and older, with the incidence peaking in women in their 50s and 60s (Sharma *et al*., 2024). For instance, Jain *et al*. (2017) reported a median age of 54 years, while Sharma and Pattanshetty (2018) found a median age of 53.2 years. Kataki *et al*. (2018) in North East India reported younger median ages (48 years), while others, like Kumar *et al*. (2020), found a higher median age of 74 years. These differences may be due to regional variations in awareness, access to healthcare, and age at diagnosis. In our study, most patients were in the 50–59 age range, suggesting that lack of awareness and screening could lead to late diagnoses.

Our study also found that 74.42% of patients live in rural areas, which is consistent with the higher prevalence of cervical cancer in rural populations due to limited access to healthcare, lower awareness, and cultural barriers to screening. This finding is supported by studies like those of Srivastava *et al*. (2018) and Thulaseedharan *et al*. (2012), which showed rural populations faced greater risks due to limited healthcare access, lower awareness, and cultural barriers to screening. In contrast, studies reported by Sharma *et al*. (2017) and Shrivastav *et al*. (2021) show a higher proportion of urban patients (58.3% and 66.9%, respectively). This could indicate that urban women have better access to healthcare services, such as screenings and vaccinations, and are more likely to seek medical care. However, the rural-urban discrepancy may also suggest underreporting in rural areas due to inadequate healthcare infrastructure, highlighting the need for targeted interventions in rural settings to improve screening and awareness.

Regarding occupation, most women in our study were homemakers (52.69%) or daily wage laborers (29.16%), which indicates that many come from lower-income backgrounds with limited access to healthcare. This may contribute to delayed detection and treatment. In contrast, Sharma *et al*. (2017) found that 49.8% of patients were manual laborers, possibly due to increased exposure to environmental carcinogens and limited healthcare access. The differences in

occupation across studies underline the need for targeted prevention and intervention strategies based on socio-economic and occupational factors.

Regarding diet, 92.84% of our patients reported consuming a mixed diet. Studies, including Sharma and Pattanshetty (2018), have shown that non-vegetarian diets are associated with an increased risk of cervical cancer. This dietary pattern could play a role in understanding the health outcomes of the women in our study, as compared to other regions like New Delhi, where 19.9% of women were non-vegetarians (Shrivastav *et al*., 2021).

These differences in age, geography, occupation, and dietary patterns across studies emphasize the need for tailored strategies to improve cervical cancer awareness, access to preventive care, and early diagnosis, particularly among rural and low-income populations in Tamil Nadu.

4.2.2. Health background and menstrual characteristics

The study examined various health and reproductive factors, including "Reproductive and Medical History Characteristics," which encompassed menstrual patterns, family history of cancer, surgical history, radiation exposure, native treatment, and comorbidities.

As shown in Table 8, the study population's health and medical profile revealed several key characteristics. Notably, a family history of cancer was reported by only 7.67% (30) of patients, indicating a relatively low genetic predisposition to cancer. In contrast, 92.33% (361) of patients had no known family history of cancer. Regarding surgical history, 30.95% (121) of patients reported no prior surgery. However, specific procedures were reported by smaller percentages, including cataract surgery (0.77%, 3), hysterectomy (0.51%, 2), lower segment cesarean section (0.51%, 2), and other surgeries (3.84%, 15). 59.59% (233) of patients' surgical history was not available, indicating a potential gap in medical record-keeping. Radiation exposure was minimal, with only 2.05% (8) of patients reporting prior radiation, while 97.95% (383) had no radiation history. Furthermore, 8.70% (34) of patients had undergone native treatment, whereas 91.30% (357) did not.

Table 8. Medical history and menstrual status of cervical cancer patients

Variables	Number (N=391)	%
Health and medical variables		
Family history of cancer		
Yes	30	7.67
No	361	92.33
History of surgery		
No	121	30.95
cataract surgery	3	0.77
Hysterectomy	2	0.51
LSCS	2	0.51
Others	15	3.84
Not applicable	233	59.59
Exposure to Prior radiation		
Yes	8	2.05
No	383	97.95
History of native treatment		
Yes	34	8.70
No	357	91.30
Comorbidities		
Yes	133	34.04
No	258	65.96
Menstrual history		
Post menopause	258	66.99
Pre menopause	133	34.01

The majority of our patients reported that they never knew of an existing family cancer history or lack thereof. Significantly, familial cervical or cancers at other sites were not indicative of increased probabilities of patient delays. This lack of association may be attributed to a considerable difference in the awareness of family health history, particularly in lower-resource settings where recall bias is more common because of limited access to family health information. Cultural factors and health anxiety may further influence recall accuracy, contributing to potential underreporting (De Vries *et al*., 2018; Awofeso *et al*., 2018). Further research is needed to understand how family history awareness might impact diagnostic delays, particularly in populations where socioeconomic and cultural factors shape health perceptions and recall.

In our study, a notable proportion of patients with cervical cancer also had chronic comorbid conditions, with 34.05% of the cohort experiencing one or more such conditions. The most frequently observed comorbidities were hypertension and diabetes, which is consistent with findings from other studies on cervical cancer populations. For instance, Khondekar *et al*. (2021) also identified hypertension and diabetes as the leading comorbid conditions among cervical cancer patients, indicating a common trend where these chronic illnesses are prevalent in similar demographic groups. Singh *et al*. (2016) further reinforced this, reporting that 41% of their patients had hypertension and 36% had diabetes. These findings underscore that, beyond the primary illness, a significant number of patients face additional health burdens, which may complicate treatment and recovery.

Kumar *et al*. (2020) observed an even higher rate of comorbidities, with 69% of their patient cohort experiencing at least one chronic condition and hypertension affecting 44.31% of their sample. The difference in comorbidity rates between studies, with ours at a moderate level and at a higher level, may be due to regional healthcare disparities, demographic variations, or lifestyle differences among the populations studied. However, despite these variations, the high prevalence of hypertension and diabetes across studies highlights an important association with cervical cancer, underscoring the need for comprehensive management of these conditions to improve patient outcomes. In contrast, Ibfelt *et al*. (2013) reported

lower rates of comorbidities in their cervical cancer patients, suggesting that some demographic factors, access to healthcare, or lifestyle differences may account for varying rates of chronic disease across populations.

Beyond comorbidities, our study also identified menopausal status as a significant risk factor, with 67% of our patients diagnosed during their post-menopausal period. This finding is aligned with other studies indicating a high incidence of cervical cancer in post-menopausal women, which suggests that hormonal changes, age-related immune decline, or other cumulative risk exposures may increase vulnerability to cervical cancer after menopause. For instance, Begoihn *et al*. (2019) reported that 76.95% of women in their study experienced disease progression in the post-menopausal phase, implying that menopause may contribute to cancer advancement. Nasreen *et al*. (2023) also found that 67.39% of their cervical cancer patients were post-menopausal, a figure close to that observed in our study, while Sharma and Pattanshetty (2018) reported an even higher rate of 90.1% in their cohort at Manipal.

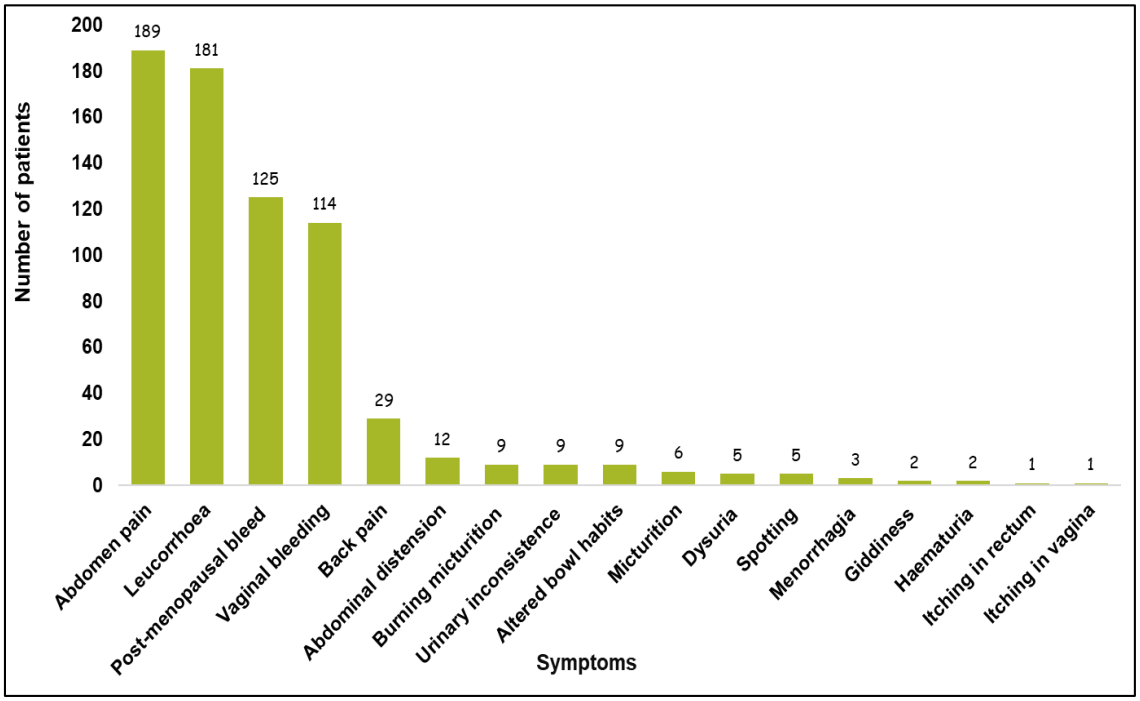
These reports suggest that post-menopausal women may constitute a higher-risk group for cervical cancer. This is possibly due to age-related physiological changes or prolonged exposure to risk factors. Given these findings, there is a strong rationale for implementing targeted screening and preventive strategies for post-menopausal women to aid in early detection and management. Overall, both the prevalence of comorbidities like hypertension and diabetes and the association of cervical cancer with post-menopausal status highlight the need for a holistic approach to managing cervical cancer, addressing not only the cancer itself but also the broader health challenges faced by these patients.

4.2.3. Signs and symptoms, and associated comorbidities

The study revealed a range of symptoms experienced by the 391 patients, providing valuable insights into the clinical presentation of cervical cancer. All patients presented with symptoms related to cervical cancer. Some of them had more than one symptom.

As shown in Figure 4, The most common symptoms reported were abdominal pain (48.33%, 189) and leucorrhoea (46.29%, 181), indicating significant gastrointestinal and reproductive health concerns. Weight loss (36.32%, 142) and post-menopausal bleeding (31.96%, 125) were also prevalent, suggesting potential advancement in disease. Vaginal bleeding (29.16%, 114) and loss of appetite (22.76%, 89) were additional prominent symptoms.

Figure 4. Clinical profile of cervical cancer patients



Other notable symptoms included back pain (7.41%, 29), abdominal distension (3.06%, 12), burning micturition (2.30%, 9), urinary inconsistency (2.30%, 9), altered bowel habits (2.30%, 9), and dysuria (1.27%, 5). Less frequent symptoms were spotting (1.27%, 5), menorrhagia (0.76%, 3), giddiness (0.51%, 2), haematuria (0.51%, 2), itching in the rectum (0.25%, 1), and itching in the vagina (0.25%, 1).

Our study observation aligns closely with previous findings, including a study by Kumar *et al*. (2020), where vaginal discharge was reported by 51.13% of patients as a primary symptom. Kaverappa *et al*. (2015) similarly found that 40% of patients experienced postmenopausal bleeding as their first sign of cervical

cancer. In earlier studies, Gundrajakuppam *et al*. (2011) reported that over half of the patients identified white discharge as their initial symptom, a finding echoed by Shruthi *et al*. (2014), where patients typically experienced white discharge followed by vaginal bleeding. These studies collectively highlight the significance of unusual vaginal discharge as a common early symptom, often unnoticed by patients.

Supporting our findings, researchers reported that a thin, watery, blood-tinged vaginal discharge is a common initial symptom among cervical cancer patients, often accompanied by mild vaginal or contact bleeding. Although early cervical cancer is frequently asymptomatic, early signs such as abnormal discharge, occasional bleeding, and, rarely, vaginal tumors may manifest (Kebebew *et al*. 2021; Maree *et al*., 2015). As the disease progresses, symptoms typically intensify, with patients experiencing excessive vaginal bleeding, pelvic and back pain, leg pain, leg swelling, urinary or fecal leakage, weight loss, and decreased appetite (Christiansen *et al*., 2022). This evidence underscores the critical importance of raising awareness about subtle early symptoms, particularly unusual vaginal discharge, which remains one of the most frequently reported initial signs of cervical cancer.

4.2.4. Cervical cancer tumour characteristics

The histological analysis of 391 cervical cancer cases provides an in-depth examination of the various tumor types within this population, revealing a range of cervical cancer subtypes (Table 9). Squamous cell carcinoma (SCC) was the predominant subtype, representing 60.61% (237 cases) of all cases. These findings align with existing studies that have consistently identified SCC as the most prevalent form of cervical cancer, with reported rates of 89% in Mumbai (Balasubramanian *et al*., 2021), 88.10% in Mumbai (Jain *et al*., 2017), 85% in Bihar (Chauhan *et al*., 2016), and 91% in Mumbai (Khondekar *et al*., 2021).

The high prevalence of SCC is likely attributable to its strong association with persistent infection by high-risk HPV strains, particularly types 16 and 18, which are known to drive cervical carcinogenesis. In countries with established screening programs, the incidence of invasive SCC has declined as preinvasive

lesions are more readily detected and managed early (Okunade, 2020). By contrast, adenocarcinoma cases may be increasing in these settings, as the endocervical location of adenocarcinoma can make it less amenable to early detection via traditional screening (Castanon *et al*., 2016). In regions where cervical cancer screening programs are limited or underutilized, SCC remains the predominant subtype, as invasive cases often go undetected until later stages (Okunade, 2020).

Adenocarcinoma (AC) was the second most common type, comprising 12.02% (47 cases), slightly higher than incidences documented in prior studies that report adenocarcinoma as less common relative to SCC. This increase may reflect changing patterns in HPV strain prevalence and improved screening practices, which could detect more adenocarcinomas than in the past. Continued tracking of these shifts in tumor subtypes could clarify trends and suggest how HPV types and screening advances affect cervical cancer epidemiology.

Table 9. Histopathological characteristics of cervical cancer

Histology	N = 391	%
Squamous cell carcinoma	237	60.61
Adenocarcinoma	47	12.02
Non keratinizing carcinoma	12	3.07
Invasive carcinoma	4	1.02
Large cell carcinoma	1	0.26
Papillary carcinoma	1	0.26
Gastro intestinal stromal tumor	1	0.26
Signet ring cell carcinoma	1	0.26
Not available	87	22.25

In addition to SCC and adenocarcinoma, rarer histological types were observed, including non-keratinizing carcinoma (3.07%, 12 cases), invasive carcinoma (1.02%, 4 cases), large cell carcinoma (0.26%, 1 case), papillary carcinoma (0.26%, 1 case), gastrointestinal stromal tumor (0.26%, 1 case), and signet ring cell carcinoma (0.26%, 1 case). Additionally, 22.25% (87 cases) needed more histological information due to incomplete records.

The predominance of SCC in our study emphasizes the need for a deeper investigation into risk factors and mechanisms driving this subtype. This trend may reflect an actual increase in SCC cases or variations in diagnostic practices. These findings highlight the need for accurate and comprehensive histological diagnosis, critical for effective treatment planning and improved patient outcomes

4.2.5. Sociodemographic characteristics of cervical cancer patients

As shown in Table 10, In a comparative study of 391 cervical cancer patients, significant differences in sociodemographic and clinical variables were observed between rural (N=291) and urban (N=100) populations. The study highlights distinct patterns that may inform targeted healthcare strategies for each group by examining lifestyle, medical history, and demographic characteristics.

Occupation was significantly associated with residence, with rural patients more likely to be employed (OR=4.42, 95% CI: 2.26–8.62, p=0.0001). This finding suggests that rural patients, potentially due to socioeconomic or lifestyle factors, are more frequently engaged in work. Employment status could affect patients' time availability and healthcare access, impacting treatment adherence and health outcomes.

Diet type showed no significant difference, with most patients reporting a mixed diet (p=0.62). This suggests that diet pattern does not strongly influence dietary habits, although dietary quality and variety could still vary, potentially impacting overall health and cancer prognosis.

Table 10. Sociodemographic and clinical profile

	Total	Rural N=291	Urban N=100	Odds ratio	95% CI	p-value	X^2	p-value
Occupation								
Employed	185	160	25	4.42	2.26 – 8.62	0.0001*	19.74	0.00003**
Home maker	206	131	75					
Diet								
Vegetarian	28	22	6	1.26	0.50 – 3.17	0.62	0.25	0.62
Mixed	363	269	94					
Tobacco history								
Yes	81	65	16	1.88	1.03 – 3.45	0.04*	4.18	0.04*
No	310	226	84					
Family history								
Yes	30	20	10	0.67	0.30 – 1.49	0.32	0.99	0.32
No	361							
History of prior Irradiation								
Yes	8	5	3	0.57	0.13 – 2.50	0.46	0.54	0.46
No	383	286	97					
History of Native Treatment								
Yes	34	28	6	1.69	0.68 – 4.20	0.26	1.27	0.26
No	357	263	94					
Comorbidities								
Yes	133	98	35	0.54	0.33 – 0.89	0.01*	6.20	0.01*
No	258	193	65					
Menstrual history								
Post menopause	258	210	48	2.81	1.71 – 4.61	0.0001**	15.83	0.0001**
Pre menopause	133	81	52					

Rural patients were slightly more likely to report a history of tobacco use (OR=1.88, 95% CI: 1.03–3.45, p=0.04), a statistically significant finding that highlights the need for smoking cessation efforts, especially in rural areas. Tobacco uses as a risk factor for cervical cancer progression means targeted prevention counseling would be beneficial for these populations.

There was no significant statistical difference between groups concerning family history of cancer (OR=0.67, 95% CI: 0.30–1.49, p=0.32), but it was a bit higher among urban patients. This may suggest that genetic predisposition has little variation between groups; however, in general, a family history should still be taken into account while screening for cancer.

Prior irradiation history also showed no significant difference (p=0.46), which may indicate similar treatment access between rural and urban groups. Likewise, the history of native treatment use, which reflects traditional or alternative practices, did not vary significantly between groups (p=0.26). This indicates that rural and urban patients may use these approaches as complementary to conventional treatment.

Comorbidities were significantly more common in urban patients (OR=0.54, 95% CI: 0.33–0.89, p=0.01). This may reflect a greater prevalence of lifestyle-related health conditions, increased life expectancy, or better diagnostic practices in urban settings. Comorbidities complicate cancer treatment, often requiring additional medical attention and influencing treatment choices, making integrated, multidisciplinary care essential for urban patients.

Menstrual history was different with significant data showing that post-menopausal status is the most prevalent in rural patients (OR=2.81, 95% CI: 1.71–4.61, p=0.0001). This provides the basis to speculate about a more aged profile of the rural patient and how this aspect will greatly influence presentation and progression of the disease, forcing the recognition of the importance of age-sensitive strategies.

These results show complex sociodemographic and clinical differences between rural and urban cervical cancer patients. Key factors such as employment, tobacco use, comorbidities, and menstrual status vary significantly,

indicating a need for customized healthcare approaches that address each group's unique characteristics. Holistic care integrating these factors could enhance cervical cancer management, enabling more effective, patient-centered treatment.

4.2.6. Cervical cancer treatment regimens: patient distribution and characteristics

The treatment regimen was unavailable for 41.43% of the 391 patients initially enrolled, as these patients defaulted either after registration or following preliminary investigations. This high rate of default highlights significant challenges in the treatment plan for cervical cancer patients, which may be attributed to factors such as socioeconomic barriers, lack of awareness about the importance of ongoing treatment, or logistical difficulties in accessing healthcare services.

As shown in Table 11, a total of 229 cervical cancer patients eventually received various types of treatment that can be classified into two main categories: monotherapy and combined therapy. Such classification is very important for the understanding of treatment trends and to determine the most effective approaches in the management of cervical cancer.

The primary approach to monotherapy was chemotherapy, used in 38 patients, making up 16.6% of the treatment group. The most commonly used agents among this subgroup were cisplatin, which was given to 30 patients, because it has been well established as an effective agent in the treatment of cervical cancer. Other chemotherapy regimens included carboplatin in 1 patient and combinations of paclitaxel + carboplatin, which were given to 6 patients, and cisplatin + etoposide, which was administered to 1 patient. These choices represent the dependence on cisplatin as a keystone in the chemotherapy treatment of cervical cancer due to its well-established efficacy against this cancer.

For 6 patients, radiotherapy was applied, accounting for a minority of the treatment group. Here, conventional radiotherapy was applied as the sole modality, indicating a targeted application for certain patient scenarios in which radiation therapy would be appropriate, possibly in localized disease or in palliative care needs.

Table 11. Treatment regime

Type of treatment	Number of patients (N=229)
Mono therapy	**44 (19.1)**
Chemotherapy	**38 (7.8)**
Carboplatin	1
Cisplatin	30
Cisplatin + etaposide	1
Paclitaxel + carboplatin	6
Radiotherapy	**6 (2.6)**
Conventional	6
Combined therapy	**185 (80.9)**
Cisplatin + conventional	115
Cisplatin + Intensity-Modulated Radiation Therapy	29
Cisplatin + Extracorporeal radiation therapy	10
Cisplatin + 3D Conformal Radiation Therapy	2
Carboplatin + Conventional	16
Carboplatin + Intensity-Modulated Radiation Therapy	6
Gemcitabine + Conventional	1
Gemcitabine + Intensity-Modulated Radiation Therapy	1
Paclitaxel + carboplatin + Intensity-Modulated Radiation Therapy	2
Paclitaxel + carboplatin + 3D Conformal Radiation Therapy	1
Paclitaxel + carboplatin + Conventional	1
Paclitaxel + carboplatin + Extracorporeal radiation therapy	1

In this study of 229 cervical cancer patients, combined therapy was the predominant approach of 185 patients (80.9%). This preference for combined regimens aligns with a growing trend in advanced cervical cancer management that integrates multiple therapeutic modalities, which aim to leverage the

synergistic effects of these combinations, ultimately enhancing treatment efficacy and improving patient outcomes (Jiang *et al*., 2021).

Cisplatin-based regimens were the most common in our study, underscoring cisplatin's central role in multimodal treatment strategies. As a potent radiosensitizer (Brown *et al*., 2019), cisplatin is frequently paired with radiation to maximize tumor cell kill while reducing resistance risks associated with monotherapy. The most prevalent regimen paired cisplatin with conventional radiation, followed by combinations with sophisticated techniques including Intensity-Modulated Radiation Therapy (IMRT) and 3D Conformal Radiation Therapy (3D-CRT). These advanced modalities allow precise tumor targeting, limiting radiation exposure to surrounding healthy tissue and improving patients' tolerability and quality of life (Petrelli *et al*., 2014).

In our study, 115 patients received cisplatin with conventional radiation, and 29 received cisplatin with IMRT, highlighting cisplatin's established role in improving outcomes in advanced cervical cancer. Recent studies, including those by Zeng *et al*. (2024) and Li *et al*. (2023), underscore cisplatin's strong radiosensitizing properties that contribute to improved survival rates and tumor control. IMRT has gained popularity due to its precision in targeting tumors, minimizing side effects, and enhancing patient quality of life (Jin *et al*., 2023).

For patients contraindicated for cisplatin, mainly reported renal toxicity—carboplatin-based combinations were an alternative (16 patients with conventional radiation, 6 with IMRT). Although carboplatin is a viable alternative, it is less effective than cisplatin as a radiosensitizer (Sebastião *et al*., 2016). A small subset (10 patients) received cisplatin with extracorporeal radiation therapy (ERT), an option typically reserved for pelvic recurrence or prior radiation exposure.

In addition, some patients received combinations of paclitaxel and carboplatin with conventional radiation or IMRT, reflecting an ongoing exploration of multi-drug regimens for enhanced efficacy. Recent studies by Mabuchi *et al*. (2017) have demonstrated the potential of paclitaxel-carboplatin-radiation combinations to improve outcomes in advanced cervical cancer cases, further

underscoring the trend toward individualized treatment plans that optimize outcomes while minimizing toxicity.

This preference for combining chemotherapy with precision radiation aligns with evolving guidelines advocating for multimodal approaches, as they improve tumor control and minimize side effects (Chen *et al*., 2017). These trends are consistent with advances in cervical cancer treatment and reflect a broader shift toward personalized, tailored therapeutic strategies. The predominance of combination therapy highlights the complexity of cervical cancer treatment, where multidisciplinary approaches are essential to meet patients' diverse needs, incorporating chemotherapy, radiotherapy, and potentially targeted therapies or immunotherapies based on tumor characteristics and patient health.

The study's findings also reveal treatment adherence challenges and emphasize the importance of ongoing support and education to keep patients informed and engaged throughout their treatment journey. This strategic emphasis on combination therapy represents a response to the multifaceted demands of cervical cancer treatment, aiming to improve overall outcomes and quality of life for patients.

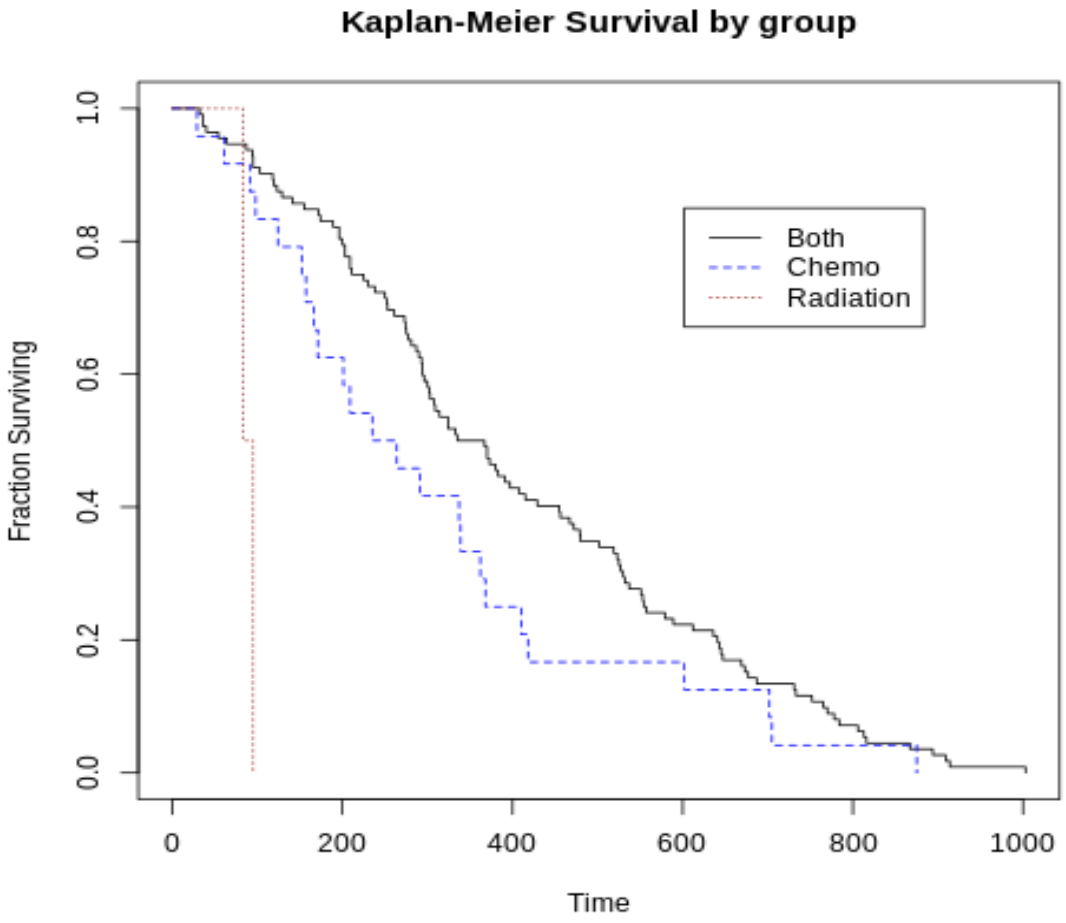
4.2.7. Survival analysis

Survival data was available for 138 out of the 229 patients who received treatment. Based on this data, a Kaplan-Meier plot was generated. The Kaplan-Meier plot reveals that patients receiving combination therapy have the highest survival probabilities over time. As shown in Figure 5, the survival curve for this group declines gradually, indicating longer survival times than the other groups.

The median follow-up time was 46.3 weeks (324) days. Patients who received only chemotherapy showed a moderate survival curve, with lower survival rates than the combination therapy group but higher than the radiation-only group. The radiation-only group exhibited the steepest drop in survival probability, especially in the early stages, suggesting that radiation alone may be the least effective option for long-term survival in this cohort.

This analysis concluded that combination therapy offers the most favorable survival outcomes, followed by chemotherapy alone, with radiation alone associated with the shortest survival times. This suggests that combination therapy might confer a survival advantage and, in certain groups, become a treatment of choice. These findings may help guide choices of treatment in different patients and indicate what outcome patients should expect with different treatments.

Figure 5: Kaplan-Meier survival analysis comparing treatment modalities in cervical cancer patients. (x-axis (time): the x-axis represents time in days, tracking the progression from 0 to 1000 days since the start of treatment. y-axis (fraction surviving): the y-axis displays the fraction of patients alive at each time point, with 1.0 representing 100% survival at the start and decreasing towards 0 as patients pass away).



Despite the valuable insights that this study has provided, several limitations need to be taken into account when interpreting the results. Survival data were available for only 138 out of 229 patients, meaning that a large proportion of the cohort did not contribute data for the survival analysis. This could potentially lead to bias if the missing data were not randomly distributed. Second, treatment adherence may have varied among patients, and inconsistencies in how strictly patients followed their prescribed treatment regimens could have affected survival outcomes, introducing variability in the results. Additionally, the follow-up period, while providing a median follow-up time of 46.3 weeks, may not be long enough to capture long-term survival trends, especially for patients who experienced delayed recurrence or death. The study failed to consider other confounding variables, including the age of the patient, comorbidities, or the stage of cancer, which can significantly impact survival. These variables may have influenced the survival analysis because they are known to influence the response to treatment and prognosis. Taking all these limitations into account, findings should be treated with caution. Further research needs to be carried out with an adequate sample size and longer follow-up to confirm these results.

Phase II results of our study reported that in the hospital-based cohort survey, occupation and other exposures significantly influenced the incidence of therapy adoption among urban women. The study also identified a prevalent unwillingness to opt for screening and vaccination within this population. These findings highlight the importance of targeted interventions for rural working women, improved access to healthcare services, and tailored treatment strategies based on histological diagnoses. There were more squamous types in the cohort. An analysis concluded further that combination treatment provides the survival benefit best with the results recommending it as one that may prove to offer better survival and may be considered by default for survival benefits in case extended survival may be required, and these should guide treatment as part of patient decisions on expected results.

Our analysis revealed that squamous cell cervical carcinoma accounted for the majority of cases in our study population. To investigate the genetic alterations underlying this subtype, we performed exome sequencing on a subset of five representative squamous cell cervical carcinoma samples with the following objectives:

1. To identify recurrent mutations and genetic variants.
2. To characterize the mutational landscape.
3. To uncover potential driver genes and pathways.

This targeted approach aims to enhance our understanding of squamous cell cervical cancer molecular mechanisms, ultimately informing diagnostic and therapeutic strategies.

Phase III

4.3. Mutational profiling of squamous cell cancer patients

Cervical cancer's genetic complexities, particularly in cases like squamous cell carcinoma (SCC), can be explored using whole exome sequencing (WES). This method focuses on the exome—the protein-coding regions of genes—where most disease-causing mutations are found. By examining these regions, WES helps to identify driver mutations, discover biomarkers, and offer potential insights for personalized treatment strategies.

In Tamil Nadu, the mutational profiles of cervical cancer-associated genes need to be better studied. In this study, we analysed representative samples from the squamous cell cervical cancer patients to map out these genetic changes.

The WES workflow includes four key steps:

1. **DNA extraction** from patient samples.
2. **Library preparation**, where the DNA is processed for sequencing.
3. **Sequencing** the exome, focusing on the protein coding regions.
4. **Data analysis** to identify genetic variants impacting protein function.

WES offers a cost-effective, clinically feasible approach to understanding cancer-related genetic changes, bringing it closer to routine use in patient diagnostics and personalized cancer care (Duppala *et al.*, 2024).

4.3.1. Patient profile: squamous cell cervical cancer

In our study, we collected biopsy samples from five patients diagnosed with squamous cell cervical cancer, represented as CC1 – CC5. Each patient had a different clinical profile, which clarified the variability of this cancer's presentation (see Table 12 for detailed characteristics). The ages of the patients ranged from 52 to 59 years, indicating that cervical cancer can affect women in their later years, a critical demographic feature for targeted screening and prevention efforts.

Patients were noted to have fair nutritional status, which is crucial for general and specific recovery in cancer patients. More importantly, none of these patients presented with signs of anemia or jaundice at the time of diagnosis. These could be complicating features in the patient's clinical picture and could, therefore, influence treatment decisions.

Blood types ranged from B +ve in CC1 to the rest, CC2-CC5, being O +ve. In this context, it may not directly affect their treatment but is significant in other clinical scenarios, such as transfusion requirements at surgery time or the potential for certain disease associations.

The symptoms reported by the patients were post-menopausal bleeding in CC1, CC4, and CC5. This symptom is very important since it is a common presenting symptom of cervical cancer and requires further evaluation in post-menopausal women. Abdominal pain was reported by CC2 and CC3, which might indicate a more advanced disease state or complications arising from tumor growth. Other symptoms included weight loss, white discharge, loss of appetite, nausea, vomiting, diarrhoea, pain, fatigue, and urinary incontinence, all of which can be common in cervical cancer patients (Andring *et al*., 2024). These symptoms highlight the multifaceted impact of the disease on patients' quality of life and the importance of thorough clinical evaluations.

Table 12. Clinical and demographic features of squamous cell cervical cancer patients

Parameters	CC1	CC2	CC3	CC4	CC5
Age	52	54	59	59	52
Histology type	SCC	SCC	SCC	SCC	SCC
Nutritional status	Fair	Fair	Fair	Fair	Fair
Anaemia	No	No	No	No	No
Jaundice	No	No	No	No	No
Blood type	B+ve	O+ve	O+ve	O+ve	O+ve
BMI	23.683	32.275	34.576	21.403	22.074
BP	110/70	120/80	110/70	110/70	110/70
Pulse	84	86	82	84	74
Presenting symptoms	Post-menopausal bleeding	Right side abdomen pain for 25 days, weight loss	Lower abdomen pain for 3 months. Post-menopausal bleeding with white discharge in the past 3 months, weight loss, loss of appetite	Urinary incontinence in the past one month	Post-menopausal bleeding with white discharge in the last 3 months.
History of smoking	No	No	No	No	No
History of alcohol	No	No	No	No	No
History of tobacco	No	No	No	No	No
Diet	Mixed	Mixed	Mixed	Mixed	Mixed
Comorbidities	Diabetes mellitus, hypertension	-	-	-	-

Interestingly, none of the patients had a history of smoking, alcohol, or tobacco consumption. This absence of common risk factors often associated with cervical cancer raises essential questions about other potential etiological factors, including genetic predispositions or the role of high-risk human papillomavirus (HPV) strains.

Dietary habits among the patients were mixed, suggesting a varied intake of food groups, which could influence their overall nutritional status and health outcomes. However, further details regarding the specific dietary components would be necessary to understand their potential impact on cancer progression or treatment response.

The only comorbidities identified in the cohort were diabetes mellitus and hypertension in CC1. The presence of these conditions may influence treatment decisions, as managing comorbidities alongside cancer therapy is crucial for optimizing patient care. Diabetes, for instance, can complicate wound healing and increase the risk of infections, while hypertension may necessitate careful monitoring during treatment, mainly if surgery or specific chemotherapeutic agents are involved (Shouman *et al*., 2024).

Overall, the diverse clinical profiles of these five patients diagnosed with squamous cell cervical cancer underscore the complexity of the disease and the necessity for personalized treatment approaches. Each patient's unique characteristics, symptoms, and comorbidities must be carefully considered when developing a comprehensive care plan to optimize outcomes and address each individual's specific needs.

4.3.2. DNA isolation, quantity, and quality check

DNA was successfully isolated from all five biopsy samples using the QIAamp DNA Mini Kit (CA, USA), a widely recognized and reliable method for extracting high-quality genomic DNA from various biological samples. The isolation process is crucial, as it ensures that sufficient and pure DNA is available for subsequent analyses, including genetic sequencing, or mutation detection that

may contribute to understanding the underlying mechanisms of squamous cell cervical cancer.

The extracted DNA from all five patient samples met the required criteria for quantity and purity, as demonstrated by the absorbance readings obtained during analysis. Table 13 depicts assessment of the quality of the isolated DNA across the cohort. This step is essential because contaminants can adversely affect downstream applications, leading to unreliable results.

Table 13. DNA sample quality and concentration

Sample Name	Purity A260/280	Concentration (ng/µl)
CC1	1.80	188.6
CC2	1.82	1788.8
CC3	1.83	484.5
CC4	1.85	247.8
CC5	1.83	125.3

The purity and concentration of DNA samples extracted from five patients diagnosed with squamous cell cervical cancer (designated as CC1 to CC5) was assessed to ensure their suitability for further molecular analyses. The A260/280 ratio, which serves as a reliable indicator of DNA purity, ranged from 1.80 to 1.85 across the samples. These values suggest that the DNA is high quality, with minimal protein contamination.

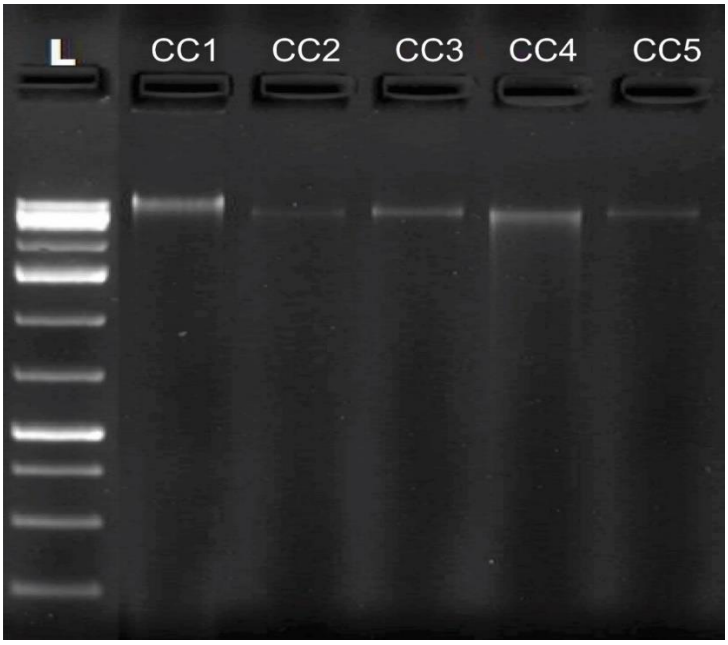
To further validate the quality of the quantified DNA, we performed electrophoresis on a 1% agarose gel, a standard technique used to assess DNA integrity and size. This method allows for visualization of DNA bands, providing insights into the presence of high molecular weight DNA and confirming the absence of degradation.

Image of the separated DNA bands, embedded below in Figure 6. The resulting gel image clearly represents the DNA quality, demonstrating distinct

bands corresponding to the DNA samples. The bands indicate that the DNA is intact and suitable for further applications.

Overall, the combination of A260/280 ratio analysis and agarose gel electrophoresis confirms the high quality of the DNA samples extracted from cervical cancer patients. These findings assure that the DNA can be confidently utilized in subsequent molecular analyses aimed at elucidating the molecular mechanism of squamous cell cervical cancer, potentially leading to advances in diagnosis, treatment strategies, and our overall understanding of the disease.

Figure 6. DNA fragmentation pattern



4.3.3. Library preparation

The table 14 presents quality control metrics for five squamous cell cervical cancer DNA libraries (CC1-CC5) prepared for next-generation sequencing (NGS).

Two key metrics were reported:

1. Concentration (ng/µl): The amount of library DNA present in each sample, measured in nanograms per microliter (ng/µl).
2. Insert Size: The average size of the DNA fragments (inserts) in each library, measured in base pairs.

Table 14. Library preparation quality control metrics

Sample Name	Concentration (ng/µl)	Insert size in bps
CC1	23.4	458
CC2	55.2	445
CC3	33.4	447
CC4	56.6	451
CC5	33.6	433

The concentrations of the DNA libraries prepared from the five squamous cell cervical cancer samples varied significantly, ranging from 23.4 ng/µl for sample CC1 to 56.6 ng/µl for sample CC4. These differences in concentration indicate the varying amounts of DNA material available for sequencing, which is an important consideration when planning sequencing runs. Higher concentrations of DNA can allow for more reads during sequencing, potentially enhancing the depth of coverage and increasing the likelihood of detecting rare variants (Sims *et al*., 2014).

In addition to concentration, the insert sizes of the DNA libraries were assessed, revealing a relatively consistent range across the samples. The insert sizes varied from 433 base pairs (bp) for sample CC5 to 458 bp for sample CC1, with an average insert size of approximately 447 bp. These insert sizes are significant because they can affect several aspects of the sequencing process.

Insert size is a crucial factor that influences sequencing efficiency and data quality. Libraries with larger insert sizes provide more context for assembling genomes or identifying structural variations, while smaller inserts limit the amount of contiguous information obtained. The relatively consistent insert sizes across the samples suggest that the library preparation process was uniform, which is favorable for obtaining reliable and comparable sequencing data (Yohe, and Thyagarajan, 2017).

The concentration and insert size metrics are integral for assessing library quality and determining the optimal conditions for sequencing. For instance, the

concentration of the DNA libraries directly impacts the sequencing depth—the higher the concentration, the greater the number of template molecules available during the sequencing process. This can lead to enhanced coverage of the targeted regions, facilitating more accurate variant calling and minimizing the risk of missing low-frequency variants that could be clinically relevant (Van Dijk *et al*., 2014).

4.3.4. Library quality check

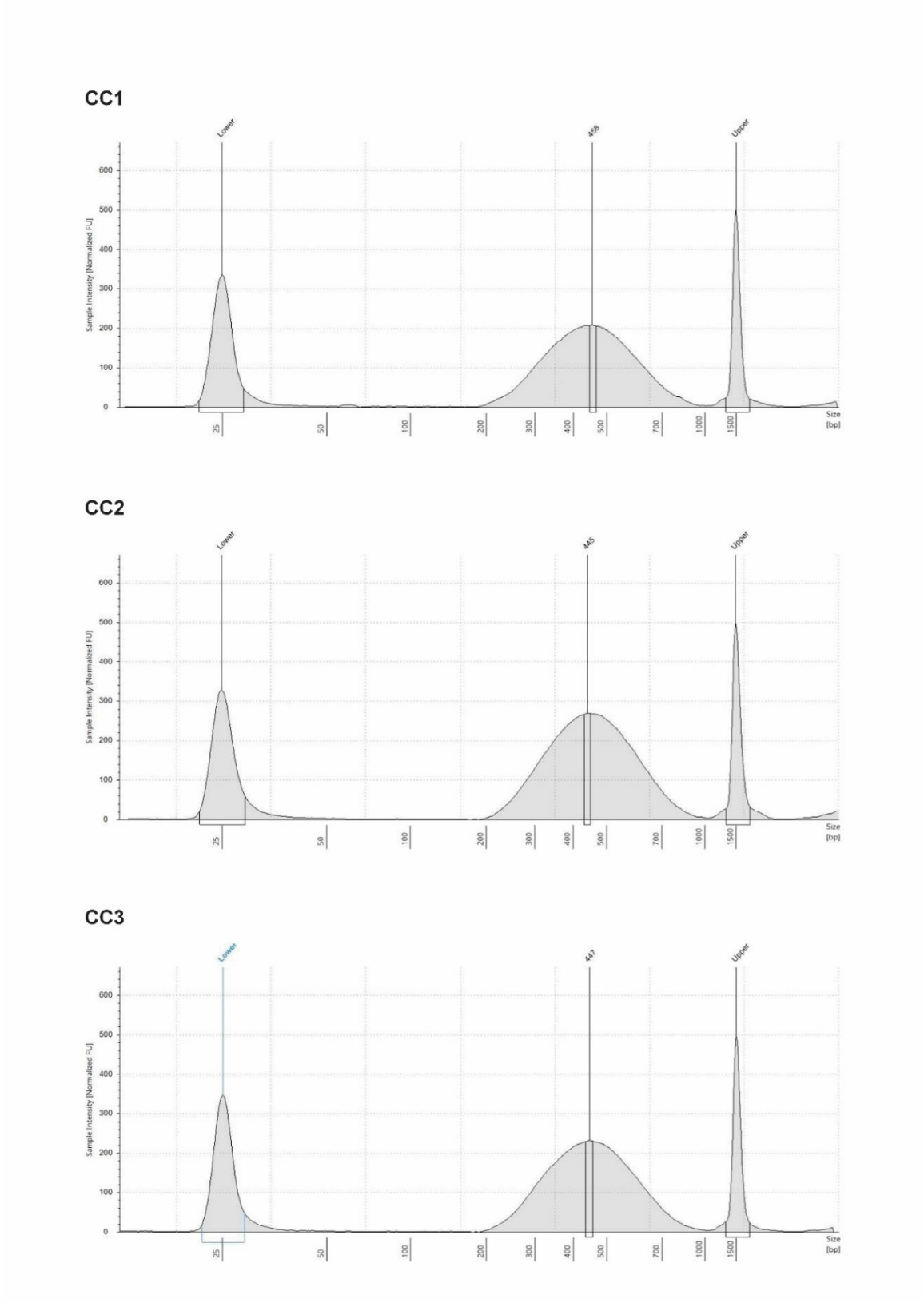
Electropherograms were essential for assessing the quality of DNA libraries before sequencing. This technique separated DNA fragments by size, generating a visual representation that showed the fragment size distribution. A successful library typically displayed a prominent peak within the desired range (300–600 bp) and included reference markers for the smallest (approximately 25 bp) and largest (around 1500 bp) fragments.

The electropherograms for the samples CC1, CC2, CC3, CC4, and CC5 provide a detailed overview of the size distribution of DNA fragments within each library (Figure 7). Each graph plots the size of the DNA fragments in base pairs (bp) along the x-axis and the sample intensity, showing the relative amount of fragments at each size on the y-axis.

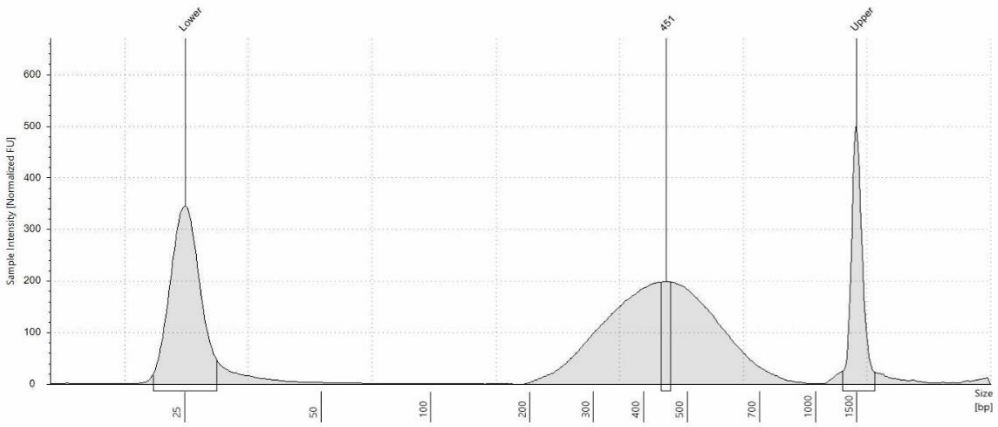
In sample CC1, a major peak appears at around 458 bp, which is the mean size of the library fragments and ensures that most of the sample is within the desired size range for sequencing. The lowest marker at about 25 bp indicates the smallest fragment, while the highest one at around 1500 bp is an indication of the maximum fragment size. This clear distribution suggests that the library preparation was successful.

Sample CC2 further supports this finding, with a major peak at 445 bp, indicating that most of the sample also falls within the desired sequencing size range, supported by the same reference markers.

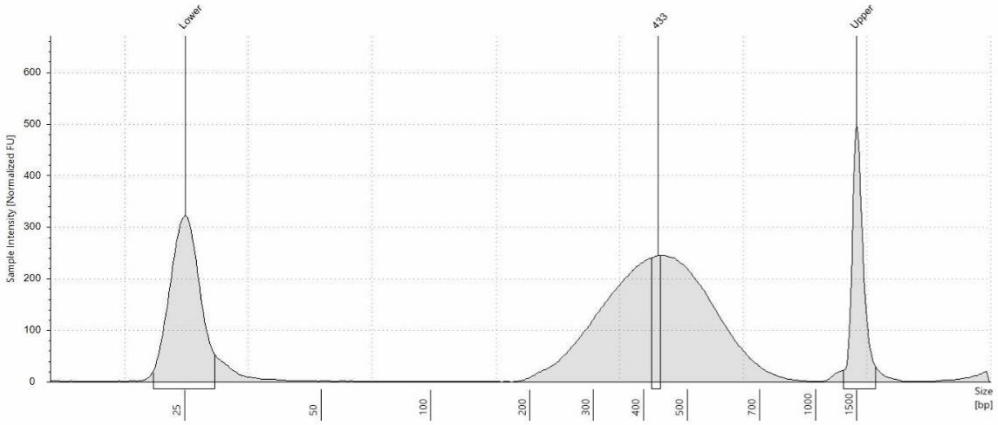
Figure 7. DNA fragment size distribution in library samples



CC4



CC5



The electropherogram of sample CC3 shows reference points at 25 bp and 1500 bp, with a main peak around 447 bp, which represents the target fragment size.

Similarly, in sample CC4, a peak around 451 bp is evident, confirming adequate library preparation within the intended size range, as indicated by the consistent reference markers.

Finally, the electropherogram for sample CC5 shows a major peak around about 433 bp, confirming once again that the library preparation was successful with DNA fragments properly sized for subsequent applications, for example, NGS. The frequent appearance of reference markers and large peaks in all samples, therefore, indicates that the DNA libraries were well prepared and appropriately sized.

The reported metrics indicate successful library preparation, with varying concentrations and consistent sample insert sizes. These results confirm the sequencing strategy and optimization, ensuring high-quality data for downstream analysis.

4.3.5. Base calling, quality control assessment, and alignment

Whole Exome Sequencing (WES) data analysis encompasses a multistep process: base calling, quality control, alignment, variant calling, annotation, and prioritization (Bao *et al*., 2014). Accurate base calling is crucial for efficient Next-Generation Sequencing (NGS)-based variant detection. Preprocessing involves adapter removal and trimming of low-quality bases, with optional removal of redundant reads and contaminants. Following quality control and preprocessing, sequence reads are aligned to a reference genome with high efficiency and accuracy (Seaby *et al*., 2016). This alignment enables reconstruction of a patient's exome data.

We conducted whole exome sequencing (WES) using the advanced Illumina NovaSeq 6000 platform, which is renowned for its capacity to generate high-throughput sequencing data with exceptional accuracy. This sequencing run produced approximately 63.7 million high-quality reads across all samples. Remarkably, 97.7% of these reads, amounting to 62.2 million, passed stringent quality control assessments (as shown in Table 15). This high percentage of reads meeting quality standards is crucial, as it indicates the reliability of the data for downstream analyses, including variant detection and functional studies.

Following the generation of these sequencing reads, the resulting exome sequencing data were aligned to the human reference genome (hg38) for each of the samples. Accurate alignment is essential for identifying variants, as it allows for the proper mapping of reads to specific genomic locations, facilitating the detection of single nucleotide variants (SNVs), insertions, deletions, and other genomic alterations associated with squamous cell cervical cancer (Bao *et al*., 2014).

The quality control report for the Next-Generation Sequencing (NGS) data from the five squamous cell cervical cancer samples (CC1-CC5) demonstrated

excellent data quality. This evaluation showed that there was a high percentage of sequencing reads that passed the quality filters, indicating effective sequencing and data processing protocols. The total number of sequencing reads produced for the samples was variable, with the lowest at 44.62 million for sample CC3 and the highest at 73.08 million for sample CC4. These variations in read numbers represent inherent differences in library quality, input DNA concentration, and sequencing efficiency among the samples.

Table 15. NGS data quality control

Sample ID	Total reads		Reads passed filters (%)
	Before filtering (Mb)	After filtering (Mb)	
CC1	62.161068	60.857428	97.90
CC2	70.797582	69.511192	98.18
CC3	44.616982	43.832664	98.24
CC4	73.082898	71.282384	97.53
CC5	67.946216	65.671788	96.65

A notable feature of this quality control test was the constantly high percentage reads passing filters with a range between 96.65% by CC5 up to 98.24% by CC3. High percentage values such as these reflect negligible data loss during the process of filtering which means that more than half the data generated remains usable for use in subsequent analysis. High-quality filtering is very important in NGS, as it helps remove low-quality reads, adapter contamination, and other artifacts that could compromise the integrity of the data (Li, 2014).

The depth, as demonstrated by the large number of reads, is quite deep and will therefore cover the exome comprehensively so that low-frequency variants are easily detected. It is crucial for cancer research where heterogeneity in tumor samples may present different variants at varying frequencies. The ability to capture these variants accurately enhances our understanding of the genetic

landscape of squamous cell cervical cancer, potentially informing personalized treatment strategies.

The data reduction during filtering was minimal, with less than 3.5% of data lost across all samples. Specifically, CC1 had 62.16 million total reads, with 97.90% passing filters, resulting in a data reduction of 1.30 Mb. Similarly, CC2, CC3, CC4, and CC5 showed comparable data reduction patterns. These results indicate that the sequencing data is of high quality, reliable, and suitable for downstream analysis, such as alignment, variant calling, and gene expression analysis.

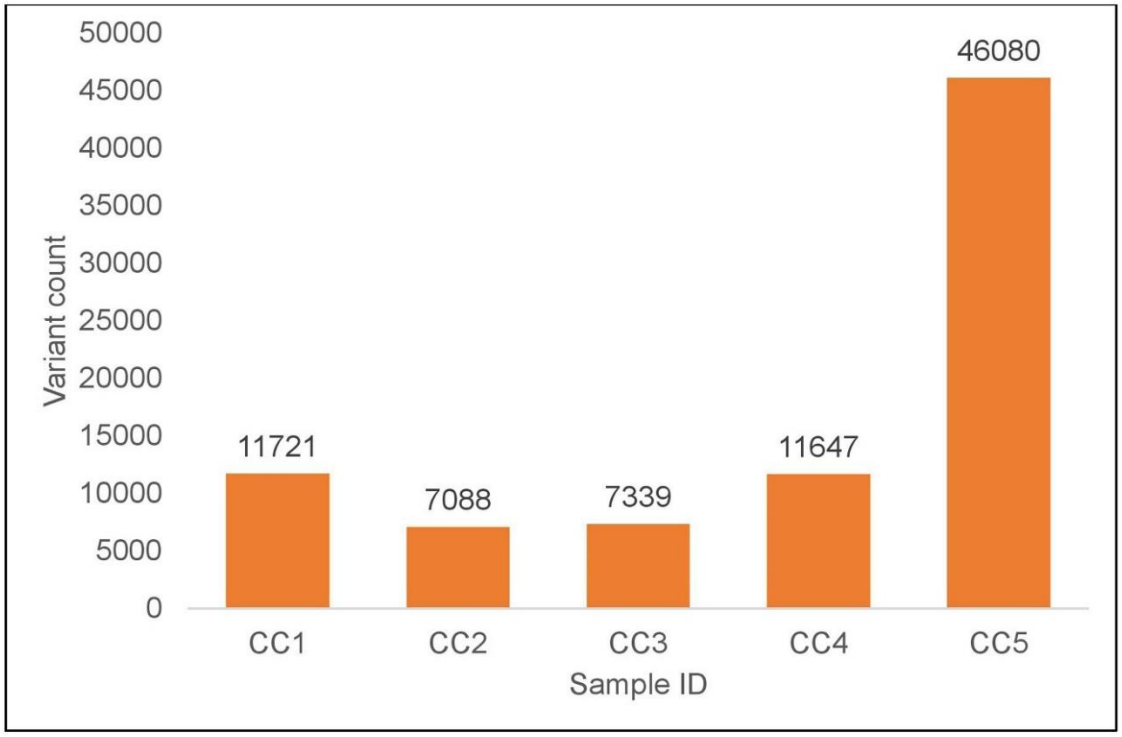
The high-quality data is crucial for squamous cell cervical cancer research, enabling accurate identification of genetic mutations, gene expression changes, and potential biomarkers. The excellent pass rates and minimal data loss demonstrate the efficiency of the sequencing and data processing protocols, ensuring reliable results for research and potential clinical applications. Furthermore, these findings provide a solid foundation for subsequent bioinformatics analysis, facilitating the discovery of novel insights into squamous cell cervical cancer biology and the development of personalized therapeutic strategies. Somatic genomic alterations drive the majority of human malignancies.

To elucidate the molecular mechanisms underlying squamous cell cervical cancer, we conducted a comprehensive somatic alteration study. Systematic characterization of somatic mutations in cancer genomes is crucial for understanding disease pathogenesis and developing targeted therapeutic strategies.

4.3.6 Profiles of somatic mutations in squamous cell cervical cancer samples

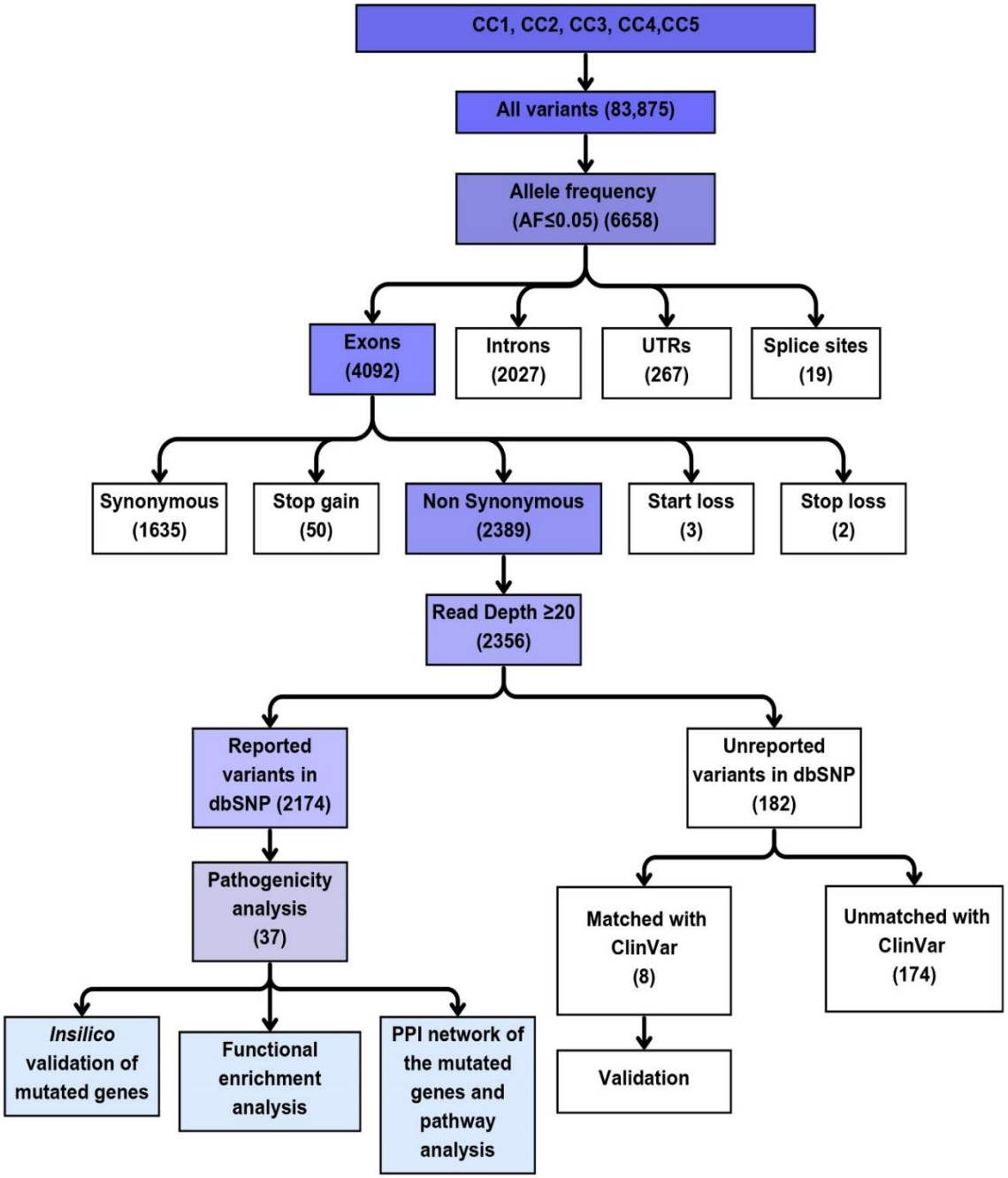
Somatic genomic alterations are a primary driver of human cancers, including squamous cell cervical cancer. To better understand the genetic landscape of squamous cell cervical cancer, we investigated somatic single-nucleotide variations (SNVs) in five squamous cell cervical cancer samples. The mutation frequencies for each sample were analysed, revealing variations in mutation rates (Figure 8).

Figure 8. Somatic gene alterations in the five squamous cell cervical cancer samples



As shown in Figure 9, a comprehensive mutational analysis was conducted on five samples of squamous cell cervical cancer (designated as CC1 to CC5), which resulted in the identification of a substantial number of variants—83,875 in total. This extensive dataset highlights the genetic complexity and heterogeneity often observed in cancer samples, as well as the potential for numerous alterations that may contribute to tumorigenesis (Supplementary File 1).

Figure 9. Variant analysis workflow: filtering, annotation, and functional insights



The flowchart illustrates the detailed steps of the filtering process, with the corresponding data for each step tabulated in Supplementary File 2. To refine our analysis and focus on high-confidence mutations that are more likely to have biological significance, we applied an allele frequency filter of less than 0.05. This stringent filtering process reduced the initial variant count from 83,875 to 6,658 variants. The rationale behind this filtering strategy is to concentrate on alleles with low population frequency, as these are more likely to represent rare, potentially pathogenic mutations rather than common polymorphisms that occur in the general population (Saint Pierre, and Génin, 2014). In cancer research, identifying these driver mutations is critical, as they are often implicated in tumor development, progression, and response to treatment.

Allele frequency is an essential metric in genetic analysis, defined as the proportion of a specific allele at a given locus in a population (Bodmer and Tomlinson, 2010). In the context of our study, understanding the allele frequency of mutations aids in distinguishing between benign variants and those that are likely to drive malignancy. By filtering out variants with higher frequencies, we can increase the likelihood of identifying mutations that play a direct role in the pathophysiology of squamous cell cervical cancer.

Once the variants were filtered based on allele frequency, we proceeded to categorize them by their functional regions within the genome. This categorization revealed a total of 4,092 exonic variants, which are alterations occurring within protein-coding regions of genes. Exonic mutations can have significant implications for protein function, potentially leading to the production of altered or dysfunctional proteins that contribute to cancer development (Ross *et al*., 2020).

In addition to exonic variants, we identified 2,027 intronic variants, which occur within non-coding regions of genes that can influence gene expression and splicing processes. While intronic mutations are often considered less impactful than exonic changes, they may still play a role in gene regulation and may contribute to cancer through mechanisms that alter mRNA splicing or gene expression levels (Diederichs *et al*., 2016).

We also identified 267 variants located in the untranslated regions (UTRs) of genes. UTR variants can affect mRNA stability, localization, and translational efficiency, thereby impacting protein production indirectly (Diederichs *et al*., 2016). 19 splice site variants were identified. These are of significant concern as they can lead to the aberrant splicing of pre-mRNA, where critical coding sequences may be either included or excluded, thus generating aberrant protein products.

Overall, this comprehensive analysis of the mutational landscape in squamous cell cervical cancer has provided a wealth of information regarding the types and locations of genetic alterations present in the samples. The identification and categorization of these variants not only enhance our understanding of the molecular mechanisms underlying cervical cancer but also pave the way for further investigations aimed at targeting these mutations for therapeutic interventions. This study, by focusing on high-confidence and low-frequency variants, makes a contribution to ongoing efforts to further detail the genetic basis of squamous cell cervical cancer.

We identified a total of 4,092 variants in the coding regions of genes in our comprehensive analysis of the mutational landscape in squamous cell cervical cancer. Of these, 2,389 were nonsynonymous variants, which are of special interest because they change the amino acids in the proteins encoded. These alterations can significantly affect protein function and are often implicated in various disease processes, including cancer (Tetreault *et al*., 2015). The classification of variants includes other types as well, such as synonymous mutations, stop gain (nonsense mutations), start loss, and stop loss mutations. Each of these categories has distinct implications for protein expression and functionality, but nonsynonymous mutations have garnered considerable attention in cancer research due to their potential roles as drivers of oncogenesis (Diederichs *et al*., 2016).

To better understand the impact of these nonsynonymous mutations, we undertook a detailed investigation into their functional characteristics, associated biological pathways, and processes. Nonsynonymous mutations are known to alter the amino acid sequence of proteins, possibly resulting in modifications to the

structure and function of proteins, which might interfere with regular cellular functions and promote carcinogenesis (Petrosino *et al*., 2021).

Applying a read-depth filter, the depth was refined to focus more on variants with depths greater than 20 reads; this filtered step, in turn improved the reliability of variant calls as our data were very high quality. From this refining step, there were 2,356 variants that met the read depth sufficiency. Using this variant to cross-reference, 2,174 variants matched and reported variants on the dbSNP database.

Subsequently, using several computational tools such as SIFT, Polyphen2v2, LRT, MutationTaster, FATHMM, Radial SVM, and LR we applied a pathogenicity analysis to determine the functional impact of these variants on genes. After elimination of variants tolerated, benign, probably damaging, or neutral by these tools, we selected 37 that were found damaging and deleterious for detailed functional analysis. For the remaining variants that were not classified as damaging or deleterious, the predictions were summarized and recorded in Supplementary File 2. This ensures all variants are included and provides some additional context to future studies.

To our knowledge, to date only few studies (Duppala *et al*., 2024; Kumari Konathala, *et al*., 2017; Das *et al*., 2016) on meta-analysis of genetic polymorphisms in cervical cancer patients were reported in India. Unlike these studies, our study was limited to studies on squamous cell cervical carcinoma with 5 representative samples. Our study on squamous cell cervical cancer was hindered by many factors including low awareness, preference of non-surgical treatment options, and approaching medical help during final stage which limited our sample size into just five cases.

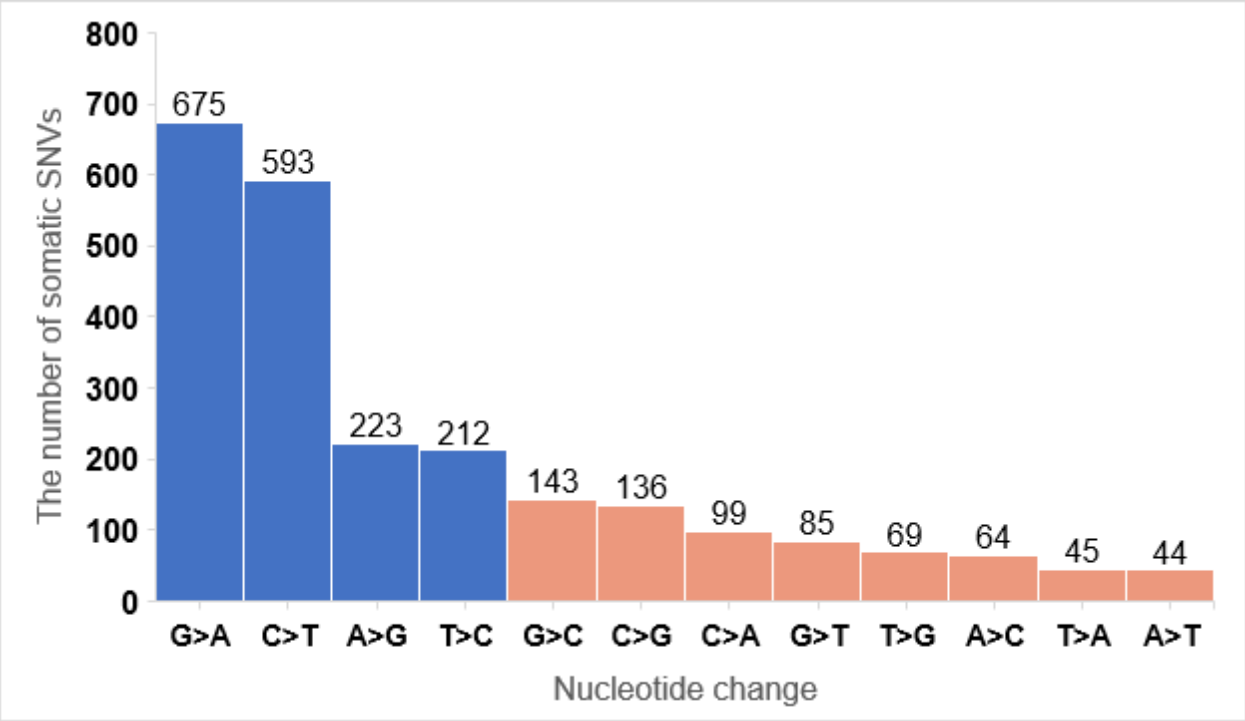
4.3.7. Comprehensive analysis of nonsynonymous variants in squamous cell cervical cancer

Nonsynonymous single nucleotide polymorphisms (nsSNPs) are genetic changes that alter the amino acid sequence of proteins. These changes can affect how proteins work, their stability, how they interact with other proteins, and the activity of key sites in enzymes and receptors. Studying nsSNPs is important because they can help us understand how certain mutations lead to diseases like cancer (Zhao *et al*., 2018; Stefl *et al*., 2013).

To gain deeper insights into the mutational landscape of cervical cancer, we performed a comprehensive mutation spectrum analysis, which highlighted distinct patterns of nucleotide transitions and transversions among the identified nsSNPs. As shown in Figure 10, the analysis revealed that among the mutations, G>A (675 occurrences) and C>T (593 occurrences) transitions were the most prevalent. These substitutions represent transition mutations, where one purine (G) is replaced by another purine (A), or one pyrimidine (C) is replaced by another pyrimidine (T). This high frequency aligns with the findings of Wu *et al*. (2024), which also reported C>T and G>A transitions as the most common in their study cohort, consistent with the present results.

Moderate frequencies of A>G (223 occurrences) and T>C (212 occurrences) transitions are also observed, reflecting additional mutational mechanisms potentially active in the studied biological context. In contrast, less frequent substitutions, such as A>T (44 occurrences) and T>A (45 occurrences), represent transversions, where a purine is replaced by a pyrimidine or vice versa. Transversions generally occur less often due to the higher molecular complexity required for such changes.

Figure 10. Frequency and distribution of nonsynonymous mutations in squamous cell cervical cancer. (blue colour denotes the transversion and orange colour denotes the transition.)



The prevalence of such types of mutations has provided insights into mechanisms driving mutagenesis in squamous cell cervical cancer, which could also be the result of environmental factors, like exposure to carcinogens or infections by viruses, specifically Human Papillomavirus (HPV), among others. The overall patterns of mutation spectrum observed here as part of this analysis contribute to more nuanced understanding of genetic alterations that characterize squamous cell cervical cancer. This would determine the types of nsSNPs and their relative frequencies, further helping in the identification of their specific roles in the tumorigenic process and how the disease evolves. Such insight not only into squamous cell cervical cancer biology but also, more importantly, into potential strategies for therapy focused on the downstream effects of the mutations could serve to improve both diagnosis and therapy for patients afflicted with this form of cancer.

4.3.8. Validation of nonsynonymous mutations through COSMIC database comparison

In our study, we considered only nonsynonymous mutations, which are known to change the amino acid sequence and may impact the function of proteins. Other types of mutations, such as synonymous mutations and intronic variants, were excluded from the analysis in order to obtain a precise evaluation of mutations with functional implications. Table 16 provides a comparison of the identified nonsynonymous mutations with data for squamous cell cervical cancer in the COSMIC (Catalogue of Somatic Mutations in Cancer) database, a widely accepted comprehensive resource for documenting somatic mutations in cancer. The purpose of the comparison was to validate and place our findings into a more relevant context.

All five samples in our study had nonsynonymous mutations for MUC12, amounting to a total of 20 mutations. The COSMIC database showed that out of 21 samples tested, 16 had nonsynonymous mutations, with a slightly higher total of 24 mutations. This would mean that our cohort had a 100% mutation rate compared to the 76% rate in COSMIC, thus possibly indicating that MUC12 mutations are more reliably observed in our focused cohort.

Similarly, for TTN, all five samples in our study were found to harbor nonsynonymous mutations, with 13 mutations identified. In the COSMIC database, nonsynonymous mutations were reported in 96 out of 130 samples, with a total of 165 mutations. Although the mutation rate in our dataset was higher (100% compared to 74% in COSMIC), the COSMIC data reflects a larger number of mutations, likely due to its broader sample size and genetic diversity.

For AHNAK, all five samples of our dataset were nonsynonymous mutations. Total mutations were found to be 7. According to COSMIC, the mutations were present in 18 of 23 samples, and the total mutations found were 20. This highlights a higher mutation rate in our study (100% vs. 78% in COSMIC), though the total mutation count was slightly higher in COSMIC.

For LRP5L, our analysis revealed nonsynonymous mutations in all five samples, resulting in a total of 6 mutations. By comparison, the COSMIC database

identified nonsynonymous mutations in 3 out of 5 samples, with a total of 3 mutations. This underscores a higher mutation burden for LRP5L in our study cohort.

Lastly, for OR2T34, the results from our study were identical to the COSMIC database, with all five samples in both datasets showing nonsynonymous mutations and a total of 5 mutations.

Table 16. Comparison of nonsynonymous mutation profiles between study data and COSMIC database

Gene	Non synonymous mutations in selected samples			Non synonymous mutations data from COSMIC database		
	Sample	Mutation	No of mutations	Sample tested	Mutated sample	No of mutation
MUC12	5	5	20	21	16	24
TTN			13	130	96	165
AHNAK			7	23	18	20
LRP5L			6	5	3	3
OR2T34			5	5	5	5

Our comparison of nonsynonymous mutations between the study data and the COSMIC database revealed consistent mutation patterns for several genes, including MUC12, TTN, AHNAK, and OR2T34, with both datasets showing high mutation rates in these genes. The high mutation rates in genes like MUC12 and TTN suggested their potential importance in squamous cell cervical cancer. These findings provide valuable insights into the genetic landscape of squamous cell cervical cancer and highlight key genes that may be critical in its molecular pathology.

4.3.9. Mutational analysis of nonsynonymous variants in squamous cell cervical cancer

Several bioinformatics tools were used to predict the potential impact of genetic variants on protein function in this study. SIFT evaluates the importance of each amino acid in a protein by comparing its conservation across different species. Variants that

affect highly conserved amino acids are more likely to be harmful. PolyPhen2 predicts the probable damaging effects of amino acid changes through consideration of 3D structure and evolutionary history of the protein. LRT allows for a comparison of the probabilities that a variant is harmful versus neutral, thus giving an idea about its potential effect on the protein. MutationTaster checks on sequence and structural information to give a prediction as to whether or not a variant may cause disease. FATHMM looks at the evolutionary context of the protein to predict whether a variant could impair protein function. RadialSVM is a machine learning approach to classify variants as harmful or neutral based on known data about genetic changes. Lastly, LR is a statistical model that considers multiple factors to predict whether a genetic change is likely to affect protein function.

Additional scores have also been applied for the assessment of the functional importance of variants. VEST3 scores predict the ability of the pattern of evolutionary conservation to make any variant potentially pathogenic. CADD phred calculates multiple annotations in scores for a particular variant that explains its ability to affect the function of the gene. GERP++_RS measures the conservation of a sequence to help identify evolutionary constraints, with higher scores indicating regions that are more likely to be important. PhyloP46way placental and phyloP100way vertebrate evaluate the evolutionary conservation of sequences across a range of species, with placental and vertebrate scores providing insights into the importance of the variant in these groups. SiPhy 29way logOdds provides a score based on the conservation of a variant across 29 species, offering a probability that the variant is deleterious. Together, these tools provided a comprehensive analysis of genetic variants, helping to identify those with the potential to contribute to squamous cell cervical cancer.

As shown in Table 17, In CC1, several deleterious variants were identified. MPEG1 was consistently predicted to be pathogenic across all tools, with a VEST3 score of 0.809 and a CADD Phred score of 13.78, indicating moderate functional impact. High conservation scores (GERP++_RS: 5.53; phyloP46way placental: 2.616) further highlight its evolutionary importance. Another gene, ABCC2, with a CADD Phred score of 33 and a VEST3 score of 0.96, suggests strong pathogenicity. Similarly, RDH12, associated with retinal dystrophy (Feathers *et al*.,

2019) (e.g., Leber congenital amaurosis), exhibited a highly deleterious CADD Phred score of 36. HADHA also showed a high CADD Phred score (35) and strong conservation (GERP++_RS: 5.43), implicating its role in mitochondrial fatty acid beta-oxidation disorders (Le-Tian *et al*., 2020). These findings suggest that the genes in CC1 are associated with metabolic, visual, and immune-related functions, with ABCC2 and RDH12 warranting prioritization for further research due to their established links to human diseases.

In CC2, notable genes include ATP2B2, with a CADD Phred score of 32 and a phyloP100way vertebrate score of 2.088, highlighting its critical role in calcium regulation. Variants in this gene could contribute to neurological and auditory disorders (Poggio *et al*., 2023; Minich *et al*., 2017). EHHADH demonstrated strong pathogenic potential, with a CADD Phred score of 21.9 and a GERP++_RS score of 5.91, highlighting its likely functional significance. This gene is associated with lipid metabolism disorders, which can lead to systemic clinical effects and emphasize its biological importance (Ren *et al*., 2021). Proper regulation of lipid metabolism is essential for cellular homeostasis, and disruptions in this pathway have been linked to various diseases, including cancer. Metabolic dysregulation is a well-established hallmark of cancer and plays a pivotal role in tumorigenesis (Hanahan and Weinberg, 2011). Alterations in EHHADH may drive metabolic reprogramming in cancer cells, potentially promoting tumor development and progression.

Another critical gene, CDH2, essential for cell adhesion and tissue integrity, had a CADD Phred score of 19.52 and high conservation (GERP++_RS: 6.16). While mutations in CDH2 are primarily linked to neural development and congenital heart conditions (László and Lele, 2022), its role in maintaining cell adhesion suggests a potential connection to cancer biology. Loss or alteration of cell adhesion molecules, including cadherins like CDH2, is a well-documented hallmark of cancer, contributing to tumor invasion, metastasis, and loss of tissue integrity.

Table 17. Assessment of genetic variants in cervical cancer using bioinformatics tools

Sample ID	Gene Name	SIFT	Polyphen2 v2	LRT	Mutation Taster	FATHMM	Radial SVM	LR	VEST3 score	CADD phred	GERP ++_RS	phyloP4 6 way placental	phyloP100 way vertebrate	SiPhy 29way logOdds
CC1	MPEG1	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.809	13.78	5.53	2.616	4.815	16.379
CC1	ABCC2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.96	33	5.18	1.582	9.855	14.697
CC1	RDH12	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.838	36	5.65	2.941	9.284	18.083
CC1	BCS1L	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.731	18.75	4.04	0.785	2.449	14.215
CC1	HADHA	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.31	35	5.43	2.557	7.76	17.823
CC1	MYOF	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.86	25.8	5.11	2.646	9.657	18.736
CC1	KCNJ1	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.904	24.4	5.71	2.69	7.805	19.866
CC1	ATP1A4	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.992	21.1	4.2	2.336	9.619	14.442
CC1	CP	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.365	26.4	5.46	2.573	6.237	17.497
CC2	ATP2B2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.439	32	5.6	2.627	2.088	19.619
CC2	EHHADH	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.844	21.9	5.91	2.793	9.322	20.299
CC2	GJC3	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.752	18.24	4.77	2.132	7.021	12.581
CC2	SLC10A5	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.453	21.8	5.51	2.759	5.75	17.267

Sample ID	Gene Name	SIFT	Polyphen2 v2	LRT	Mutation Taster	FATHMM	Radial SVM	LR	VEST3 score	CADD phred	GERP ++_RS	phyloP4 6 way placental	phyloP100 way vertebrate	SiPhy 29way logOdds
CC2	CCDC127	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.836	18.12	5.68	2.696	4.298	17.277
CC2	CDH2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.87	19.52	6.16	2.937	9.476	20.86
CC2	CGB1	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.599	11.11	1.6	1.215	3.211	9.226
CC3	CRYBG2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.245	24.4	4.99	2.47	2.434	16.416
CC3	CAPN9	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.84	13.86	4.41	1.249	5.776	14.159
CC3	CCT4	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.809	19.19	-0.098	0.057	0.648	4.166
CC3	TNS3	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.57	32	5.66	2.662	9.25	17.237
CC4	ACVR2B	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.48	20.1	5.01	2.598	6.576	6.643
CC4	CACNA1I	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.815	16.94	3.78	2.033	1.829	10.831
CC4	HEPHL1	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.397	23.6	5.95	2.279	8.741	16.424
CC4	RFC4	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.895	26.3	5.16	2.871	4.831	14.631
CC4	VIT	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.9	20.3	4.64	1.336	9.827	14.291
CC4	CACNB2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.658	24.1	5.88	2.78	9.837	20.219
CC5	CHD4	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.817	21.7	5	2.832	6.187	17.925

Sample ID	Gene Name	SIFT	Polyphen2 v2	LRT	Mutation Taster	FATHMM	Radial SVM	LR	VEST3 score	CADD phred	GERP ++_RS	phyloP4 6 way placental	phyloP100 way vertebrate	SiPhy 29way logOdds
CC5	ACSF3	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.375	20.7	3.67	0.956	8.848	12.153
CC5	CACNA1I	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.815	16.94	3.78	2.033	1.829	10.831
CC5	SLC4A3	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.969	28.3	4.75	2.471	9.505	17.954
CC5	GALC	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.855	17.44	3.62	2.326	3.615	6.789
CC5	RFC4	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.895	26.3	5.16	2.871	4.831	14.631
CC5	VIT	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.9	20.3	4.64	1.336	9.827	14.291
CC5	ARSJ	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.964	21.9	5.64	2.657	9.742	19.701
CC5	CLDN11	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.96	23.1	5.66	2.675	9.441	19.75
CC5	CACNB2	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.658	24.1	5.88	2.78	9.837	20.219
CC5	CGB1	Deleterious	Probably Damaging	Deleterious	Disease causing	Deleterious	Deleterious	Deleterious	0.599	11.11	1.6	1.215	3.211	9.226

CC3 harbors deleterious mutations in CRYBG2 and TNS3. CRYBG2 is primarily involved in lens structure and is only moderately conserved (GERP++_RS: 4.99), suggesting it may play a role in maintaining structural integrity and could be associated with diseases like cataracts. Although proteins like CRYBG2 are mainly linked to the eye, they likely contribute to more universal cellular functions, particularly in stress responses. Disruptions in these processes can affect cytoskeletal organization and cell adhesion, both of which are critical for cancer development and progression (Djomehri *et al*., 2020). Mutations in CRYBG2 may also influence cancer-related pathways, warranting further research to understand their role. TNS3, with a high phyloP100way vertebrate score of 9.25 and a CADD Phred score of 32, plays a crucial role in cytoskeletal and focal adhesion formation. Abnormal TNS3 function can impair cellular signalling and structural integrity, disrupting cytoskeletal organization (Mainsiouw *et al*., 2023). In cervical cancer, this may lead to alterations in cell adhesion properties, increasing motility and invasiveness—all features of cancer progression. Its role in focal adhesions also implicates TNS3 in tumor-related pathways, such as integrin signalling, which is often dysfunctional in cervical cancer.

In CC4, ACVR2B had a deleterious variant predicted to be pathogenic (CADD Phred: 20.1; GERP++_RS: 5.01), which underscores its critical function in the TGF-beta signalling pathway, which controls cell growth, differentiation, and apoptosis (Du *et al*., 2024). The dysregulation of this pathway is a hallmark of many cancers, including cervical cancer, where altered TGF-beta signalling can contribute to tumor growth, immune evasion, and metastasis. Mutations in ACVR2B could potentially interfere with this regulatory network to promote tumorigenesis and disease progression. HEPHL1, with predictions for pathogenicity (CADD Phred: 23.6; GERP++_RS: 5.95), is involved in iron metabolism and has a potential relationship to anemia. Iron metabolism has a very close relationship to the biology of cancer since disturbance in iron homeostasis can induce oxidative stress, DNA damage, and growth of tumors (Morales and Xue, 2021). Anemia is also another common comorbidity in cervical cancer patients (Hufnagel *et al*., 2021), that could be due to mutations like these. RFC4, a highly critical protein involved in the process of

DNA replication and repair (Morimoto *et al*., 2024), had a high CADD Phred score of 26.3, which essentially highlighted its association with genomic instability- one of the major events driving cancer development. Mutations within RFC4 could compromise DNA replication and repair processes, thereby increasing mutation rates and chromosomal anomalies that increase susceptibility to cancer. While ACVR2B is a central gene because of its role in TGF-beta signalling, HEPHL1 and RFC4 are also clinically relevant for their contributions to metabolic dysregulation and genomic instability, respectively, both of which are pivotal in cervical cancer progression.

In CC5, deleterious variants were found in CHD4, CACNB2, and SLC4A3, genes each of which plays an important role in cellular functions that are relevant to cancer. CHD4 (CADD Phred score 21.7) is involved in chromatin remodeling and gene regulation (Xia *et al*., 2017), and mutations in this gene can lead to cancer by disrupting normal cell division and gene expression. CACNB2 (CADD Phred score 24.1) regulates calcium channels, which are crucial for processes like cell growth and death. Abnormal calcium signalling can promote cancer by encouraging tumor growth and metastasis (James, and Brackenbury, 2022). SLC4A3 (CADD Phred score 28.3) helps regulate pH by transporting bicarbonate. Disruptions in pH regulation are common in cancer cells and can support tumor growth and invasion (White and Swietach, 2024). Together, these genes highlight chromatin remodeling, calcium signaling, and pH regulation as crucial areas of interest in the development of cancer and how their mutations could result in cancer development.

37 variants were consistently predicted to be deleterious across all five samples (Table 18), thus underlining their potential pathogenicity. Among the high-priority mutated genes, ABCC2, RDH12, ACVR2B, HEPHL1, and CHD4 stand out due to their strong associations with human diseases, high conservation scores, and functional significance. These results provide a foundation for understanding genotype-phenotype associations in squamous cell cervical cancer and the implications of these associations in tumorigenesis and disease progression

through pathways that disrupt protein stability, alter protein-protein interactions, and impair key signalling processes.

As shown in Table 18, variants are spread across multiple chromosomes, including regions on chr1, chr2, chr3, and others, with exact genomic positions and reference (Ref) and alternate (Alt) alleles given. Each mutation is accompanied by a given gene, revealing its possible biological significance. For instance, immune regulation, the visual process, and mitochondrial function all involve genes that include MPEG1 (chr11), RDH12 (chr14), BCS1L (chr2), and ATP1A4 (chr1), respectively. Indeed, calcium channel-related genes, such as CACNB2 (chr10) and CACNA1I (chr22), point to possible participation in neural or cardiac physiology. The sequencing depth ranges from low (e.g., 29 for CRYBG2 in CC3) to high (e.g., 368 for CGB1 in CC2), underscoring varying levels of confidence in variant detection.

The clinical significance annotations (CLNSIG) indicate that a majority of the variants are of uncertain significance, reflecting incomplete data on their pathogenicity. Some, like CP in CC1, appear to be benign/likely benign. Others, for example, GALC in CC5, have conflicting views on pathogenicity, so again functional validation and expert consensus are required. A few genes are recurrent among samples. Such recurrent ones include RFC4, CACNA1I, and VIT; shared genetic factors might be present, or the prevalence of the mutation may be at the population level. This study delivers important insights about the genetic variability found among the samples, most importantly on low frequency and presumably pathogenic variants that may implicate in phenotypes of this disease.

Table 18. Characterization of genetic variants in squamous cell cervical cancer: clinical significance and impact

S. No	Sample ID	Chr	Position	Ref	Alt	Gene Name	avsnp150	Read Depth	CLNSIG
1.	CC1	chr11	59211953	G	A	MPEG1	rs768506219	74	.
2.	CC1	chr10	99818821	G	A	ABCC2	rs536840524	79	VUS
3.	CC1	chr14	67725105	G	A	RDH12	rs745471670	82	VUS
4.	CC1	chr2	2.19E+08	C	T	BCS1L	rs748859459	85	VUS
5.	CC1	chr2	26201217	C	T	HADHA	rs758726265	91	.
6.	CC1	chr10	93389122	G	A	MYOF	rs761425525	94	.
7.	CC1	chr11	1.29E+08	C	T	KCNJ1	rs777296437	100	.
8.	CC1	chr1	1.6E+08	G	A	ATP1A4	rs138546868	131	.
9.	CC1	chr3	1.49E+08	G	A	CP	rs61733458	180	Benign / Likely benign
10.	CC2	chr3	10400988	C	T	ATP2B2	rs200591536	41	.
11.	CC2	chr3	1.85E+08	G	C	EHHADH	rs551915910	81	VUS
12.	CC2	chr7	99929052	A	T	GJC3	rs121908693	129	.
13.	CC2	chr8	81693891	G	A	SLC10A5	rs780908683	133	.

S. No	Sample ID	Chr	Position	Ref	Alt	Gene Name	avsnp150	Read Depth	CLNSIG
14.	CC2	chr5	205541	C	T	CCDC127	rs370948312	138	.
15.	CC2	chr18	27990212	G	C	CDH2	rs200059562	158	.
16.	CC2	chr19	49036146	C	T	CGB1	rs762432991	368	.
17.	CC3	chr1	26339303	C	T	CRYBG2	rs375184181	29	.
18.	CC3	chr1	2.31E+08	C	T	CAPN9	rs370370799	31	.
19.	CC3	chr2	61878947	C	A	CCT4	rs528619072	34	.
20.	CC3	chr7	47303112	G	A	TNS3	rs185351500	59	.
21.	CC4	chr3	38479792	C	T	ACVR2B	rs769170500	63	VUS
22.	CC4	chr22	39649833	C	T	CACNA1I	rs760155220	64	.
23.	CC4	chr11	94105997	A	G	HEPHL1	rs531038953	66	.
24.	CC4	chr3	1.87E+08	G	A	RFC4	rs574664855	67	.
25.	CC4	chr2	36808934	G	A	VIT	rs185199684	89	.
26.	CC4	chr10	18500855	G	C	CACNB2	rs772195020	129	.
27.	CC5	chr12	6582931	G	A	CHD4	rs771875799	33	.

S. No	Sample ID	Chr	Position	Ref	Alt	Gene Name	avsnp150	Read Depth	CLNSIG
28.	CC5	chr16	89145269	G	A	ACSF3	rs535695991	72	VUS
29.	CC5	chr22	39649833	C	T	CACNA1I	rs760155220	73	.
30.	CC5	chr2	2.2E+08	G	T	SLC4A3	rs34807189	76	.
31.	CC5	chr14	87965582	T	C	GALC	rs183105855	85	Conflicting interpretations of pathogenicity
32.	CC5	chr3	1.87E+08	G	A	RFC4	rs574664855	85	.
33.	CC5	chr2	36808934	G	A	VIT	rs185199684	88	.
34.	CC5	chr4	1.14E+08	G	T	ARSJ	rs549191458	88	.
35.	CC5	chr3	1.7E+08	G	A	CLDN11	rs559887015	89	.
36.	CC5	chr10	18500855	G	C	CACNB2	rs772195020	99	.
37.	CC5	chr19	49036146	C	T	CGB1	rs762432991	259	.

A whole exome sequencing study would, therefore, begin with the identification of a large number of genetic variants, but to derive meaningful biological insights, it is crucial to evaluate how these variants affect genes and their functions. Since genes are the functional units of the genome, they provide a direct link between genetic alterations and cellular pathways or biological processes that may contribute to disease.

The study shifts its focus to the mutated genes that might be contributing to the pathogenesis of diseases, rather than focusing on key players in mechanisms, which is individual variants. This approach facilitates a better integrated understanding of the disease and can map it to pathways and biological processes by uncovering molecular underpinnings of disease. The study identifies those most likely to drive the disease process by targeting genes harboring potentially damaging variants. Moreover, focusing on mutated genes allows for the identification of potential therapeutic targets or biomarkers because genes are more directly linked to cellular functions and can be modulated pharmacologically or genetically. Thus, studying mutated genes would mean a move from the description of raw data (variants) to the interpretation of biologically significant implications; the study will be more focused and relevant to understanding and dealing with the disease.

4.3.10. Functional Insights of mutated genes

Our bioinformatics analysis, we had discovered mutant genes; some may carry potentially risky or damaging variants. We subsequently analyzed the differential expression for assessing changes in the expression of these mutant genes across the various experimental conditions tested. Through the present study, we could conclude if these gene mutations were correlated with significant alterations in gene expression and therefore served to provide some insights into possible functional implications in disease.

37 of these variants represent 31 genes, where all underwent comprehensive analysis with functional enrichment analysis, pathway analysis and the construction of the protein-protein interaction (PPI) network.

Table 19. Gene ontology enrichment analysis of squamous cell cervical cancer-linked mutated genes.

Term	Genes	Count	P-Value	Benjamini value
Biological process				
transmembrane transport	SLC10A5, GJC3, ABCC2, SLC4A3, MPEG1	5	4.40E-04	9.60E-02
transport across blood-brain barrier	ABCC2, ATP1A4, SLC4A3	3	6.70E-03	7.40E-01
copper ion transport	HEPHL1, CP	2	1.90E-02	1.00E+00
iron ion transport	HEPHL1, CP	2	3.90E-02	1.00E+00
visual perception	CACNB2, CRYBG2, RDH12	3	4.00E-02	1.00E+00
Cellular components				
cell projection	ATP1A4, TNS3, CCT4	3	3.10E-02	1.00E+00
voltage-gated calcium channel complex	CACNB2, CACNA1I	2	3.80E-02	1.00E+00
presynaptic active zone membrane	CDH2, ATP2B2	2	5.00E-02	1.00E+00
endoplasmic reticulum lumen	CDH2, ARSJ, CP	3	7.30E-02	1.00E+00
plasma membrane	ABCC2, MYOF, ATP1A4, ATP2B2, SLC4A3, CP, ACVR2B, CLDN11, CACNB2, CACNA1I, CDH2, HEPHL1, KCNJ1	13	8.00E-02	1.00E+00

Term	Genes	Count	P-Value	Benjamini value
Molecular function				
ATP hydrolysis activity	RFC4, ABCC2, ATP1A4, CHD4, ATP2B2, BCS1L, CCT4	7	4.70E-05	5.20E-03
ATP binding	RFC4, ABCC2, ATP1A4, CHD4, ATP2B2, BCS1L, ACSF3, ACVR2B, KCNJ1, CCT4	10	3.90E-04	2.20E-02
long-chain-3-hydroxyacyl-CoA dehydrogenase activity	HADHA, EHHADH	2	3.10E-03	1.20E-01
3-hydroxyacyl-CoA dehydrogenase activity	HADHA, EHHADH	2	9.30E-03	2.00E-01
enoyl-CoA hydratase activity	HADHA, EHHADH	2	1.10E-02	2.00E-01

To further support our results, we validated some hub genes, analysed additional cancer genomic data to ensure deeper insights into the relevance of the variant in cancer. Table 19 tabulates the functional enrichment analysis results for further ease in interpretation. It does bring to the surface profound biological processes, cellular components, and molecular functions that get disrupted in squamous cell cervical cancer through this all-encompassing gene enrichment analysis.

a. Biological process

Transmembrane transport revealed significant enrichment among other biological processes and was characterized by a number of genes including SLC10A5, GJC3, ABCC2, SLC4A3, and MPEG1 (P = 4.40E-04). This process plays an important role in the process governing the passage of ions and molecules inside and outside the cell, maintaining homeostasis and supporting various cellular functions (Johnson and Langton, 2023). Transmembrane transport has been shown to be disrupted in various diseases involving impairments in ion balance and molecular trafficking, such as cancer and metabolic disorders.

Transport across the blood-brain barrier, facilitated by ABCC2, ATP1A4, and SLC4A3, is another key process (P = 6.70E-03). This is extremely important in regulating the entry and exit of substances between the bloodstream and the brain, providing the brain with a protective environment. Changes in that pathway may cause neurotoxicity and impaired drug delivery through the brain, which indirectly would lead to other systemic pathologies and those related to cancer progression (Achar *et al*., 2021). Also, in cervical cancer, the transport pathway may influence drug delivery and progression of disease as the body responds to treatment. Exploring these pathways may even help identify better methods of delivering treatment and improve outcomes from the patient's standpoint; just like research on transport systems is advancing to help assist therapies for brain metastases from breast cancer (Terceiro *et al*., 2023).

Also, copper transport of ions via HEPHL1 and CP, and iron transport of ions with the same genes were additionally confirmed as enacting a biological

process with a level of enrichment (P = 1.90E-02 and P = 3.90E-02). Copper and iron are essential trace elements that modulate multiple enzymatic functions, oxidative stress, and metabolism (Lelièvre *et al*., 2020; Salnikow, 2021). Transport dysregulation can cause oxidative damage, alter immune responses, and affect mitochondrial function, leading to different cancers or other chronic diseases.

Lastly, the enrichment of the genes related to visual perception (P = 4.00E-02), including CACNB2, CRYBG2, and RDH12, demonstrates the disturbances of processes concerning the retina and light signal processing. Although this may seem unrelated to cervical cancer, the genes involved, particularly CACNB2, are also implicated in calcium ion signalling and neurological function. This suggests possible overlapping molecular mechanisms in which disruptions in calcium channels and signalling could foster oncogenic processes.

These findings collectively point to the involvement of genetic variants in pathways that regulate ion transport, transmembrane signalling, and cellular homeostasis, shedding light on their potential contributions to disease progression.

b. Cellular components

Functional enrichment analysis pointed out significant cellular components that are associated with the genetic variants identified that can inform about the subcellular locations and structures that might have important roles in diseases. Participation of genes, like ATP1A4, TNS3, and CCT4, was noted in the development of cell projections, P = 3.10E-02, meaning their roles may involve cellular extensions, such as cilia, filopodia, and lamellipodia. These structures are integral to the process of cell migration, adhesion, and occur chiefly during/in intracellular signal transduction. Dysregulation in the formation or functioning of cell projections may cause the progression of cancer by allowing cancer cells to invade surrounding tissues and metastasize.

The voltage-gated calcium channel complex category has been obtained by two candidate genes, CACNB2 and CACNA1I (P = 3.80E-02). They are linked to cell signalling, muscular contraction, neuronal activity, and calcium ion homeostasis (Potier-Cartereau *et al*., 2020). Their involvement to the development

of squamous cell cervical cancer is increased when calcium channel function is altered, leading to abnormal signalling pathway control over cell proliferation, proliferation, and metastasis during cancer.

Enrichment of CDH2 and ATP2B2 in the presynaptic active zone membrane category (p=5.00E-02) suggests involvement in synaptic transmission and neuron communication. These genes are vital for vesicle docking and neurotransmitter release (Emperador-Melero and Kaeser 2020). Though these genes are predominantly associated with the nervous system, either a mutation or altered expression of such genes might interfere with cellular communication and adhesion mechanisms in various other tissue types that could positively or negatively affect cancer development.

The relevant genes, which most expressed are CDH2, ARSJ, and CP, are primarily involved with the endoplasmic reticulum (ER) lumen (P = 7.30E-02)-essentially a folding and modification site for proteins transport. The malfunctioning ER-associated processes could culminate in the accumulation of misfolded protein, which prompts ER stress and unfolded protein responses (Sun *et al*., 2021). The latter is mostly taken advantage of by cancer cells to withstand hostile conditions, such as hypoxia or starvation.

A total of 13 genes, namely ABCC2, MYOF, ATP1A4, ATP2B2, SLC4A3, CP, ACVR2B, CLDN11, CACNB2, CACNA1I, CDH2, HEPHL1, and KCNJ1, which are widely known to be associated with the plasma membrane, have been found to be enriched for their association with the endoplasmic reticulum (P = 8.00E-02). The plasma membrane functions as the main interface between the cell and external environment regarding signal transduction, ion transport, and cell adhesion processes (Sun *et al*., 2021; Firdous *et al*., 2023). The mutation of plasma membrane-associated genes can impair normal cell signalling and adhesion, subsequently causing unrestricted proliferation and escape from apoptosis and metastasis-an essential aspect of cancer.

The findings clearly express that, most of the affected cellular components, including calcium channels, cell projections, and plasma membranes, are said to

amplify the activity of dysfunctional cell communication and signalling pathways that uphold the survival of squamous cell cervical carcinoma, besides being the basis of normal cellular operations, and engagement on these will provide intervention targets against the progression of cancer.

c. Molecular function

The molecular function analysis offers a brief insight into the biochemical activities related to the identified genetic variants, which thereafter predict potential molecular mechanisms underlying squamous cell cervical cancer. The following molecular functions have been enriched. Seven genes (RFC4, ABCC2, ATP1A4, CHD4, ATP2B2, BCS1L, CCT4) were identified as involved in ATP hydrolysis, (P = 4.70E-05). This activity refers to the digestion of ATP to ADP and inorganic phosphate, releasing vast amounts of energy required to power numerous cell processes (Seyfried *et al*., 2020). These include, but are not limited to, DNA replication and repair (e.g., RFC4), ion transport (e.g., ABCC2, ATP1A4), and chromatin remodeling (e.g., CHD4). This is one target of ATP that can dysregulate ATP hydrolysis and provide an energy deprivation pathway in most cancer hallmarks like genomic instability and signalling alteration and resistance to apoptosis.

Ten genes (RFC4, ABCC2, ATP1A4, CHD4, ATP2B2, BCS1L, ACSF3, ACVR2B, KCNJ1, CCT4) are associated with ATP binding (P=3.90E-04). This activity initiated by ATP binding is one of the key elements of energy transfer, signalling, and metabolism, supporting enzymes and any protein in their efficient function by virtue of being ATPases, kinases, and/or transporters (Tang *et al*., 2022).

The HADHA and EHHADH genes are themselves enriched for long-chain-3-hydroxyacyl-CoA dehydrogenase activity, which is critical to mitochondrial fatty acid β-oxidation, a metabolic pathway that degrades fatty acids to generate energy (Wang *et al*., 2022). Impairments within this process are likely to lead to metabolic dysregulation, which is becoming a hallmark of practically all cancers. In turn, this has the potential for sustaining cancer cell proliferation in the energy-deprived conditions of tumors and for inhibiting apoptosis under these conditions.

HADHA and EHHADH relate to 3-hydroxyacyl-CoA dehydrogenase activity (P = 9.30E-03), yet again, a step in fatty acid β-oxidation. This activity is crucial for the degradation of medium- and long-chain fatty acids (Fang *et al*., 2022). Disruptions in the pathway may create energy imbalances and re-program metabolism in cancer cells, allowing those cells to rapidly proliferate and survive in the tumor microenvironment.

HADHA and EHHADH also show enoyl-CoA hydratase activity (P = 1.10E-02), an enzymatically important step of fatty acid β-oxidation. This involves the hydration of enoyl-CoA intermediates into 3-hydroxyacyl-CoA. Any defect in enzyme function may lead to the disturbance of ATP production and lipid metabolism, thus contributing to metabolic disorders that drive the growth of cancer cells via energy shifts. The present analysis on molecular function indicates a central role of ATP-dependent processes and lipid metabolism in squamous cell cervical cancer.

Genes involved in ATP hydrolysis and binding regulate fundamental cellular processes such as DNA repair, ion transport, and chromatin remodelling, all of which are crucial in cancer development. Meanwhile, genes involved in fatty acid β-oxidation highlight the importance of metabolic reprogramming in cancer cells. Dysregulation of these pathways can enhance tumor progression by providing cancer cells with the energy and metabolic flexibility required for survival and growth.

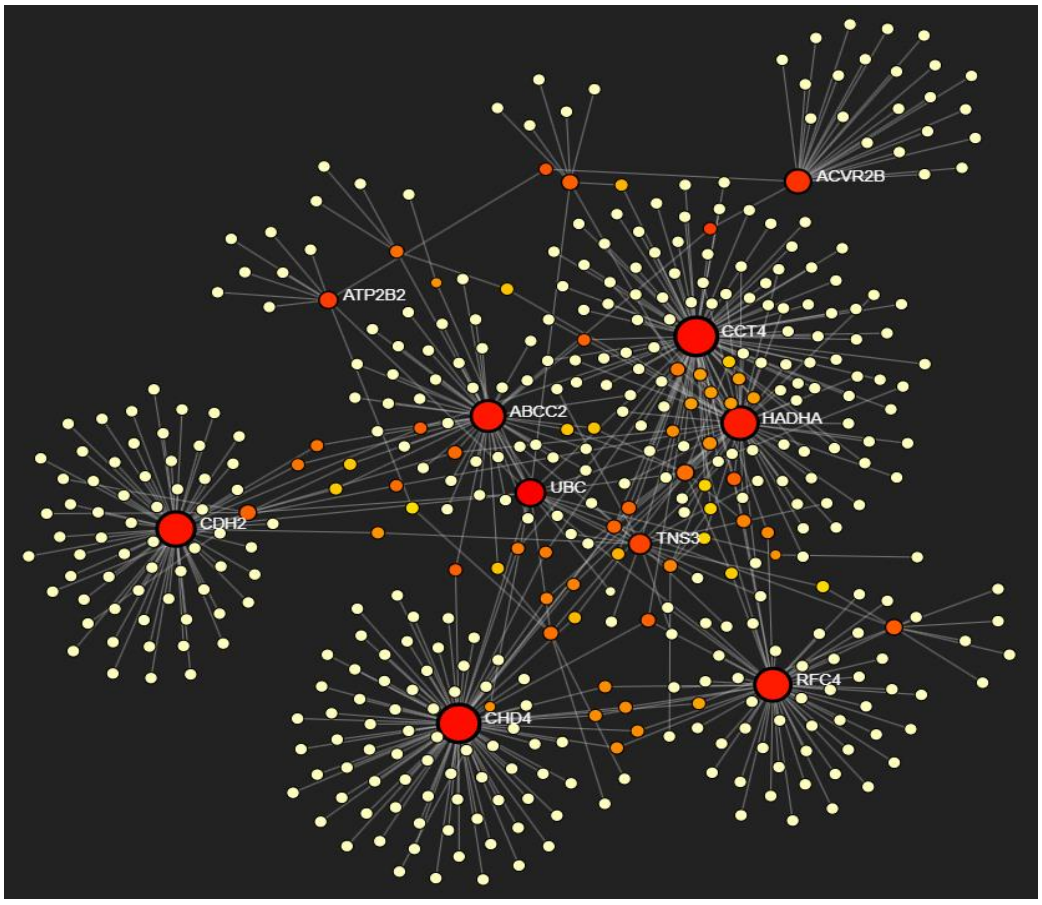
These findings suggest potential therapeutic targets, including enzymes involved in ATP utilization and lipid metabolism, for future cancer treatment strategies.

4.3.11. Construction of Protein-protein interaction network for the mutated genes and pathway analysis

As shown in Figure 11, a protein-protein interaction (PPI) network was constructed from mutated genes identified in squamous cell cervical cancer samples. The network is made of 452 nodes that represent the proteins encoded by the mutated genes and their interacting genes. These proteins interact through 509 edges, representing physical or functional connections. Such interactions

consider the relationships among proteins that may participate in shared biological pathways or processes affected by mutations. The network was generated using 20 mutated genes as seed proteins out of 31 genes, initially known to be mutated in the corresponding cancer samples and thus likely considered biologically relevant for the disease. Adding these seed proteins provided a basis for growing the subnetwork to encompass their direct and indirect interactions with other proteins to help uncover more general functional relationships. This process allows for identifying more proteins that can contribute to the disease mechanisms by interacting with seed proteins.

Figure 11. Protein-Protein Interaction network of mutated genes



It allows researchers to derive functional insight of the mutations discovered in squamous cell cervical cancer by analyzing this PPI network. The network's key proteins with many connections may be vital in tumorigenesis or cancer progression. This analysis provides a useful means of prioritizing genes and

proteins for further studies, ultimately facilitating the identification of potential therapeutic targets or biomarkers in squamous cell cervical cancer.

The analysis highlights the significance of network topology in understanding gene interactions and identifies key genes that may serve as potential targets for therapeutic intervention. Furthermore, this study demonstrates the power of PPI network analysis in unravelling the complex biological mechanisms underlying human diseases, providing a valuable resource for future research endeavours.

The integration of the mutated genes and their interacting partners with the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways revealed strong enrichment in different biological processes, providing deep insight into the intricate interactions underlying the mutated genes. As illustrated in Table 20, this systems-level analysis demonstrated that the mutant genes and their interactors converged to critical pathways critical for cellular functions, which included Hippo signalling pathway, TGF-beta signalling pathway, Adherens junction, Hepatitis C and Tight junction.

Table 20. KEGG pathway enrichment of mutated genes

Pathways name	Hits	P value	Adj P value
Hippo signalling pathway	26/154	1.45E-11	4.60E-09
TGF-beta signalling pathway	19/92	2.54E-10	4.03E-08
Adherens junction	14/72	1.38E-07	1.46E-05
Hepatitis C	20/155	3.76E-07	2.99E-05
Tight junction	20/170	1.67E-06	0.000106

The analysis of pathway enrichment highlights several key molecular mechanisms potentially disrupted by mutated genes identified in squamous cell cervical cancer. Among these, the Hippo signalling pathway showed the highest level of significance, with 26 out of 154 genes affected (P-value: 1.45E-11; Adjusted P-value: 4.60E-09). This pathway is essential for regulating organ size,

tissue homeostasis, and cell proliferation (Fu *et al*., 2022). Mutations in its components may impair its ability to control cell growth and apoptosis, leading to unregulated proliferation and tumorigenesis, making it a critical area for further exploration in cervical cancer.

The TGF-beta signalling pathway, with 19 of 92 genes affected (P-value: 2.54E-10; Adjusted P-value: 4.03E-08), is another significantly enriched pathway. It plays a pivotal role in regulating cell growth, differentiation, and immune system modulation. Disruptions in this pathway are often associated with immune evasion, epithelial-to-mesenchymal transition (EMT), and increased metastatic potential, highlighting its involvement in the progression and aggressiveness of cervical cancer (Knobloch *et al*., 2019).

The Adherens junction pathway was also significantly impacted, with 14 of 72 genes mutated (P-value: 1.38E-07; Adjusted P-value: 1.46E-05). Adherens junctions are critical for maintaining cell-cell adhesion and tissue integrity (Lessey *et al*., 2022). Mutations in genes within this pathway could compromise cell adhesion, promote cellular detachment, and facilitate invasive behaviour, contributing to cancer progression.

Additionally, the Hepatitis C pathway and Tight junction pathway were enriched, with 20 of 155 genes (P-value: 3.76E-07; Adjusted P-value: 2.99E-05) and 20 of 170 genes (P-value: 1.67E-06; Adjusted P-value: 1.06E-04) affected, respectively. While the Hepatitis C pathway's enrichment suggests possible overlap with viral infection mechanisms and chronic inflammation (Chaudhari *et al*., 2021), the Tight junction pathway highlights disruptions in epithelial barrier functions, which may facilitate tumor invasion and metastasis (Bhat *et al*., 2019). These findings collectively provide insights into the molecular basis of cervical cancer, offering potential avenues for targeted therapeutic strategies.

4.3.12. Hub gene verification through UALCAN

In this study, we focused on squamous cell carcinoma of the cervix, specifically from patients aged between 50 and 55 years. Due to the unavailability of normal sample sequencing data from our own collection, we opted to use the

normal data from the UALCAN database for comparison. We compared the statistical significance of gene expression across three conditions: Normal vs. Primary tumor, Squamous-cell vs. Normal, and Normal vs. Age (41-60 years). The statistical significance was assessed for various mutated genes, and the results provided insights into the differential expression of genes in different comparisons (Figure 12, 13, and 14). This approach helped establish a broader context by including external data, which aided in validating the findings and ensuring comprehensive analysis of gene alterations in cervical cancer.

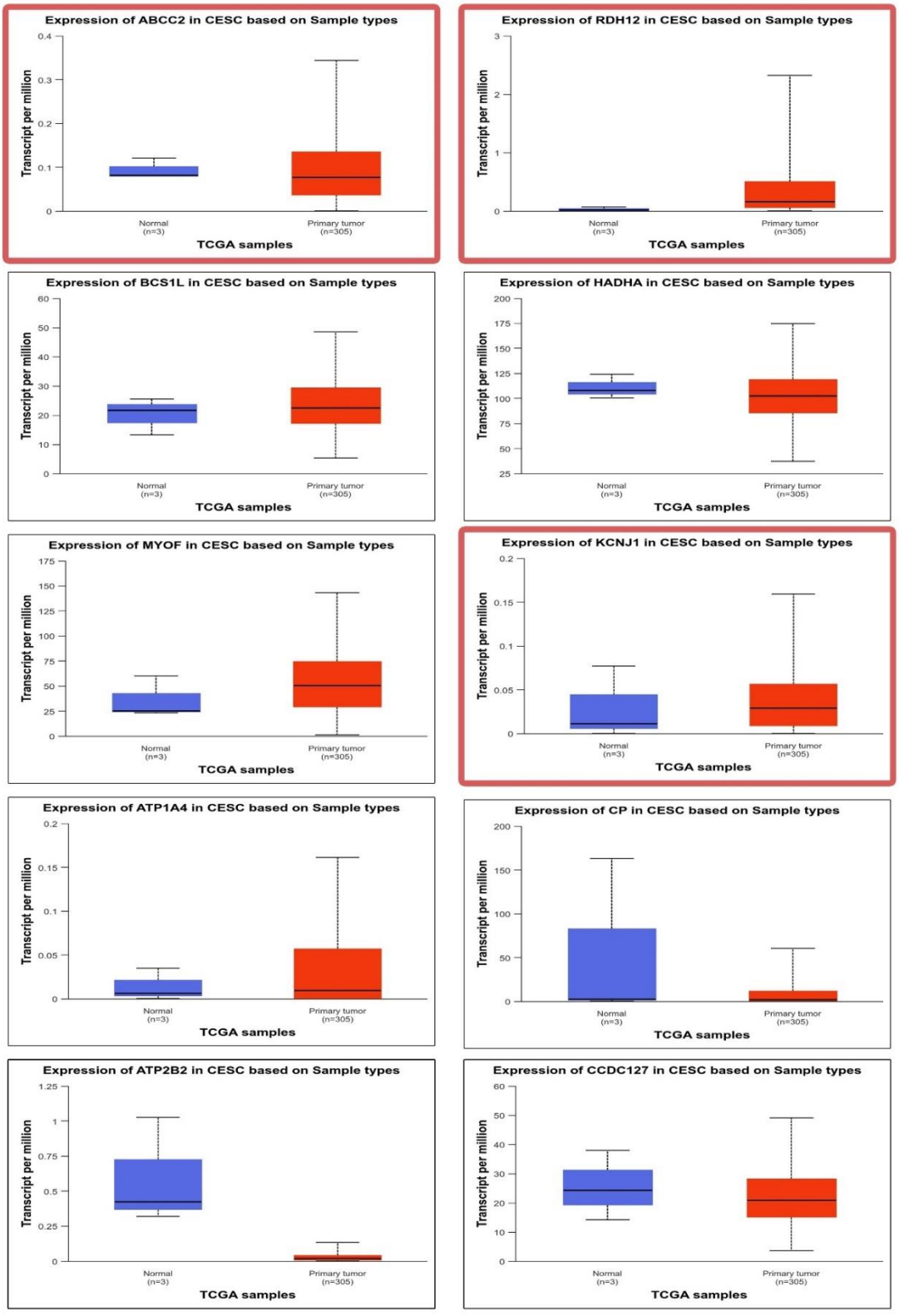
Out of the 31 mutated genes identified, 20 were found to interact effectively with other proteins, significantly disrupting critical biological processes. This led us to narrow the focus to 20 hub genes for further analysis. As shown in Table 21, to validate the expression levels of these key genes in squamous cell cervical carcinoma, we utilized the UALCAN platform, which integrates data from TCGA and GTEx. The results highlight key insights into the potential roles of these genes in disease progression.

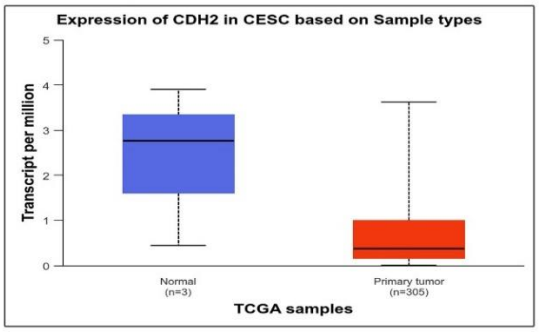
In this study, several genes were identified as potential biomarkers for squamous cell cervical carcinoma based on their differential expression across multiple comparisons. Among these, RDH12 showed highly significant changes in expression across all comparisons ($p < 1E-09$), indicating a strong association with the tumor phenotype and suggesting its potential role in disease-specific processes. This consistent dysregulation across tumor and age conditions makes RDH12 a strong candidate for further exploration as a diagnostic or prognostic biomarker. Similarly, RFC4 exhibited exceptionally low p-values across all comparisons, particularly in the Normal vs. Primary and Squamous-cell vs. Normal comparisons ($p < 1E-12$), highlighting its pivotal role in the tumor progression and suggesting its potential as a biomarker for early detection or therapeutic targeting. AIM1L also displayed highly significant changes in expression ($p < 1E-12$), further supporting its role as a key player in cervical cancer. Both RFC4 and AIM1L could serve as biomarkers for tumor progression and may have therapeutic potential in targeting cervical carcinoma.

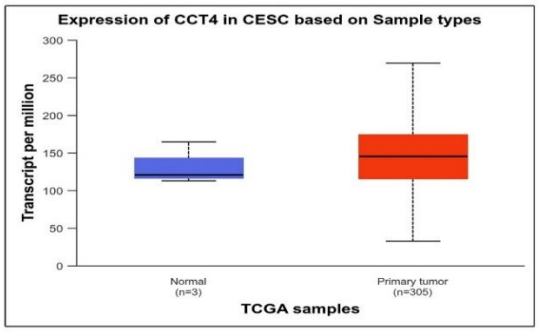
Table 21. Statistical significance of gene expression in mutated genes across different comparisons

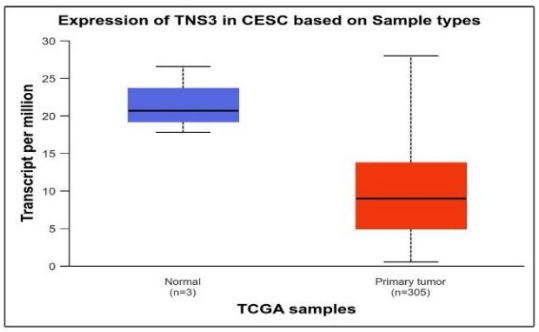
Gene name	Statistical significance		
	Normal-vs-Primary	**Squamous-cell-VS-Normal**	**Normal-vs-Age (41-60Yrs)**
ABCC2	7.477900E-03	1.868450E-02	**9.901000E-03**
RDH12	2.48209997000259E-09	3.71940001020477E-09	**9.67050000000302E-05**
BCS1L	4.472000E-01	5.141800E-01	5.209400E-01
HADHA	7.933400E-01	7.628600E-01	6.710600E-01
MYOF	3.216200E-01	2.588000E-01	3.358800E-01
KCNJ1	4.448200E-02	6.145600E-01	**1.950380E-02**
ATP1A4	1.153620E-01	1.250810E-01	1.974240E-01
CP	9.812600E-01	7.044400E-01	8.685400E-01
ATP2B2	2.400000E-01	1.617690E-04	**4.389500E-02**
CCDC127	8.517600E-01	8.718600E-01	9.359400E-01
CDH2	4.899200E-01	6.184800E-01	4.458400E-01
CCT4	5.753800E-01	5.837000E-01	5.266600E-01
TNS3	1.977200E-01	7.330400E-02	1.985240E-01
ACVR2B	4.884200E-01	3.343600E-01	5.417400E-01
RFC4	<1E-12	1.62447832963153E-12	**<1E-12**
CHD4	3.057600E-01	3.179400E-01	2.541400E-01
SLC4A3	1.775570E-01	1.799970E-01	1.852210E-01
TUBA8	9.574000E-01	3.161200E-01	7.886600E-01
ACSF3	9.462000E-01	7.644200E-01	9.849000E-01
AIM1L	1.62458935193399E-12	<1E-12	**<1E-12**

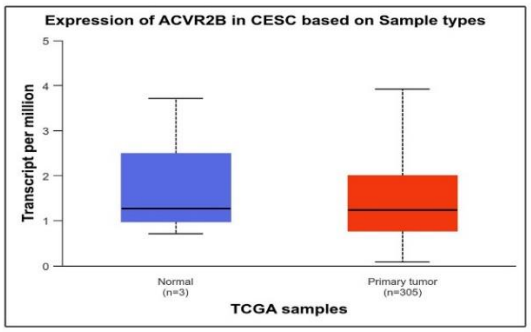
Figure 12. Differential expression of key genes in squamous cell cervical carcinoma vs. normal tissues based on TCGA data (the red box represents significant results, while the black box denotes non-significant results.)

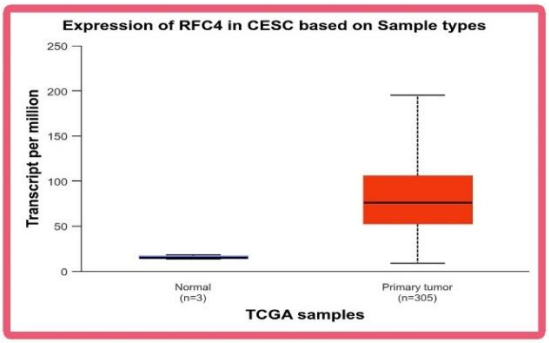


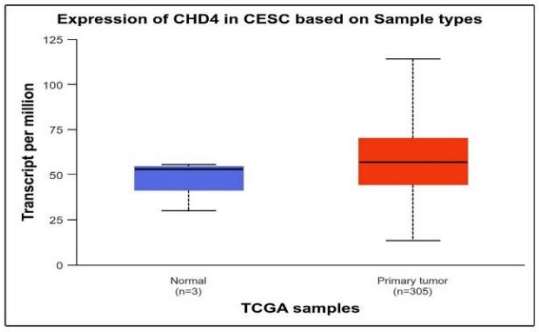

Expression of CDH2 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples

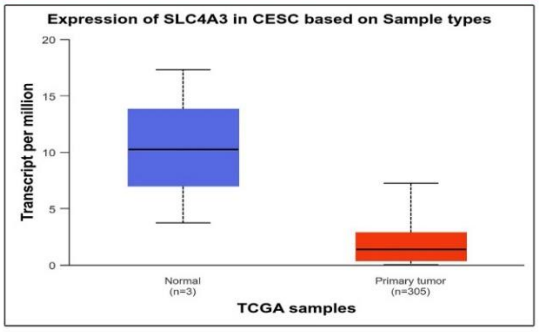

Expression of CCT4 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples

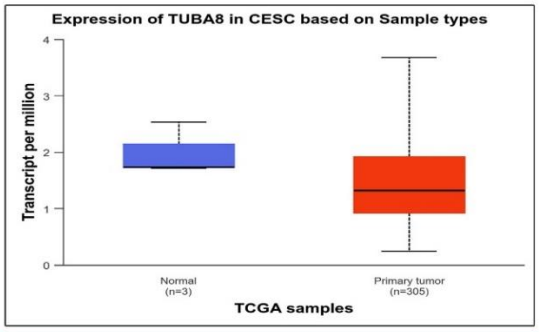

Expression of TNS3 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples

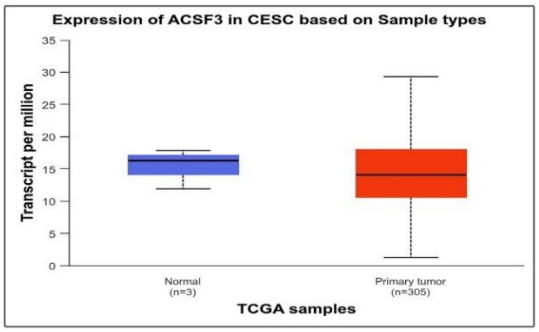

Expression of ACVR2B in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples


Expression of RFC4 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples


Expression of CHD4 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples


Expression of SLC4A3 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples


Expression of TUBA8 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples


Expression of ACSF3 in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples

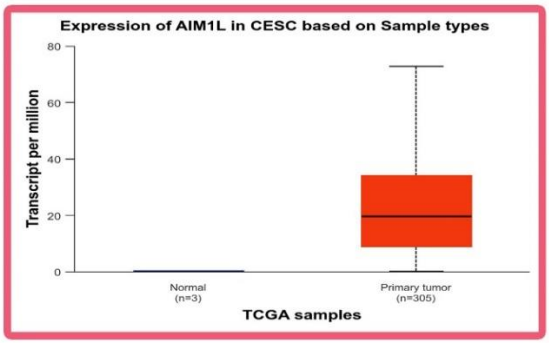
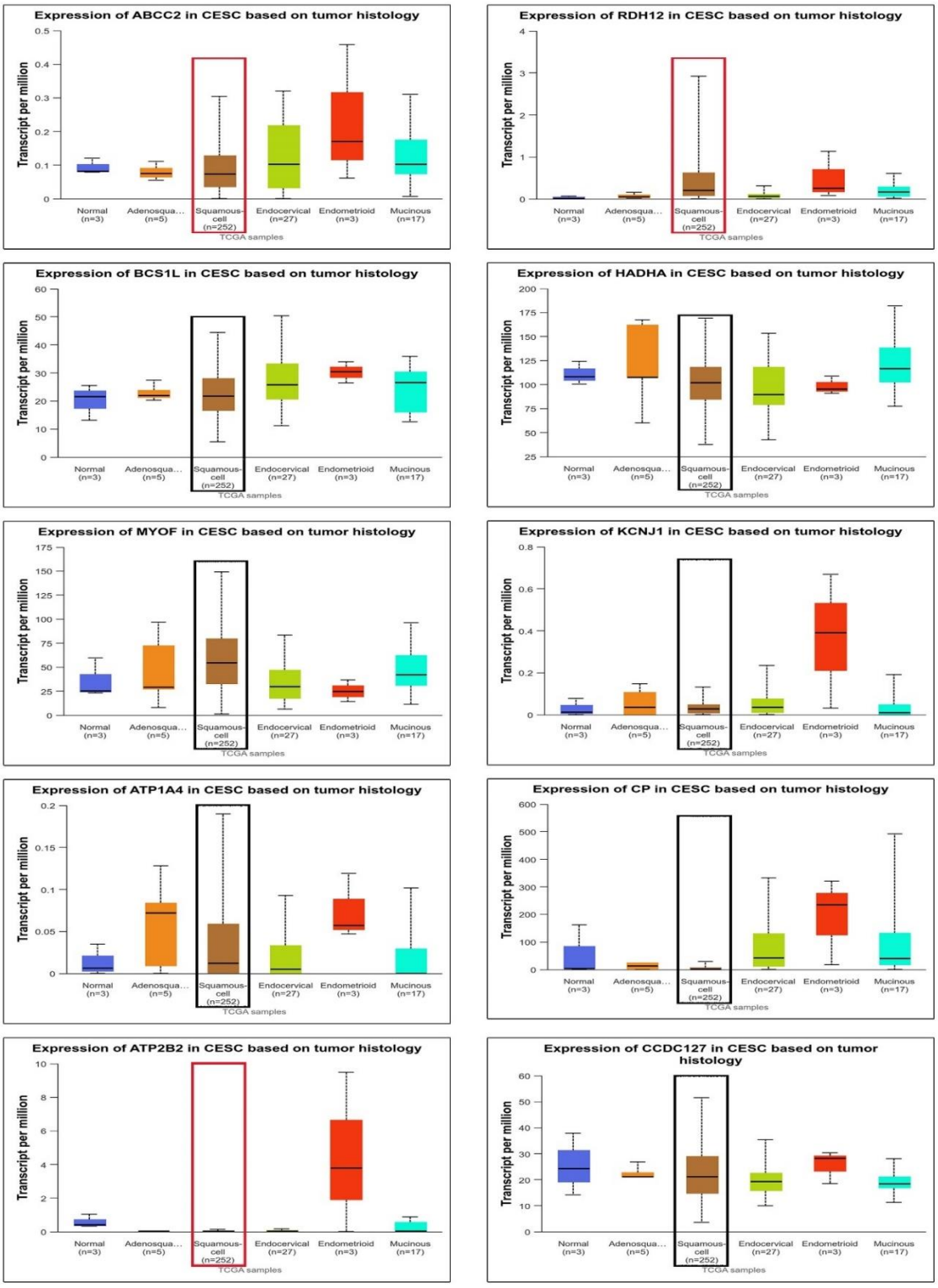
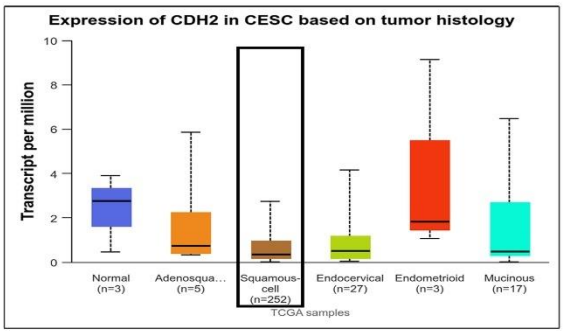
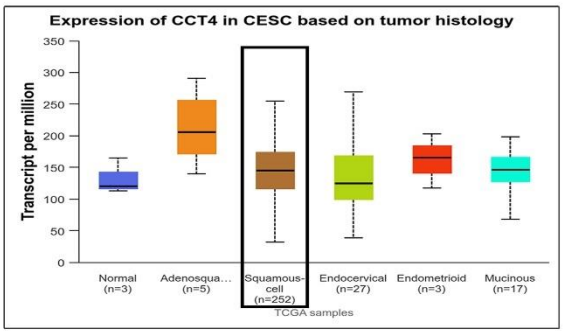
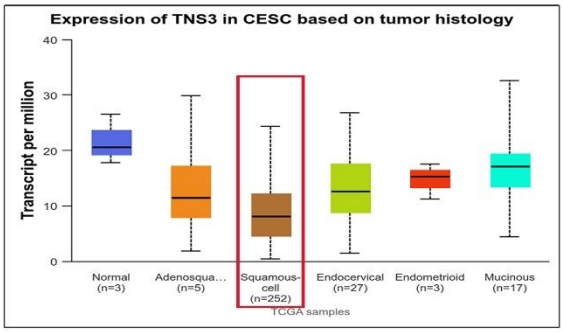

Expression of AIM1L in CESC based on Sample types
Transcript per million
Normal (n=3)
Primary tumor (n=305)
TCGA samples

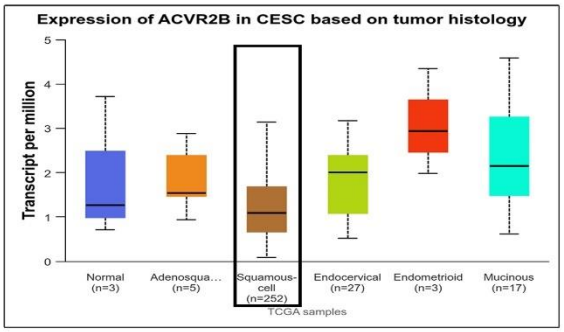
Figure 13. Differential expression of key genes in squamous cell cervical carcinoma vs. squamous cell cervical cancer based on TCGA data (the red box represents significant results, while the black box denotes non-significant results)

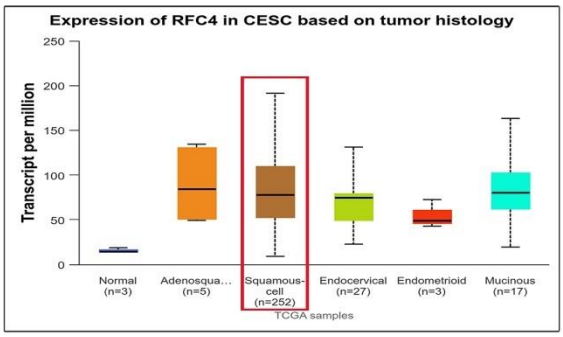


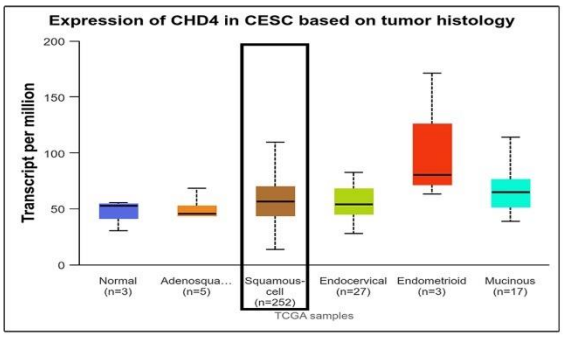

Expression of CDH2 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples

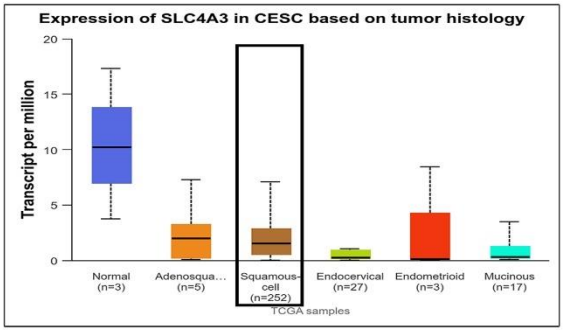

Expression of CCT4 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples

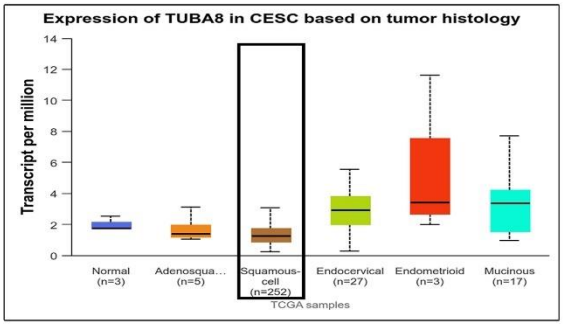

Expression of TNS3 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples

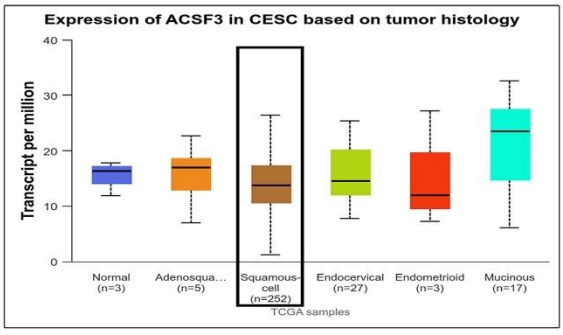

Expression of ACVR2B in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples


Expression of RFC4 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples


Expression of CHD4 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples


Expression of SLC4A3 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples


Expression of TUBA8 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples


Expression of ACSF3 in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples

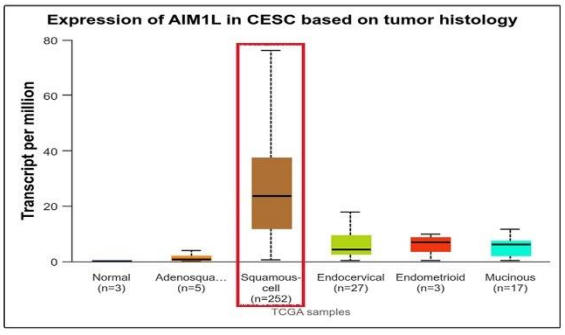
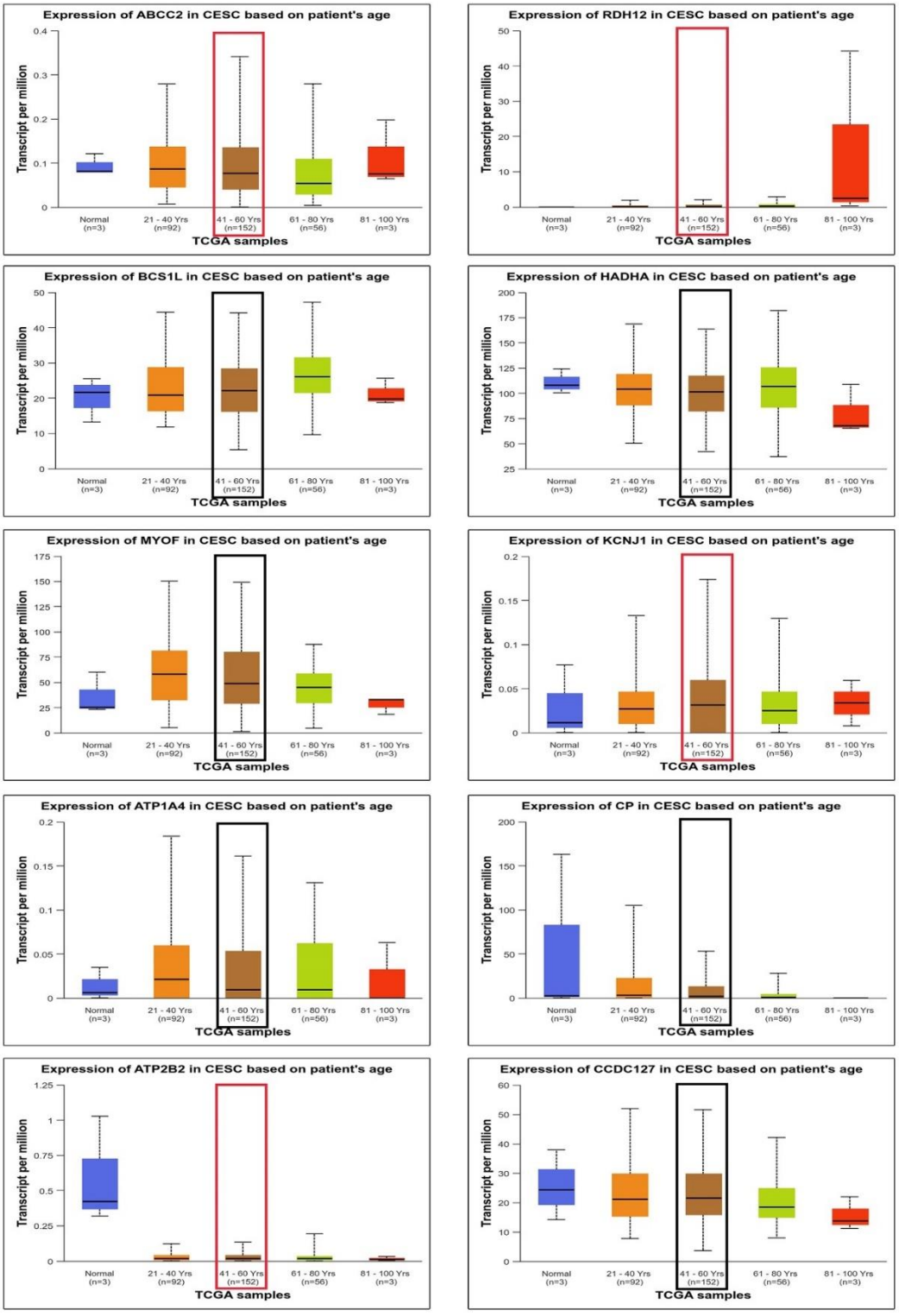
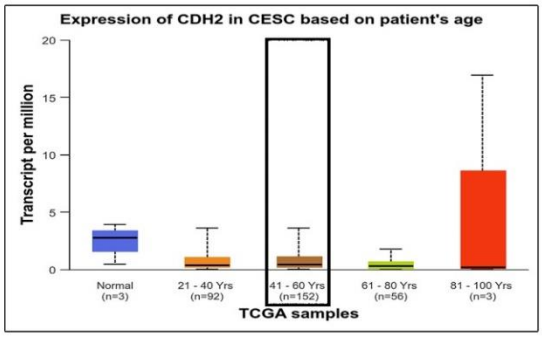
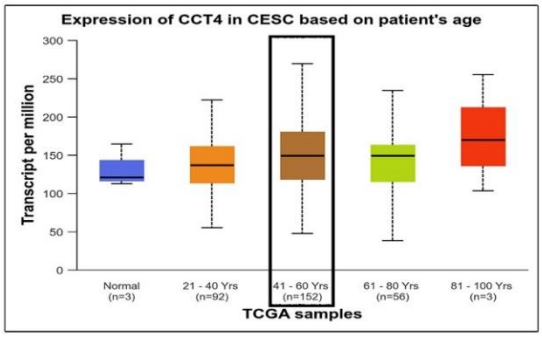
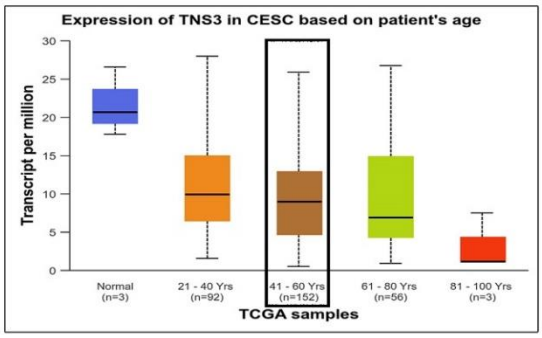

Expression of AIM1L in CESC based on tumor histology
Transcript per million
Normal (n=3)
Adenosqua... (n=5)
Squamous-cell (n=252)
Endocervical (n=27)
Endometrioid (n=3)
Mucinous (n=17)
TCGA samples

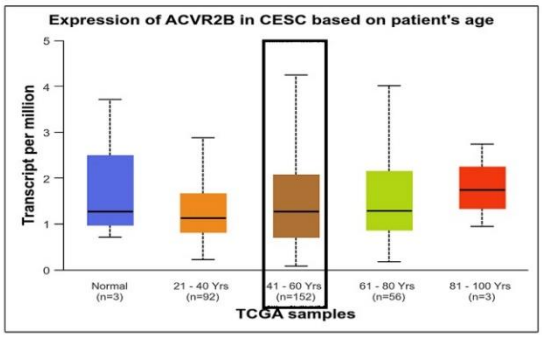
Figure 14. Differential expression of key genes in squamous cell cervical carcinoma vs. patients age based on TCGA data (the red box represents significant results, while the black box denotes non-significant result)

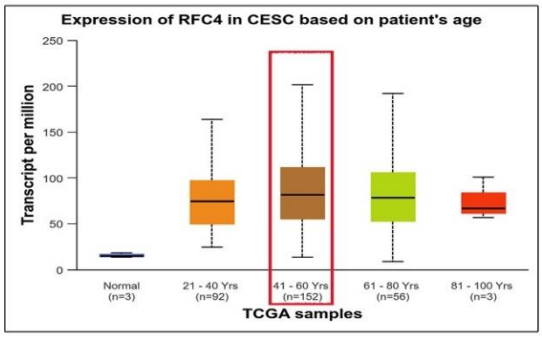


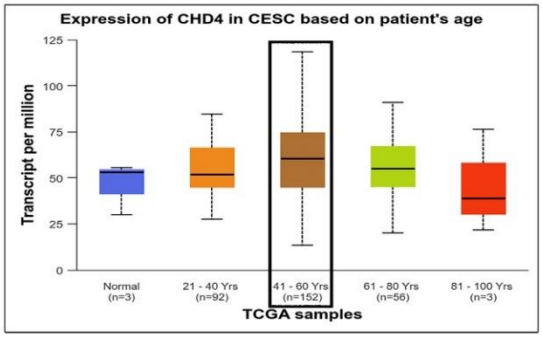

Expression of CDH2 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples

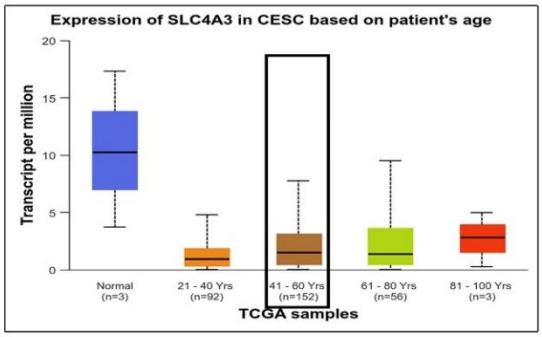

Expression of CCT4 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples

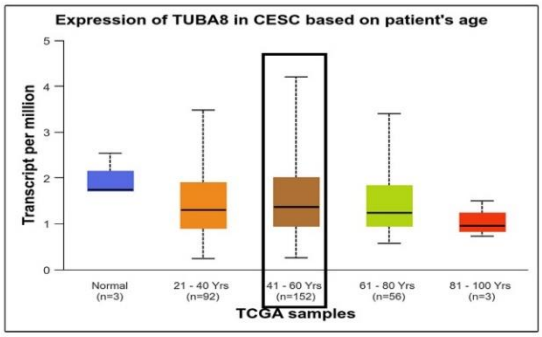

Expression of TNS3 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples

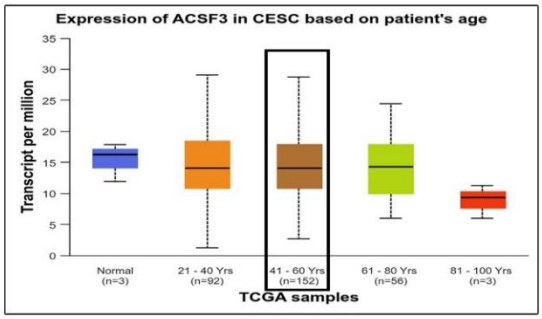

Expression of ACVR2B in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples

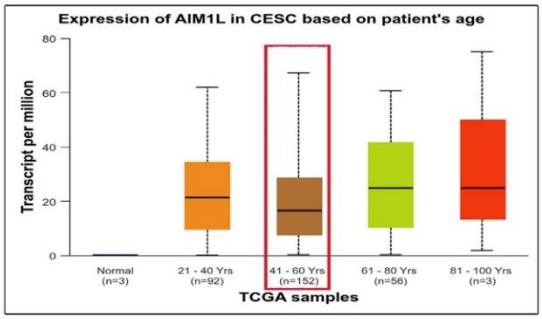

Expression of RFC4 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples


Expression of CHD4 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples


Expression of SLC4A3 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples


Expression of TUBA8 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples


Expression of ACSF3 in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples


Expression of AIM1L in CESC based on patient's age
Transcript per million
Normal (n=3)
21 - 40 Yrs (n=92)
41 - 60 Yrs (n=152)
61 - 80 Yrs (n=56)
81 - 100 Yrs (n=3)
TCGA samples

Additionally, ATP2B2 demonstrated significant differences in expression between squamous cell carcinoma and normal tissues (p = 1.62E-04) and between normal tissues and the age group 41-60 years (p = 4.39E-02), suggesting that it may be involved in age-related regulatory processes as well as tumor-specific alterations. This gene holds promise as a potential biomarker for both age-related tumor susceptibility and cancer treatment. The gene KCNJ1 showed moderate significance in the Normal vs. Primary comparison (p = 4.45E-02), indicating a possible role in the tumorigenic process and warranting further investigation to confirm its potential as an early tumor marker.

In contrast, ABCC2 displayed significant expression differences across the Normal vs. Primary (p = 7.48E-03), Squamous-cell vs. Normal (p = 1.87E-02), and Normal vs. Age (p = 9.90E-03) comparisons, suggesting that it may be influenced by both tumor-specific and age-related factors. This makes ABCC2 a potential biomarker for understanding the interplay between tumor development and aging. On the other hand, genes like TNS3 (p = 7.33E-02) exhibited moderate significance, and genes such as ACSF3, TUBA8, and CHD4 showed no significant differential expression across comparisons (p > 0.05), suggesting a lower likelihood of their involvement in the pathogenesis of squamous cell cervical carcinoma.

This study identifies RDH12, RFC4, AIM1L, ATP2B2, and ABCC2 as promising biomarkers for squamous cell cervical carcinoma, with potential applications in early detection, prognosis, and therapeutic targeting. Their consistent and significant dysregulation implies they may serve as biomarkers or therapeutic targets. On the other hand, the lack of statistical significance for certain genes suggests they may not exhibit differential expression in this context or that their roles require further functional validation. However, further validation of these genes through functional studies and clinical trials is necessary to confirm their clinical utility and to better understand their roles in tumor biology and aging.

In our comprehensive analysis of the mutational landscape of squamous cell cervical cancer, we initially identified 83,875 genetic variants. Of these, 4,092 variants were located within the coding regions of genes. Among the coding

variants, 2,389 were classified as nonsynonymous, which are particularly significant as they lead to amino acid changes in the encoded proteins. To enhance the reliability of our findings, we applied a read depth filter, retaining only variants with a depth greater than 20 reads. This filtering process yielded 2,356 high-confidence variants. When these were cross-referenced with the dbSNP database, 2,174 were identified as previously reported variants. Importantly, we identified 37 harmful nonsynonymous mutations affecting key genes involved in immune response, metabolism, DNA repair, and cell signalling pathways. Additionally, 182 novel mutations were discovered, providing opportunities for further research. A protein-protein interaction (PPI) network constructed from the mutated genes revealed significant disruptions in critical pathways, including the Hippo signalling and TGF-beta pathways, which are known to play pivotal roles in cancer progression. This study focused on the analysis of 31 mutated genes in squamous cell cervical carcinoma, specifically in patients aged 50–55 years. Due to the unavailability of normal sample sequencing data from the study's collection, the UALCAN database, which integrates TCGA and GTEx data, was used for comparison. Gene expression was evaluated across three key conditions: Normal vs. Primary Tumor, Squamous Cell vs. Normal, and Normal vs. Age (41–60 years). Of the 31 mutated genes, 20 were identified as hub genes based on their significant interactions with other proteins, which disrupt critical biological processes. These hub genes were validated using the UALCAN platform to ensure robust findings. The study identified several key genes as highly significant biomarkers. RDH12 emerged as a strong candidate for diagnostic or prognostic use, as it demonstrated consistent dysregulation across all comparisons ($p < 1E-09$). Similarly, RFC4 exhibited exceptionally low p-values across all comparisons ($p < 1E-12$), underscoring its pivotal role in tumor progression and its potential for early detection or therapeutic targeting. Another notable gene, AIM1L, also showed highly significant changes in expression ($p < 1E-12$), further confirming its critical involvement in cervical cancer.

In addition, moderately significant biomarkers were identified. ATP2B2 demonstrated significant differential expression in Squamous Cell vs. Normal

(p = 1.62E-04) and Normal vs. Age (p = 4.39E-02), suggesting its involvement in both tumor-specific and age-related regulatory processes. ABCC2 was also significant across all comparisons (p = 7.48E-03 to 1.87E-02), indicating its role in tumor progression and aging-related changes. KCNJ1 showed moderate significance in the Normal vs. Primary comparison (p = 4.45E-02), pointing to a potential role in the tumorigenic process.

Conversely, some genes, such as TUBA8, ACSF3, and CHD4, did not exhibit significant expression changes across the comparisons (p > 0.05), suggesting they may not play prominent roles in squamous cell cervical cancer or that their roles require further functional validation. Overall, this study highlighted RDH12, RFC4, AIM1L, ATP2B2, and ABCC2 as promising biomarkers for early detection, prognosis, and therapeutic targeting in cervical cancer. However, further functional studies and clinical validation are necessary to confirm their clinical utility and to better understand their roles in tumor biology and aging.

Phase IV

4.4. Validation of novel genetic variants

In phase IV, mutational profiling and confirmation of novel variant were focused. We started with exome sequencing, identifying novel missense variants in squamous cell cervical cancer, thereby enabling us to analyze all protein-coding regions of the genome and capture genetic variations relevant to cancer development. Thereafter, we prioritized variants based on their potential impact on protein functions and disease progression. The following variants were further validated using Sanger sequencing and by the precise design of the primers coupled with optimization of PCR to avoid amplifying the wrong template or sequencing during amplification and subsequent sequencing steps. These confirmations establish the true presence of new variants, showing the reliability of the findings presented here. These variants contribute knowledge regarding the etiology of cervical cancer and thus might help advance therapeutic strategies against cervical cancer towards efficient diagnosis, therapy, and prevention.

4.4.1. Molecular genetics of squamous cell cervical cancer: novel insights

Among the 2389 variants identified by WES, 182 nonsynonymous variants were not reported in the dbSNP database, indicating that these mutations may be novel genetic alterations. This suggests the possibility of discovering previously uncharacterized mutations that may contribute to the pathogenesis of squamous cell cervical cancer. Further investigation of these unreported variants may provide new insights into the mechanisms of the disease and open avenues for future research into targeted therapies.

Due to the number of variants identified and the size of the sample, we were unable to confirm the 182 unreported variants. Thus, we adopted a strategy of prioritizing variants that matched the ClinVar database. Of the 182 variants, 8 have been reported in ClinVar as benign, of uncertain significance, or pathogenic. The remaining unreported variants are provided in the Supplementary File 3.

As a result, eight nonsynonymous mutations across seven genes—POM121C, PRICKLE1, SRPX2, KIF1B, BRAT1, ALOX12B, and GLIS3 (Table 22)—were selected for validation. These mutations alter the amino acid sequences of proteins, potentially affecting their structure and function and thereby influencing biological processes linked to cancer progression. The discovery of these specific variants provides valuable insights into the genetic basis of squamous cell cervical cancer and opens up new avenues for research into how genetic mutations drive the disease.

The table outlines seven nonsynonymous mutations identified in squamous cell cervical cancer samples (CC1–CC5) through whole-exome sequencing. These mutations are distributed across seven genes: POM121C, PRICKLE1, SRPX2, KIF1B, BRAT1, ALOX12B, and GLIS3, and are described in terms of their genomic locations, nucleotide changes, amino acid substitutions, and sequencing characteristics.

The mutations occur on various chromosomes, with each sample harbouring specific alterations. In sample CC1, mutations were detected in the POM121C gene on chromosome 7 (position 75424148) and the PRICKLE1 gene

on chromosome 12 (position 42464916). These mutations involved a change from T to G in POM121C, resulting in a threonine-to-proline substitution (p.T317P), and a change from C to T in PRICKLE1, leading to an arginine-to-glutamine substitution (p.R373Q). Both mutations occurred in exons, specifically in exon 12 for POM121C and exon 7 for PRICKLE1, with read depths of 65 and 110, respectively, indicating reliable sequencing coverage.

Table 22. Identification of novel nonsynonymous variants associated with squamous cell cervical cancer by whole exome sequencing

Sample ID	Chr	Position	Ref	Alt	Gene name	Read Depth	Exon position	AA change	dbSNP id
CC1	chr7	75424148	T	G	POM121C	65	12	p.T317P	-
CC1	chr12	42464916	C	T	PRICKLE1	110	7	p.R373Q	-
CC2	chrX	100664789	C	T	SRPX2	156	5	p.A124V	-
CC3	Chr1	10342161	C	G	KIF1B	30	31	p.L1163V	-
CC3	chr7	2543882	C	T	BRAT1	31	5	p.V171I	-
CC3	chr17	8072817	T	C	ALOX12B	51	15	p.Y687C	-
CC4	chr9	3828361	G	A	GLIS3	78	10	p.R747C	-
CC5	chr9	3828361	G	A	GLIS3	63	10	p.R747C	-

In CC2, a mutation was identified in the SRPX2 gene on the X chromosome (position 100664789), where a C-to-T transition caused an alanine-to-valine substitution (p.A124V) in exon 5, supported by a high read depth of 156. Similarly, CC3 had mutations in three genes: KIF1B at chromosome 1, position 10342161; BRAT1 at chromosome 7, position 2543882; and ALOX12B at chromosome 17, position 8072817. These mutations caused the following amino acid changes: leucine to valine (p.L1163V) in KIF1B, valine to isoleucine (p.V171I) in BRAT1, and tyrosine to cysteine (p.Y687C) in ALOX12B, with read depths ranging from 30 to 51. Interestingly, the GLIS3 gene on chromosome 9 at position 3828361 had a recurrent mutation of G to A in two different samples, CC4 and CC5. This mutation was found in exon

10, resulting in an arginine-to-cysteine substitution at p.R747C, with read depths of 78 and 63, respectively. The recurrence of this mutation in several samples suggests its possible implication in the pathogenesis of squamous cell cervical cancer.

Notably, none of the identified mutations had been reported previously in the dbSNP database as indicated by the absence of any dbSNP IDs. These mutations are likely novel and may thus represent unique genetic alterations in this cancer type. The amino acid substitutions resulting from these mutations are likely to greatly affect protein structure and function, thus altering biological processes involved in cancer progression. These findings lay a foundation for further investigation into the functional consequences of these genetic changes and their role in the disease.

4.4.2. Mutational analysis of novel nonsynonymous variants

These novel variants were then selected for mutational analysis using a different set of bioinformatic tools that we have used in already reported variants. These tools provide an evaluation of the amino acid substitution at the protein function level, providing significant insights into the molecular mechanism involved in disease pathogenesis.

The Table 23 presents an in-depth analysis of seven nonsynonymous mutations identified in the genes POM121C, PRICKLE1, SRPX2, KIF1B, BRAT1, ALOX12B, and GLIS3 across various squamous cell cervical cancer samples. Functional prediction tools were employed to assess the potential impact of these mutations on protein function, with an emphasis on their possible contribution to disease progression.

Functional Prediction:

The SIFT tool predicted that most of the mutations, POM121C, PRICKLE1, and GLIS3, were tolerated (T), suggesting that these mutations might not drastically disrupt protein functions. However, ALOX12B mutation was predicted to be deleterious (D), meaning its structure and function are significantly impacted, which in turn may lead to cancer.

Table 23. Predictive analysis of nonsynonymous variants in squamous cell cervical cancer using functional prediction tools

Sample ID	Gene.ref Gene	SIFT	Polyphen2	LRT	Mutation Taster	FATHMM	Radial SVM	LR	CADD	GERP ++_RS	phyloP46way placental	phyloP100 way vertebrate	SiPhy 29way logOdds
CC1	POM121C	T	B	N	N	T	T	T	10.08	-1.66	-0.082	-0.239	5.473
CC1	PRICKLE1	T	P	D	N	T	T	T	10.4	3.76	1.505	3.558	12.945
CC2	SRPX2	.	B	N	N	T	T	T	9.91	3.43	1.064	1.123	5.936
CC3	KIF1B	T	B	D	D	T	T	T	1.013	3.78	1.11	4.676	13.146
CC3	BRAT1	T	B	N	N	T	T	T	0.018	-9.02	-1.908	-0.545	4.478
CC3	ALOX12B	D	D	D	D	D	D	D	23.9	4.89	1.851	7.312	13.536
CC4	GLIS3	T	B	D	N	T	T	T	7.674	4.04	1.458	1.1	3.617
CC5	GLIS3	T	B	D	N	T	T	T	7.674	4.04	1.458	1.1	3.617

Mutations in PRICKLE1 were classified as probably damaing by PolyPhen-2, and the mutation in ALOX12B was classified as possibly damaging, implying that these changes might have a significant effect on the proper functioning of the protein. Conversely, mutations in POM121C, SRPX2, and BRAT1 were classified as benign, implying that these alterations are unlikely to impact protein function or disease pathogenesis.

LRT was predicted to be deleterious (D) for mutations in PRICKLE1, KIF1B, ALOX12B, and GLIS3, consistent with the expectation that these genes may be implicated in critical cancer-related pathways. Mutation Taster classified the ALOX12B mutation as disease-causing (D), whereas all other mutations were neutral (N). FATHMM predicted various mutations to be either deleterious or tolerated. While ALOX12B raised again as the high-potential pathogenic mutation. RadialSVM and LR mainly showed most mutations from POM121C, PRICKLE1, and GLIS3 are tolerated (T), whereas, on the other hand, ALOX12B was reported by all of these tools as consistently deleterious (D).

Conservation Scores:

The CADD scores, which combine several sources to predict the deleteriousness of a mutation, were highest for ALOX12B at 23.9, indicating it is most likely to have a functional impact. The score for the mutation in PRICKLE1 was relatively high at 10.4, and thus further supports the likelihood of pathogenicity. However, mutations in BRAT1 and SRPX2 both possessed much lower scores, indicating these changes are unlikely to have a major impact on protein function.

Evolutionary conservation scores, including GERP++_RS, phyloP46, and phyloP100, provided additional evidence for the functional importance of these mutations. For instance, ALOX12B showed a high conservation score at all points, meaning that it is a very conserved gene with crucial biological functions. On the other hand, BRAT1 had low GERP++ and phyloP scores, suggesting that the mutation may not be functionally significant when compared to other variants.

The predictive analysis predicted that the ALOX12B variant would likely be deleterious and probably damaging, whereas GLIS3 variants might contribute to

the disease pathogenesis. Other variants were predicted to be benign or tolerated by at least two tools. Such findings could provide a foundation for future studies on functional characterization, targeted therapeutic strategies, and genetic counseling.

Such predictions made in silico must be validated through experimentation to determine their accuracy. This integration of computational and experimental approaches will further advance our understanding of the molecular mechanisms behind squamous cell cervical carcinoma and eventually lead to effective diagnostic and therapeutic strategies.

4.4.3. Specific primers for novel variants

To confirm these novel genetic variants identified in samples of squamous cell cervical cancer, specific primers were designed using Primer3plus. This would help researchers design personalized primers, which could be utilized to amplify certain target gene regions surrounding the variant. Designed primers are significant for precise validation of variants using PCR amplification followed by Sanger sequencing.

A summary of the primer design details is presented in Table 24, outlining target genes, forward primers, reverse primers, and amplicon size of each variant. Amplicons are designed with a range of 170 to 247 base pairs so that PCR amplification is effective. In addition, the primers have been designed to flank the variant regions for validation purposes. Furthermore, the primer sequences are specific to target genes to avoid off-target amplification.

Primer3Plus was employed to design the primers. These parameters include primer length (18-27 nucleotides), GC content (40-60%), melting temperature (58-62°C), and amplicon size (100-300 bp). These settings ensure optimal primer performance and specificity.

Table 24. Mutation-specific primers for PCR amplification

Target gene	Primers	Amplicon size(bp)
POM121C	AATCTCCAGCGACTGACAGG	248
	GAGACCCCACCTACCACTCA	
PRICKLE1	CCAAACTTGAGGAGGAGCTG	195
	CAAGTTTCCTGGCCTCTCAG	
SRPX2	ACTGAGCTTGCCACTTGGTT	179
	ATCTCCCATCTTCCATGCAG	
KIF1B	AGCAAGACCTTGCCTCAAAA	221
	TGTCCTTGCAGATGAAGTGG	
BRAT1	CGTGATCCATGATCTTCTGG	171
	CGGTCGACACCATCTTCTC	
ALOX12B	CACAGAATGGGGAGAGGAGA	204
	CACTTCCCGGACATTCACTT	
GLIS3	CAGAGGAGAGCTGGCTAGGA	226
	CTTAGCCCCCAGTCCTCACT	

The successful design of these primers enables accurate confirmation of the novel variants. This validation is critical for confirming the presence and accuracy of the variants, providing valuable insights into the genetic basis of squamous cell cervical cancer. The implications of this study are profound, as it may lead to improved understanding of squamous cell cervical cancer genetics, potential identification of biomarkers or therapeutic targets, and enhanced diagnostic and treatment strategies.

Validation of designed primers using *In silico* PCR and touchdown PCR

i. *in silico* PCR validation:

Figure 15. *In silico* validation of specifically designed primers

UCSC In-Silico PCR

>chr7:75423955+75424202 248bp AATCTCCAGCGACTGACAGG GAGACCCCACCTACCACTCA
AATCTCCAGCGACTGACAGGcacgacgcccttattccaacccaagccggg
gatgctgctcagaaggctggatccacggccccactccagctcacctgggc
agggtggcaggctcgggggagtctgcatcttcttcaagctctctaacagt
gggttggtgcttggggccaggagggaggtgggtggggaggcagttgcagc
agcaggcagggtaaaggtaaatgaaggcTGAGTGGTAGGTGGGGTCTC

UCSC In-Silico PCR

>chr12:42464766+42464960 195bp CCAAACTTGAGGAGGAGCTG CAAGTTTCCTGGCCTCTCAG
CCAAACTTGAGGAGGAGCTGcgtcatataatcttcatgatcagcccattc
ttcagggtcttctggagtttcctgctctactctgcctttccaaaattctt
cactggcaaaacttgttccttgtctggagagactcagatcatccaatttt
cgagaaagggtgtcatcagcattgcCTGAGAGGCCAGGAAACTTG

UCSC In-Silico PCR

>chrX:100664745+100664923 179bp ACTGAGCTTGCCACTTGGTT ATCTCCCATCTTCCATGCAG
ACTGAGCTTGCCACTTGGTTttctcttagagatgagatgccacgcactac
cattcatcactagtggcacttacacctgcacaaatggagtgcttcttgac
tctcgctgtgactacagctgttccagtggctaccacctggaaggtgatcg
cagccgaatCTGCATGGAAGATGGGAGAT

UCSC In-Silico PCR

>chr1:10341936+10342156 221bp AGCAAGACCTTGCCTCAAAA TGTCCTTGCAGATGAAGTGG
AGCAAGACCTTGCCTCAAAAaaaaaaaaaaaaacccaaaatacgtactaa
agttctgattgaagcaaaaatgtgtattttcttgtaatcttttcctaatc
ttgcttggctttagattgcagtggagatcactgaatcatttgtggattac
atcaaaaccaagcctattgtatttgaagtctttgggcattatcagcagca
cCCACTTCATCTGCAAGGACA

UCSC In-Silico PCR

>chr7:2543789+2543959 171bp CGTGATCCATGATCTTCTGG CGGTCGACACCATCTTCTC
CGTGATCCATGATCTTCTGGgcacacgcgggccagtcacccccggcagg
cagggctgcccctcggctccacctcgcatggacaaagccaggacgtgcac
caggagctgactggccgccgaggccacaaacaggctggagtctccctgca
ggGAGAAGATGGTGTCGACCG

UCSC In-Silico PCR

>chr17:8072735+8072938 204bp CACAGAATGGGGAGAGGAGA CACTTCCCGGACATTCACTT
CACAGAATGGGGAGAGGAGAgacgggaagcgcgctcctaaatagaaatgc
tgttctcaatcagcaccgggtccaggtagtagtaggggatgggaaggcac
ttgttgcgctggcggatgtcgtgtgagatctggttcaggcgctggcggaa
cgcctctatgctcctccgcggggcctcctccacgAAGTGAATGTCCGGGA
AGTG

UCSC In-Silico PCR

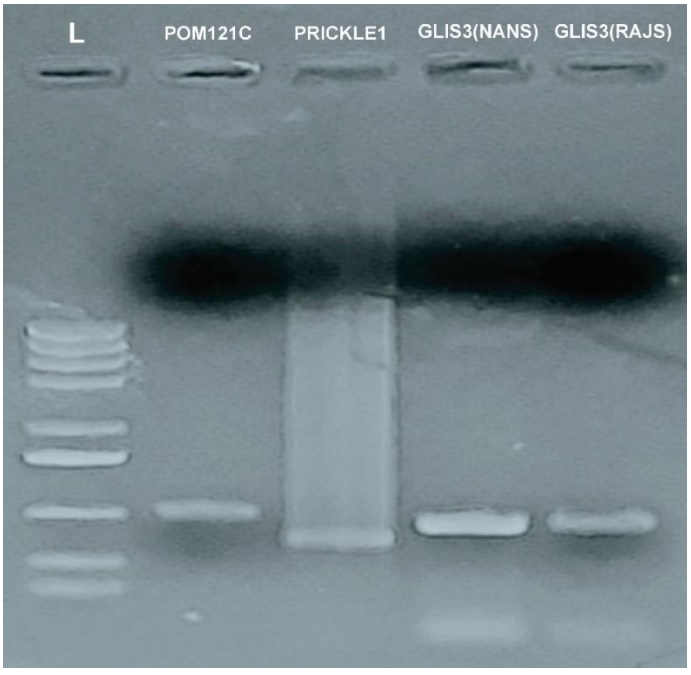
>chr9:3828289+3828514 226bp CAGAGGAGAGCTGGCTAGGA CTTAGCCCCCAGTCCTCACT
CAGAGGAGAGCTGGCTAGGAcagcggtccacggtgctgatctgcaagaag
gtagcatcttcagccccgctgcggagagactccccaaagaggctcgagga
acttgaaggtaaatcatacactggaagagaaagaacgcagttaagtcagt
aactcctgccccatgccacgcccactggaaatgcactacagtaatgcagc
tcaccaAGTGAGGACTGGGGGCTAAG

We validated the specifically designed primers using the UCSC Genome Browser tool for *in silico* PCR. Figure 15 illustrates the virtual validation of primers, where the UCSC Genome Browser tool was employed to simulate PCR amplification by comparing the primer sequences against the reference genome. This computational approach allowed us to predict the binding sites of the primers on the target sequence and simulate the amplification process. By performing *in silico* PCR, we were able to evaluate the specificity of the primers, ensuring that they bound exclusively to the intended target region without forming non-specific bindings or amplifying unintended sequences.

ii. touchdown PCR

PCR (Polymerase Chain Reaction) amplification was successfully achieved for three out of six genetic variants, despite limitations imposed by insufficient DNA quantity. This partial success highlights the challenges associated with working with limited DNA samples, which can hinder the efficiency and reliability of PCR amplification. The DNA quantity limitation likely resulted in insufficient template DNA, increased risk of PCR bias, and reduced sensitivity, making it difficult to generate sufficient amplicons for detection.

Figure 16. Gel electrophoresis of amplified DNA products.



The successful amplification of three variants, as illustrated in Figure 16, demonstrates that the designed primers were highly specific to the target regions of the DNA, ensuring accurate amplification of the desired sequences. Primer specificity is a critical factor in PCR (Polymerase Chain Reaction), as it ensures that only the target DNA segments are amplified without off-target binding, which can lead to nonspecific products. This successful amplification of the three variants confirms that the PCR protocol was well-designed and reliable for the detection of these specific genetic variants.

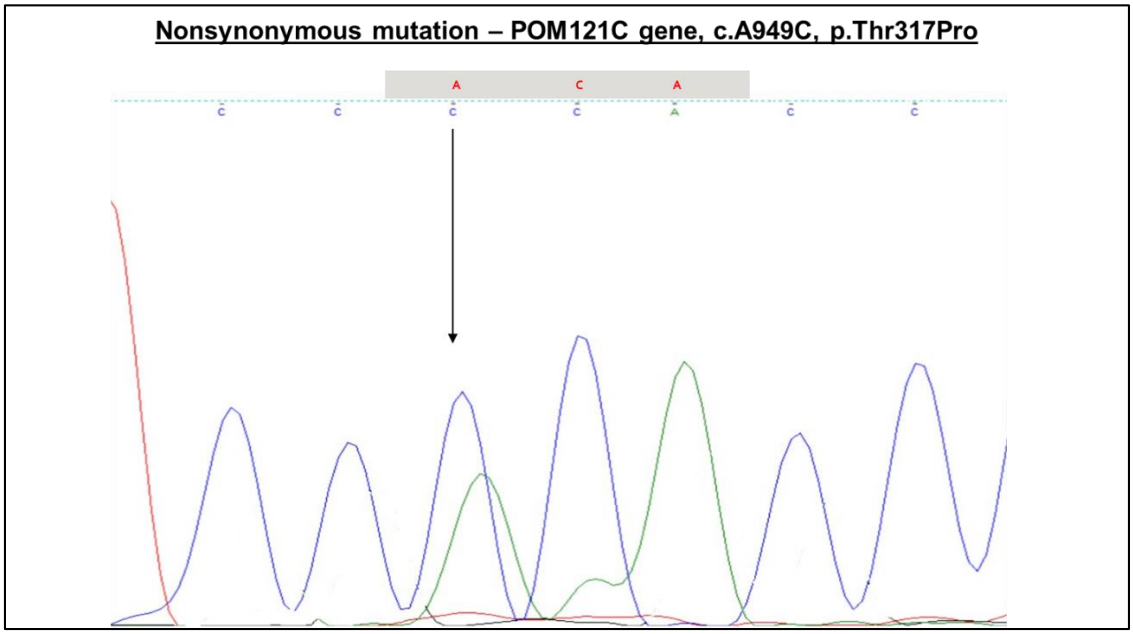
4.4.4. Sanger confirmation of novel variants in squamous cell cervical cancer

In order to confirm the novel variants in squamous cell cervical cancer, we performed sanger-sequencing. In the chromatogram, each peak represents one of the four nucleotides (A, T, G, C), and a part of the DNA that may be linked to critical genes involved in cellular processes.

Variant in POM121C gene: As shown in Figure 17, a nonsynonymous mutation was identified in the POM121C gene, specifically at nucleotide position 949, where an adenine (A) base was replaced by a cytosine (C) base, denoted as c.A949C. This point mutation leads to a change in the amino acid sequence of the resulting protein, altering the threonine (Thr) residue at position 317 to proline (Pro), represented as p.Thr317Pro. The POM121C gene encodes a component of the nuclear pore complex, essential for regulating nucleocytoplasmic transport.

The mutation has occured within a highly conserved region of the gene, suggesting potential functional significance. At the molecular level, the substitution of threonine to proline alters the chemical properties and structure of the protein. Threonine is a polar, hydrophilic amino acid, whereas proline is a non-polar, hydrophobic amino acid with a distinct ring structure. This change may disrupt local protein folding, potentially affecting protein stability and degradation, nuclear pore complex assembly and function, and transport of molecules across the nuclear envelope (Yu *et al*., 2024).

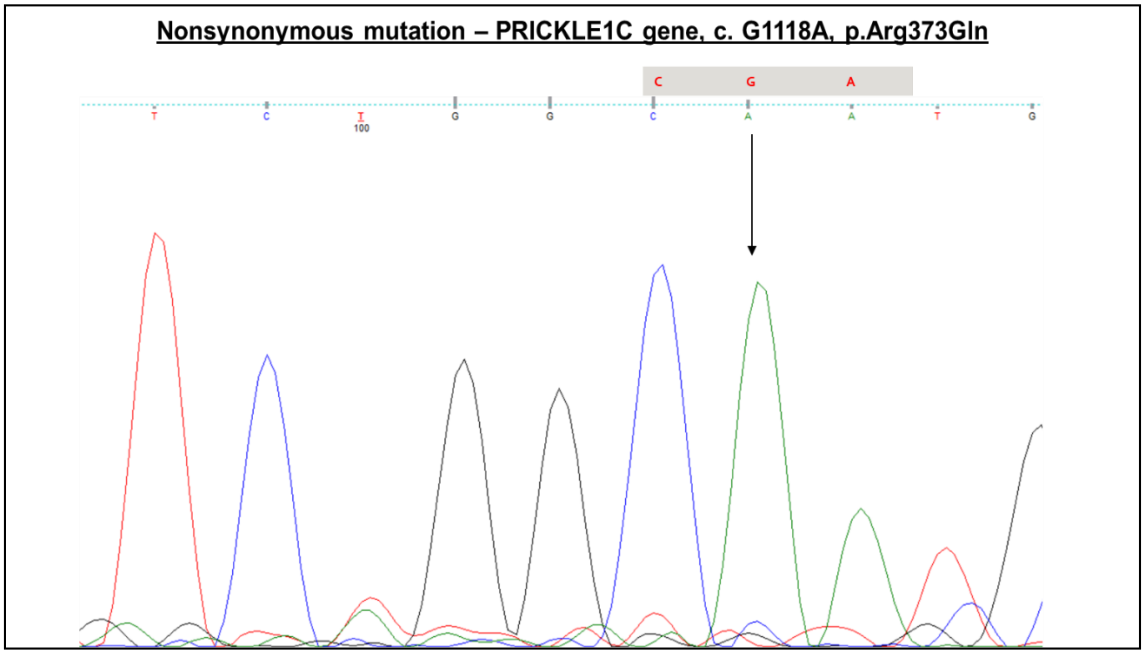
Figure 17. Sanger sequencing confirmation of POM121C variants in squamous cell cervical cancer sample



The key aspects of this mutation include its nonsynonymous nature, location in the POM121C gene, and potential functional impact on protein structure and function. Understanding the consequences of this mutation can provide valuable insights into the molecular mechanisms underlying squamous cell cervical cancer development and progression, highlighting the importance of continued research into the genetic basis of this disease.

Variant in PRICKLE1C gene: As shown in Figure 18, a nonsynonymous mutation was identified in the PRICKLE1 gene, specifically at nucleotide position 1118, where a guanine (G) base was replaced by an adenine (A) base, denoted as c.G1118A. This point mutation leads to a change in the amino acid sequence of the resulting protein, altering the arginine (Arg) residue at position 373 to glutamine (Gln), represented as p.Arg373Gln. The PRICKLE1 gene encodes a protein essential for planar cell polarity (PCP) signalling, regulating cellular orientation and migration during embryonic development and tissue maintenance.

Figure 18. Sanger sequencing confirmation of PRICKLE1C variants in squamous cell cervical cancer

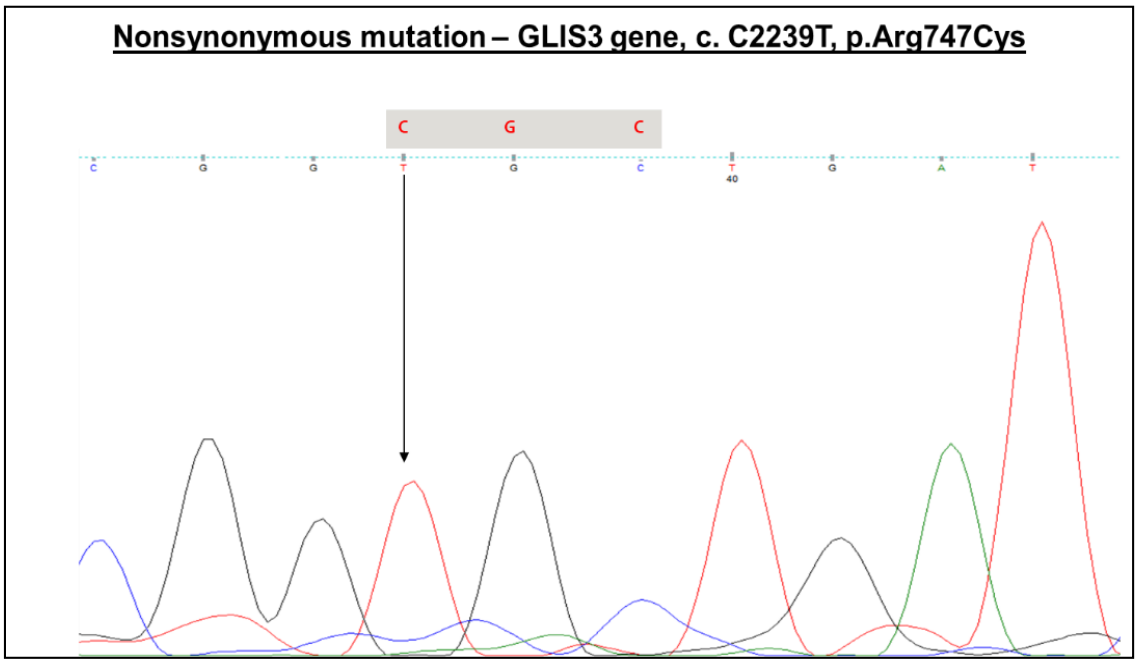


The mutation occurs within a highly conserved region of the gene, suggesting potential functional significance. At the molecular level, the substitution of arginine to glutamine alters the chemical properties and structure of the protein. Arginine is a positively charged, hydrophilic amino acid, whereas glutamine is a polar, hydrophilic amino acid. This change may disrupt local protein folding, potentially affecting protein-protein interactions, PCP signalling pathway regulation, and cellular orientation and migration (Radaszkiewicz *et al*., 2024). By exploring the molecular mechanisms underlying this mutation, researchers can uncover novel avenues for disease treatment and prevention, shedding light on the complex relationships between genetic mutations and human disease.

Variant in GLIS3 gene: As shown in Figure 19, a nonsynonymous mutation was identified in the GLIS3 gene, specifically at nucleotide position 2239, where a cytosine (C) base was replaced by a thymine (T) base, denoted as c.C2239T. This point mutation leads to a change in the amino acid sequence of the resulting protein, altering the arginine (Arg) residue at position 747 to cysteine (Cys), represented as

p.Arg747Cys. The GLIS3 gene encodes a transcription factor essential for pancreatic beta-cell function, insulin gene expression, and glucose metabolism.

Figure 19. Sanger sequencing confirmation of GLIS3 variants in squamous cell cervical cancer sample



When mutation occurs within a highly conserved region of the gene, it suggests potential functional significance. The substitution of arginine to cysteine alters the chemical properties and structure of the protein, potentially disrupting protein-DNA interactions, transcriptional regulation, and insulin signalling. This change may impair glucose-stimulated insulin secretion, pancreatic beta-cell function, and increase the risk of developing diabetes (Scoville and Jetten, 2021). The functional implications of this mutation are critical, highlighting the importance of continued research into the genetic basis of metabolic disorders. Understanding the consequences of the p.Arg747Cys mutation can provide valuable insights into the complex relationships between genetic mutations and disease development, ultimately informing personalized medicine approaches and improving disease management. The key aspects of this mutation include its nonsynonymous nature, location in the GLIS3 gene, and potential functional impact on protein structure

and function, emphasizing its relevance to diabetes, pancreatic disorders, and metabolic syndrome (Calderari *et al*., 2018).

The mutations SRPX2 c.C371A (p.A124V), BRAT1 c.G512A (p.V171I), KIF1B c.C3625G (p.L1163V), and ALOX12B c.A2060G (p.Y687C) are potentially significant in squamous cell cervical cancer. Although DNA quantity limitations hindered validation, understanding their functional implications can provide valuable insights into SCC pathogenesis.

Variant in SRPX2: SRPX2 encodes a sushi repeat-containing protein essential for nerve development and cell growth, widely expressed in normal tissues (Schirwani *et al*., 2019). The p.A124V mutation may disrupt interactions between SRPX2 and extracellular matrix proteins, enhancing cell migration and invasion. Overexpression of SRPX2 has been linked to poor prognosis and advanced tumor stages in other cancers (Wu *et al*., 2020; Guo *et al*., 2023). Although its role in SCC is unclear, the p.A124V mutation could intensify SRPX2's oncogenic effects, making it a potential therapeutic target for controlling tumor invasiveness.

Variant in BRAT1: BRAT1 encodes BRCA1-associated ataxia-telangiectasia mutated activator 1, a protein involved in transcription regulation, DNA repair, cell cycle progression, and mitochondrial homeostasis (Engel *et al*., 2023; So and Ouchi, 2013). Dysregulation of BRAT1 has been linked to tumor aggressiveness and therapy resistance in SCC. The p.V171I mutation may further impair BRAT1’s tumor suppressor functions, contributing to genomic instability and marking it as a critical factor in aggressive SCC and a valuable biomarker for tumor progression.

Variant in KIF1B: KIF1B, a kinesin superfamily member, is crucial for cytoskeletal organization and intracellular vesicular transport (Al-Qahtani *et al*., 2017). While KIF1B mutations have been associated with various cancers, the precise mechanisms of its involvement in squamous cell cervical carcinoma are not fully understood. Investigating the impact of this mutation could shed light on the molecular mechanisms driving SCC development and progression, emphasizing the need for ongoing research into the genetic factors underlying this disease.

Variant in ALOX12B: ALOX12B encodes a lipoxygenase involved in arachidonic acid metabolism, a pathway critical for inflammation and cancer-associated tissue remodeling (Mashima and Okuyama, 2015). The p.Y687C mutation may enhance cancer-associated fibrosis and tissue remodeling, contributing to poor prognosis. Overexpression of ALOX12B has already been associated with SCC progression, and this mutation might intensify its oncogenic potential. Mutations in ALOX12B have also been identified in lung (Shen *et al*., 2009) and breast cancer (Lee *et al*., 2009). No mutational studies have been reported for ALOX12B associated with squamous cell cervical carcinoma.

Although these mutations have not been extensively validated due to DNA quantity, they are likely critical molecular mechanisms driving the progression of SCC, such as enhanced cell migration and invasion through SRPX2, disruption of transcription regulation and DNA repair by BRAT1, and alteration of inflammation and eicosanoid metabolism through ALOX12B.

Further validation of these mutations is required, followed by investigation into their functional role in SCC development. WES data confirmed by Sanger sequencing showed critical variants that had potential as either biomarkers or therapeutic targets. These could be used to design new personalized treatments based on the specific mutations causing SCC in individual patients.