UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

ANTONIO CARLOS FACCIOLO FILHO

IDENTIFICATION OF MUTATIONS POTENTIALLY ASSOCIATED WITH CHEMORADIOTHERAPY TREATMENT RESPONSE IN CERVICAL CANCER

PATOS DE MINAS-MG
JUNHO DE 2023

ANTONIO CARLOS FACCIOLO FILHO

IDENTIFICATION OF MUTATIONS POTENTIALLY ASSOCIATED WITH CHEMORADIOTHERAPY TREATMENT RESPONSE IN CERVICAL CANCER

Artigo Científico apresentado ao Instituto de Biotecnologia da Universidade Federal de Uberlândia como requisito final para a obtenção do título de Bacharel em Biotecnologia.

Prof. Dr. Matheus de Souza Gomes

PATOS DE MINAS-MG

JUNHO DE 2023

ANTONIO CARLOS FACCIOLO FILHO

Identification of Mutations Potentially Associated with Chemoradiotherapy Treatment Response in Cervical Cancer

Artigo Científico apresentado ao Instituto de Biotecnologia da Universidade Federal de Uberlândia como requisito final para a obtenção do título de Bacharel em Biotecnologia.

Banca examinadora:

Dr. Matheus de Souza Gomes - UFU
Presidente

Dra. Carolina Pereira de Souza Melo – IMP
Membro

Me. Carlos Bruno Araújo - UFU
Membro

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa que se encontra no Sistema Eletrônico de Informações (SEI) da Universidade Federal de Uberlândia.

Patos de Minas – MG, 23 de junho de 2023.

DEDICATÓRIA

Dedico este trabalho aos meus pais, minha família, amigos, aos mestres e mentores que me influenciaram, a todos que depositaram sua confiança em mim, bem àqueles que poderão fazer uso do mesmo e assim, juntos, promovermos mais ciência.

“We can judge our progress by the courage of our questions and the depth of our answers, our willingness to embrace what is true rather than what feels good”.

- Carl Sagan

AGRADECIMENTOS

Agradeço primeiramente meus pais, por todo o amor e apoio que recebi em toda minha vida. Com eles compartilho o crédito de todas as minhas conquistas, que sempre me incentivaram, prezaram pelos meus estudos e me deram as condições para seguir esse caminho. As oportunidades que tive na vida são fruto de suas ações.

A todo o Laboratório de Bioinformática e Análises Moleculares (LBAM), agradeço pelo acolhimento e auxílio que recebi durante minha iniciação no mundo acadêmico-científico, em especial ao meu orientador Dr. Matheus Gomes, por todo o incentivo e oportunidades que proporcionou.

Agradeço o grupo de pesquisa do Instituto Mario Penna pela confiança que tiveram em mim, e pela oportunidade de trabalhar nesse projeto, além de todo o apoio que recebi durante nossa colaboração.

Aos meus colegas de faculdade, sou grato pelo tempo que passamos juntos durante a graduação. Tanto pelos momentos bons quanto os de desespero, virando a noite estudando juntos ou terminando trabalhos em grupo. Aqui fiz amizades incríveis que levarei para o resto da vida.

Identification of Mutations Potentially Associated with Chemoradiotherapy Treatment Response in Cervical Cancer

Antonio Carlos Facciolo Filho, Tamires Caixeta Alves, Angelo Borges de Melo Neto, Laurence Rodrigues do Amaral, Pedro Luiz Lima Bertarini, Matheus de Souza Gomes

Laboratory of Bioinformatics and Molecular Analysis, Federal University of Uberlândia, Patos de Minas, MG, Brazil

Corresponding Author: Matheus de Souza Gomes, Federal University of Uberlandia, Institute of Genetics and Biochemistry, Laboratory of Bioinformatics and Molecular Analysis, Getúlio Vargas Avenue, 230, Patos de Minas, MG, Brazil. Phone: +55 34 38233714 – matheusgomes@ufu.br.

ABSTRACT

Background

Chemoradiotherapy resistance can significantly influence treatment outcomes in cervical cancer patients. Somatic variants play a crucial role in tumor development and treatment response. This study aimed to identify potentially clinically significant variants associated with chemoradiotherapy response using RNA-seq data analysis.

Methods

RNA-seq data from 31 cervical cancer patients were subjected to variant calling and subsequent analysis using a machine learning decision tree algorithm. Variants within coding genes were selected based on their potential to segregate non-responder and

responder groups. The selected variants were further assessed using variant impact prediction tools.

Results

We identified and investigated variants that exhibited potential for distinguishing between non-responder and responder patients. Comprehensive analysis using variant impact prediction tools indicated that these variants may have deleterious effects on gene function. The genes harboring these variants were found to be involved in crucial cellular processes associated with cancer progression, including cell cycle regulation, proliferation, and notably, drug resistance.

Conclusions

Our findings suggest that specific variants, namely MUC6(NM_005961.3):c.5992C>A, NAP1L1(NM_004537.7):c.1064_1065delinsGG, PABPC1(NM_002568.4):c.1255C>T, and NLRP4(NM_134444.5):c.1488_1489insT, have the potential to serve as indicators of poor chemoradiotherapy response in cervical cancer patients.

Keywords

Cervical Cancer, RNA-seq, Chemoradiotherapy, Somatic Mutations, MUC6, NAP1L1, PABPC1, NLRP4.

Background

Cervical cancer (CC) continues to pose a significant global health burden, ranking as the fourth most prevalent cancer among women worldwide [1]. Locally advanced CC is typically treated with chemoradiotherapy (CRT), which is considered the standard therapeutic approach [2, 3]. CRT has shown a 5-year survival rate of approximately 70% and is associated with low rates of local and distant recurrences [4]. However, the efficacy

of CRT is closely intertwined with the development of radio-resistance and drug resistance, both of which are influenced by multiple cellular mechanisms, including proliferation, apoptosis, and cell cycle regulation [5].

Somatic mutations play a pivotal role in the development and progression of cancer by disrupting the cell cycle, promoting tumor growth, and contributing to treatment resistance [6]. Oncogenes and tumor suppressor genes are essential for maintaining critical cellular processes, including DNA repair, apoptosis, and cell cycle control [7]. Mutations in these genes can directly impair these processes, thereby promoting tumor formation. Additionally, somatic mutations can confer resistance to cancer treatments through various mechanisms. Mutations that affect drug targets or the drug metabolism pathway can render tumor cells resistant to chemotherapy, while the dysregulation of DNA repair pathways can impede treatment strategies reliant on inducing DNA damage [5].

The introduction of next-generation sequencing (NGS) has revolutionized cancer research and clinical practice, particularly in the field of RNA sequencing (RNA-seq). RNA-seq has emerged as a highly effective diagnostic tool, allowing for a comprehensive analysis of the transcriptome. By profiling gene expression patterns, RNA-seq provides valuable insights into the molecular landscape of cancer and aids in the identification of genetic biomarkers [8].

In addition to RNA-seq, computational tools have significantly enhanced the identification and analysis of genetic biomarkers associated with cancer. Machine learning algorithms, for example, can be leveraged to construct predictive models for classifying different cancer subtypes or predicting treatment response [9]. These

algorithms can integrate diverse genomic data, such as gene expression profiles, somatic mutations, and clinical variables, to generate robust and accurate predictions. Such approaches hold great promise in improving cancer diagnosis, treatment selection, and personalized therapy [9].

The integration of NGS data and computational tools into clinical decision-making processes offers significant potential. Genetic biomarkers identified through NGS can provide valuable information for treatment selection, prognosis assessment, and the development of personalized therapies [10]. By analyzing the genomic landscape of a patient's tumor, including somatic mutations, gene expression patterns, and other genomic alterations, NGS allows for a more precise and individualized approach to cancer treatment. To facilitate the translation of NGS data into clinical practice, concerted efforts are underway to establish standardized pipelines and databases. Initiatives like The Cancer Genome Atlas (TCGA) and the International Genome Consortium (ICGC) aim to compile comprehensive genomic data from various cancer types and make it accessible to researchers and clinicians [11]. These resources provide a wealth of information for analyzing and interpreting NGS data, enabling researchers and clinicians to make more informed decisions in the context of cancer diagnosis and treatment. Overall, the integration of NGS data and computational tools, along with the establishment of standardized pipelines and databases, holds great promise for advancing precision medicine and improving patient outcomes in cancer care.

The primary objective of this study was to gain deeper insights into the molecular landscape of CC and identify somatic mutations that hold potential as genetic biomarkers regarding CRT response. This was achieved through the analysis of RNA-seq data obtained from CC patients who underwent CRT. The discovery and characterization of

these specific mutations have the capacity to significantly augment the clinical decision-making process, facilitating the selection of personalized treatment strategies and ultimately leading to improved patient outcomes.

Materials and methods

Patients Recruitment and Samples Selection

A total of 31 CC patients were enrolled in this study at the Mario Penna Institute in Belo Horizonte, Brazil, between August 2017 and May 2019. The study protocol received ethical approval from the local Institutional Review Board (Approval No. 1.583.784). Inclusion criteria encompassed patients with no prior history of cancer or immune diseases, a histopathological diagnosis of cervical adenocarcinoma or squamous cell carcinoma, and a cervical cancer diagnosis classified as stage II or III according to the FIGO classification. Written informed consent was obtained from all participants. After diagnosis and biopsy, patients underwent CRT, and comprehensive clinical data were collected and analyzed. Patient classification was based on two primary criteria: treatment response (R) and non-response (NR). The determination of treatment response for each patient involved clinical examinations, pathological analyses, and imaging evaluations. Patients were categorized as R if no cervical neoplastic lesions were detected within 8 months post-chemoradiation. Conversely, patients were classified as NR if cervical neoplastic lesions persisted, including cases of partial response, tumor progression, or stable disease, up to 8 months after treatment.

Fluorescence-Activated Cell Sorting (FACS) and RNA sequencing

Tumor cell samples obtained from cervical biopsy underwent fluorescence-activated cell sorting (FACS) in order to differentiate cervical cancer stem-like cells (CCSCs) and non-stem cervical cancer cells (NSCCs) from a complex mixture of tumor cells based on their

light scatter and fluorescent staining profile. NSCCs were selected and processed using the SMART-Seq v4 Ultra-low Input RNA Sequencing Kit (Takara Bio USA, CA) following the manufacturer's instructions for cDNA synthesis. Subsequently, the RNA sequencing library was prepared using the Nextera® XT Library Prep Kit (Illumina™) along with the Nextera® XT Index Kit V2 Set A (Illumina™), as per the manufacturer's instructions. Detailed descriptions of these procedures were previously provided by our research group [12].

Transcriptome analysis

Our research group previously conducted the analysis of CCSCs [12]. For the analysis of NSCCs, the quality of the sequencing data was assessed using the FastQC tool [13]. Reads with low quality (Phred score ≤ 25) were filtered out, and adapters were trimmed using the Cutadapt tool [14]. After trimming, the remaining sequences underwent another round of quality assessment using FastQC to ensure the desired quality standards. Subsequently, the sequences were mapped and aligned to the human genome (Homo sapiens GRCh38) using the STAR software [15].

Variant calling, somatic filtering and variant data analysis

Variant calling was performed using BCFtools 1.16 [16]. In order to select high confidence variants, filtering was performed by requiring a minimum Phred quality score 30 and a read depth of 100 at the position. In order to exclude potentially germline variants, common population single nucleotide variations (SNVs) present in the dbSNP database were removed from the pool of detected variants.

The remaining variants were annotated using Ensembl Variant Effect Predictor (VEP) [17]. Vep includes population allele frequency from databases such as 1000 Genomes

Project and Aggregation Database (gnomAD). Variants with population frequencies of 1% or higher were filtered out to further remove germline variants. Multiple impact prediction tools available on VEP were employed, including SIFT, PolyPhen-2, CADD and LOUEF. SIFT Indel [18] and SpliceAI-visual [19] were employed to assess the effects of indels and splice region variants respectively.

Vcf2maf [20] was used to generate MAF files from the final pool of variants for further analysis. Maftools and GenVisR R packages were used for data processing and visualization.

Association of variants to treatment response

Using the full training approach, decision tree analysis was conducted on somatic variants of coding genes. The analysis was performed using the WEKA software (Waikato Environment for Knowledge Analysis, version 3.6.11, University of Waikato, New Zealand) [21].

Results

Cohort Clinical Description

Samples from 31 women diagnosed with CC were subjected to NGS analysis. Patient clinical information and histopathological analysis are summarized in Table 1. Overall response was evaluated 8 months after ending treatment, 21 patients were classified as Responders (R) and 10 as Non-responders (NR).

Table 1: Clinicopathological characteristics and chemoradiotherapy response.

Patient characteristics		N	Value
Age	Years	31	52.3±16.8
Diagnostic	SCC	30	96.8
	ICA	1	3.2
	IIA	1	3.2

FIGO Stage	IIB	15	48.4
	IIIB	15	48.4
Histological grade	II	13	41.9
	III	13	41.9
	NA	5	16.1
Parametrial involvement	Free	1	3.2
	Unilaterally	7	22.6
	Bilaterally	22	71.0
	NA	1	3.2
Vaginal involvement	Yes	28	90.3
	No	1	3.2
	NA	2	6.5
Lymph node status	NX	8	25.8
	N0	9	29.0
	N1	14	45.2
Metastasis	MX	10	32.3
	M0	9	29.0
	M1	12	38.7
Tumor size (cm)	<4	1	3.2
	≥4	29	93.5
	NA	1	3.2
Overall response	R	21	67.7
	NR	10	32.3

The values represent the mean + standard error or %. SCC, Squamous Cell Carcinoma; ADC, Adenocarcinoma; NR, Non-Responders, R, Responders; NX, Regional lymph nodes cannot be assessed; N0, No regional lymph node metastasis; N1, Regional lymph node metastasis; MX, Distant metastasis cannot be assessed; M0, No distant metastasis; M1, Distant metastasis. NA, Not Available.

Cohort Mutation Landscape

In total, 34,617 unique variants were found within the cohort, comprised of 30,851 (89.12%) base substitutions and 3,766 (10.88%) indels. 24,432 (79.58%) variants were classified as novel, while 10,185 (29.42%) were previously annotated in variant databases. Only 522 (1.51%) variants were located in coding genes, the vast majority of the identified variants were located in non-coding transcripts of varying types, with long non-coding RNA being the most predominant. Only coding gene variants were selected for further analysis in this study, and a comprehensive statistical summary of this data is available in Supplementary Figure 1.

In the top 10 mutated genes, we observed the recurrence of genes MUC3A, MUC6, HLA-DRB1 and GOLGA6L10 [22,23,24], which were previously reported in CC, and novel genes SPATA22, FER1L5, SPDYE1, ACTR3B, ADAMTSL3 and TRIM45. Figure 1 displays an overview of the most mutated genes per patient, including the presence of variants in non-coding regions.

Variants associated to non-response

A total of 522 coding gene variants were included in the decision tree analysis. The decision tree successfully segregated responders and non-responder patients based on four junction points, each corresponding to a different variant (Figure 2).

These variants were labeled as NLRP4(NM_134444.5):c.1488_1489insT MUC6(NM_005961.3):c.5992C>A, PABPC1(NM_002568.4):c.1255C>T, and NAP1L1(NM_004537.7):c.1064A>G. It is worth noting that although the decision tree only reported one variant per gene, further investigation showed the co-occurrence of two additional SNVs in the NAP1L1 gene. One was labeled as NAP1L1(NM_004537.7):c.1062T>C, while the other was adjacent to NAP1L1(NM_004537.7):c.1064A>G, revealing them to be a deletion-insertion (delins) which we labeled NAP1L1(NM_004537.7):c.1064_1065delinsGG. The additional variants have been included in our analysis in order to provide a more comprehensive overview of the genetic variations within that gene. A comprehensive summary report of the variants, including known functions of affected genes, is available in Table 2. The deleterious potential of missense variants was assessed using scores from different variant impact prediction tools available on VEP (Table 3). Prediction tools SIFT Indel and SpliceAI were employed on suitable variants.

Table 2: Summary of variants identified as potential indicators of poor CRT response

Variant	Consequence	Patients	Database IDs	Population frequency	Gene functions
MUC6(NM_005961.3):c.5992C>A	Missense	17, 48, 123	rs756062369 COSV70132617 COSV70143795	0.00000798	Extracellular matrix structural constituent, maintenance of gastrointestinal epithelium, β-catenin degradation
NAP1L1(NM_004537.7):c.1064_1065delinsGG	Missense	63, 97	novel	Not found	Histone chaperone, nucleossome assembly, nucleotide excision repair, mitotic DNA double strand break repair, cell cycle progression
NAP1L1(NM_004537.7):c.1062T>C	Synonymous & Splice region		novel	Not found	
PABPC1(ENST00000318607.10):c.1255C>T	Missense	17, 63	rs766099049	0.00428	mRNA Translation, mRNA Decay, miRNA-mediated repression, NMD, L1 Retrotransposition, mRNA localization, local translation
NLRP4(NM_134444.5):c.1488_1489insT	Frameshift insertion	103, 121	novel	Not found	Inflamatory response, Regulation of inflamatory response, NF-kB inhibition

Table 3: Missense variants and respective assigned scores from variant impact prediction tools.

Variant	SIFT	PolyPhen-2	BLOSUM62	CADD	LOEUF	PhyloP100way
MUC6(NM_005961.3):c.5992C>A	tolerated(0.12)	probably_damaging (0.912)	-1	8.197	0.97	-1.725
NAP1L1(NM_004537.7):c.1064_1065delinsGG	tolerated(0.11)	probably_damaging (0.999)	-1	#	#	8.517*
						5.574**
PABPC1(NM_002568.4):c.1255C>T	deleterious(0.03)	possibly_damaging (0.865)	-3	26.3	0.27	6.313

Values considered to deem each parameter as evidence of deleteriousness were: SIFT < 0.05; PolyPhen > 0.85; CADD ≥ 20; LOUEF < 0.35; BLOSUM62 < 0; phyloP100way ≥ 8.0. Conservation scores in NAP1L1(NM_004537.7):c.1064_1065delinsGG were calculated for each individual base affected at postions 1065* and 1064**.

MUC6(NM_005961.3):c.5992C>A

MUC6 emerged as a highly mutated gene in multiple patients within the cohort (Supplementary figure 2A). Six unique variants were identified in this gene. Apart from the reported variant, patients 48 and 123 harbored two other missense variants in MUC6, and patient 17 harbored three. The reported variant showed evidences of having functional impact on the protein, but the consensus from multiple prediction tools and

conservation analyses would provide a more comprehensive assessment of its significance.

NAP1L1(NM_004537.7):c.1064_1065delinsGG and NAP1L1(NM_004537.7):c.1062T>C

NAP1L1 was another highly mutated gene in our cohort (Supplementary Figure 2B), displaying nine unique variants. On patient 63, the reported variants were the only ones present in NAP1L1, while patient 97 exhibited seven other variants within the gene. The delins variant was not evaluated with as much scrutiny as the others due to some prediction tools not supporting delins data input, nonetheless, amino acid substitution and conservation scores indicated it as potentially damaging to protein structure. NAP1L1(NM_004537.7):c.1062T>C is a synonymous mutation located on the splice region of an exon, however, results from SpliceAI indicate it has no effect on splicing (Supplementary Figure 3).

PABPC1(NM_002568.4):c.1255C>T

PABPC1 exhibited a moderate level of mutation in a few samples within the cohort (Supplementary Figure 2C). The reported variant achieved a consensus among most prediction scores as potentially damaging to the protein product. Interestingly, patient 63 harbored a nonsense mutation (PABPC1(NM_002568.4):c.874C>T) that also appeared in patient 108, a responder. The presence of a nonsense mutation in PABPC1 suggests potential disruption of protein function and highlights its possible involvement in treatment response or disease progression.

NLRP4(NM_134444.5):c.1488_1489insT

In contrast to the other genes analyzed, NLRP4 did not exhibit any additional mutations beyond the one reported by the decision tree (Supplementary Figure 2D). The identified variant in NLRP4 is a frameshift insertion involving a single nucleotide. This variant has been classified as deleterious by the SIFT Indel prediction tool with a confidence score of 0.858. The frameshift insertion in NLRP4 has a significant impact on the protein structure and function. It results in a disruption of almost half of the amino acid sequence, which can lead to profound changes in protein folding, interaction with other molecules, and overall protein function. Furthermore, according to SIFT Indel the variant is predicted to trigger nonsense-mediated decay, a cellular mechanism that targets mRNA with premature termination codons for degradation. These findings suggest that the frameshift insertion in NLRP4 is likely to have significant functional consequences and may contribute to poor CRT response or CC characteristics within the cohort.

Discussion

With the advancements in next-generation sequencing technologies, RNA-seq has emerged as a popular and cost-effective tool for studying gene expression and providing valuable insights into the phenotype. Despite its widespread use, accurately identifying variants from RNA-seq data can be challenging due to its inherent limitations, such as lower accuracy compared to DNA sequencing. However, several studies have successfully employed rigorous filtering criteria and other strategies to obtain reliable variant calling results from RNA-seq data [25, 26]. Considering these factors, in our study, we performed variant calling, followed by stringent filtering and processing steps, in order to map out the mutational landscape of the cohort. This analysis served two primary purposes. Firstly, it aimed to provide a comprehensive overview of the mutational profile of each patient within the cohort, shedding light on the specific genetic

alterations present in CC. Secondly, by comparing our findings to previously validated CC variant data from other studies, we aimed to assess the accuracy and reliability of our variant calling pipeline. This approach allowed us to gain valuable insights into the mutational landscape of CC in our cohort and evaluate the robustness of our variant calling methodology. By corroborating our findings with validated data from other studies, we aimed to ensure the accuracy and reliability of our results, thus strengthening the scientific foundation of our research and contributing to the growing body of knowledge in the field of CC genomics.

In our cohort, we observed the presence of several genes that belong to families frequently associated with CC development, including HLA genes and mucins. The HLA region genes play a crucial role in the immune response against tumor antigens, highlighting their significance in CC [27]. Specifically, HLA-A and HLA-B have been identified as driver genes in CC [22], and mutations in HLA-DRB1 have been associated with the development of this type of cancer [28]. Mucins, on the other hand, are known to participate in inflammatory responses, and their dysregulation is often associated with tumor progression. However, it is important to note that some mucins have been reported to act as tumor suppressors in various cancers, such as MUC2 and MUC6 [29]. In our cohort, we identified frequent mutations in MUC6 and MUC12, which have been previously reported in cervical carcinomas [24]. Additionally, MUC3A has displayed tumorigenesis potential in cervical adenocarcinoma [30]. Furthermore, we also detected other genes commonly found to be associated with CC. These findings support the involvement of these genes in CC development and emphasize their potential relevance as therapeutic targets or biomarkers for this disease. By identifying these genes within our cohort, we contribute to the growing understanding of the molecular landscape of CC

and provide valuable insights into the potential mechanisms driving its development and progression. Further investigations are warranted to explore the functional implications of these genetic alterations and their clinical significance in CC.

Through the implementation of a decision tree-based machine-learning algorithm, we successfully identified specific variants that have the potential to impact response in CRT treatment. To ascertain the significance of these variants, we conducted a comprehensive analysis considering various factors. Firstly, we assessed the potential deleterious effects of the variants by examining factors such as amino acid substitution, nucleotide conservation, and population frequencies. These criteria allowed us to gauge the likelihood of these variants causing functional alterations at the protein level and their relative rarity within the population. Furthermore, we delved into the functional characteristics of the affected genes to ascertain their involvement in cellular mechanisms associated with CRT response. This evaluation provided insights into the potential interactions between these genes and the relevant molecular pathways underlying treatment outcomes. By integrating these analyses, we aimed to establish a comprehensive understanding of the potential impact of the identified variants on CRT response. This information is crucial for guiding further investigations and developing personalized treatment approaches in order to enhance patient outcomes and optimize therapy strategies for CC.

MUC6, although commonly used as an immunohistochemical marker to differentiate subtypes of endocervical carcinomas [31], its specific role in CC remains relatively unexplored. However, its tumor suppressor role has been well-documented in other cancer types such as pancreatic cancer, renal cancer, gastric cancer, and colorectal cancer [32-35]. Studies have shown that MUC6 acts as a tumor suppressor by inducing β-catenin

degradation through the autophagy pathway [33]. β-catenin plays a crucial role in regulating cell growth and intracellular adhesion in epithelial cells. In CC therapy, targeting β-catenin has gained considerable attention due to its involvement in cell proliferation and its contribution to chemotherapy and radiotherapy resistance [36]. Understanding the functional significance of MUC6 in CC could provide valuable insights into its potential role as a therapeutic target or prognostic marker. By elucidating the molecular mechanisms by which MUC6 influences cellular processes and affects the behavior of CC cells, we can potentially identify new avenues for personalized treatment approaches and develop strategies to overcome therapeutic resistance.

NAP1L1, a nucleosome assembly protein, plays a crucial role in a diverse range of cellular functions. It has been implicated in regulating DNA replication, chromatin formation, cell cycle progression, and proliferation, among others [37]. Although its association with CC has not yet been established, multiple studies have demonstrated its involvement as a promoter of tumor cell proliferation in various cancer types [38-40]. Furthermore, there is evidence suggesting that NAP1L1 may contribute to chemotherapy resistance [41]. Understanding the precise role of NAP1L1 in CC could shed light on its potential as a therapeutic target and provide insights into mechanisms underlying tumor progression and treatment response. Given the multifaceted nature of NAP1L1's cellular functions, further investigation is warranted to elucidate its specific involvement in CC development and progression. This knowledge could pave the way for the development of targeted therapies aimed at modulating NAP1L1's activity and potentially overcoming chemoresistance in CC treatment.

PABPC1, a highly conserved poly A-binding protein, is involved in regulating various aspects of mRNA metabolism, thereby connecting it to a wide range of cellular processes

[42]. In the context of CC, PABPC1 has been found to play a role in the inhibition of the Wnt/β-catenin pathway. A study uncovered the involvement of the long non-coding RNA (lncRNA) PDHB-AS as a tumor suppressor in CC. PDHB-AS was found to recruit PABPC1, leading to the inhibition of nuclear import of β-catenin and subsequent inactivation of the Wnt/β-catenin pathway [43]. The same study revealed that PDHB-AS is downregulated in CC, suggesting a potential role in disease progression. Additionally, the overexpression of PDHB-AS was shown to reduce cisplatin resistance, indicating its potential as a therapeutic target to enhance treatment response in CC [43]. The findings highlight the significance of PABPC1 in modulating the Wnt/β-catenin pathway and its potential implication in CC development and treatment resistance.

NLRP4 is a gene involved in the regulation of inflammatory response and functions as a negative regulator of pyroptosis, a form of programmed cell death triggered by pro-inflammatory signals [44]. Pyroptosis, despite causing cell death, has a dual role in CC progression due to its impact on the tumor microenvironment. In early stages of CC, pyroptosis acts as a suppressor by eliminating tumor cells. However, in advanced stages, tumor cells can exploit the tumor microenvironment, including the release of inflammasomes and pro-inflammatory molecules associated with pyroptosis, to promote invasion and metastasis [45, 46]. The intricate relationship between NLRP4, pyroptosis, and CC suggests the potential involvement of inflammatory pathways in tumor progression and metastasis.

The decision tree analysis conducted in our study identified specific variants that demonstrated potential for distinguishing between non-responder and responder groups. The genes where these variants are located have been extensively associated with various types of cancer, and their mechanisms have been well-documented. Based on the

predictions of variant impact prediction tools, we found evidence that four out of these mutations are likely to have deleterious effects on their associated genes: MUC6(NM_005961.3):c.5992C>A, NAP1L1(NM_004537.7):c.1064_1065delinsGG, PABPC1(NM_002568.4):c.1255C>T and NLRP4(NM_134444.5):c.1488_1489insT. These variants suggest a potential role in hindering CRT response. However, it is important to note that further research is necessary to validate and confirm these findings. Expanding the sample size and including a larger cohort of patients would enable us to determine the exclusivity of these variants to non-responder patients. Such data could then be utilized in the decision-making process for selecting appropriate treatment strategies for CC patients. Conducting additional investigations and studies will enhance our understanding of the clinical significance of these variants and their potential implications for treatment outcomes in CC.

Conclusions

CC is one of the most common gynecological cancers and has a high lethality rate, therefore, choice of treatment is a crucial decision in ensuring the health and well-being of CC patients. Despite being the recommended form of treatment in locally advanced CC, poor CRT response remains an obstacle to be overcome. Our data suggests that the somatic variants MUC6(NM_005961.3):c.5992C>A, NAP1L1(NM_004537.7):c.1064_1065delinsGG, PABPC1(NM_002568.4):c.1255C>T and NLRP4(NM_134444.5):c.1488_1489insT may be potential indicators of non-response to CRT. Further research is required in order to validate these findings and to explore their biological and clinical implications. Precise understanding of the interactions between these variants and CC may contribute to guiding treatment selection for CC patients.

List of abbreviations

CC: Cervical cancer

CRT: Chemoradiotherapy

NGS: Next generation sequencing

RNA-seq: RNA sequencing

R: Responder

NR: Non-responder

CCSC: Cervical cancer stem-like cells

NSCCs: Non-stem cervical cancer cells

VEP: Variant Effect Predictor

SNV: Single nucleotide variation

delins: Deletion-insertion

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA A Cancer J Clin. 2021 May;71(3):209–49.

2. Monk BJ, Tewari KS, Koh WJ. Multimodality Therapy for Locally Advanced Cervical Carcinoma: State of the Art and Future Directions. JCO. 2007 Jul 10;25(20):2952–65.

3. Gaffney DK, Erickson-Wittmann BA, Jhingran A, Mayr NA, Puthawala AA, Moore D, et al. ACR Appropriateness Criteria® on Advanced Cervical Cancer Expert Panel on Radiation Oncology—Gynecology. International Journal of Radiation Oncology*Biology*Physics. 2011 Nov;81(3):609–14.

4. Green JA, Kirwan JJ, Tierney J, Vale CL, Symonds PR, Fresco LL, et al. Concomitant chemotherapy and radiation therapy for cancer of the uterine cervix. Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group, editor. Cochrane Database of Systematic Reviews [Internet]. 2005 Jul 20 [cited 2023 Jun 6]; Available from: https://doi.wiley.com/10.1002/14651858.CD002225.pub2

5. Liu Y, Zheng C, Huang Y, He M, Xu WW, Li B. Molecular mechanisms of chemo- and radiotherapy resistance and the potential implications for cancer treatment. MedComm. 2021 Sep;2(3):315–40.

6. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009 Mar;9(3):153–66.

7. Lane DP. p53, guardian of the genome. Nature. 1992 Jul;358(6381):15–6.

8. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet. 2019 Apr 5;10:317.

9. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2007 Feb 11;2:59–77.

10. Kirschner MB, Cheng YY, Badrian B, Kao SC, Creaney J, Edelman JJB, et al. Increased Circulating miR-625-3p: A Potential Biomarker for Patients With Malignant Pleural Mesothelioma. Journal of Thoracic Oncology. 2012 Jul;7(7):1184–91.

11. The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113–20.

12. Zuccherato LW, Machado CMT, Magalhães WCS, Martins PR, Campos LS, Braga LC, et al. Cervical Cancer Stem-Like Cell Transcriptome Profiles Predict Response to Chemoradiotherapy. Front Oncol. 2021 May 7;11:639339.

13. Andrews S. FastQC: a quality control tool for high throughput sequence data. (2010) Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

14. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1), 10 (2011).

15. Dobin A, Davis CA, Schlesinger F et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1), 15–21 (2013).

16. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov 1;27(21):2987–93.

17. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Dec;17(1):122.

18. Hu J, Ng PC. SIFT Indel: Predictions for the Functional Effects of Amino Acid Insertions/Deletions in Proteins. Toland AE, editor. PLoS ONE. 2013 Oct 23;8(10):e77940.

19. De Sainte Agathe JM, Filser M, Isidor B, Besnard T, Gueguen P, Perrin A, et al. SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. Hum Genomics. 2023 Feb 10;17(1):7.

20. Cyriac Kandoth. mskcc/vcf2maf: vcf2maf v1.6.19. (2020). doi:10.5281/zenodo.593251

21. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor Newsl. 2009 Nov 16;11(1):10–8.

22. The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017 Mar;543(7645):378–84.

23. Chung TKH, Van Hummelen P, Chan PKS, Cheung TH, Yim SF, Yu MY, et al. Genomic aberrations in cervical adenocarcinomas in Hong Kong Chinese women: Cervical adenocarcinoma genomics. Int J Cancer. 2015 Aug 15;137(4):776–83.

24. Dash S, Kinney NA, Varghese RT, Garner HR, Feng W chun, Anandakrishnan R. Differentiating between cancer and normal tissue samples using multi-hit combinations of genetic mutations. Sci Rep. 2019 Jan 30;9(1):1005.

25. Jessen E, Liu Y, Davila J, Kocher JP, Wang C. Determining mutational burden and signature using RNA-seq from tumor-only samples. BMC Med Genomics. 2021 Dec;14(1):65.

26. Hashimoto S, Noguchi E, Bando H, Miyadera H, Morii W, Nakamura T, et al. Neoantigen prediction in human breast cancer using RNA sequencing data. Cancer Sci. 2021 Jan;112(1):465–75.

27. Pagliuca S, Gurnari C, Rubio MT, Visconte V, Lenz TL. Individual HLA heterogeneity and its implications for cellular immune evasion in cancer and beyond. Front Immunol. 2022 Sep 5;13:944872.

28. Kamiza AB, Kamiza S, Mathew CG. HLA-DRB1 alleles and cervical cancer: A meta-analysis of 36 case-control studies. Cancer Epidemiology. 2020 Aug;67:101748.

29. Kufe DW. Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer. 2009 Dec;9(12):874–85.

30. Guo M, Li G, Jiang X, Lv S. A tumor of composite cervical adenocarcinoma with MUCIN gene mutation：a case report [Internet]. In Review; 2021 Jul [cited 2023 Jun 14]. Available from: https://www.researchsquare.com/article/rs-394664/v2

31. Fulgione C, Raffone A, Travaglino A, Arciuolo D, Santoro A, Cianfrini F, et al. Diagnostic accuracy of HIK1083 and MUC6 as immunohistochemical markers of

endocervical gastric-type adenocarcinoma: A systematic review and meta-analysis. Pathology - Research and Practice. 2023 Jan;241:154261.

32. Yuki A, Fujii C, Yamanoi K, Matoba H, Harumiya S, Kawakubo M, et al.

33. Glycosylation of MUC6 by α1,4-linked N -acetylglucosamine enhances suppression of pancreatic cancer malignancy. Cancer Science. 2022 Feb;113(2):576–86.

34. Liu BH, Liu GB, Zhang BB, Shen J, Xie LL, Liu XQ, et al. Tumor Suppressive Role of MUC6 in Wilms Tumor via Autophagy-Dependent β-Catenin Degradation. Front Oncol. 2022 Apr 28;12:756117.

35. Betge J, Schneider NI, Harbaum L, Pollheimer MJ, Lindtner RA, Kornprat P, et al. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. Virchows Arch. 2016 Sep;469(3):255–65.

36. Shi D, Xi X xia. Regulation of MUC6 Methylation Correlates with Progression of Gastric Cancer. Yonsei Med J. 2021;62(11):1005.

37. Wang B, Li X, Liu L, Wang M. β-Catenin: oncogenic role and therapeutic target in cervical cancer. Biol Res. 2020 Dec;53(1):33.

38. Zlatanova J, Seebart C, Tomschik M. Nap1: taking a closer look at a juggler protein of extraordinary skills. FASEB j. 2007 May;21(7):1294–310.

39. Gan H, Xu X, Bai Y. Trametes robiniophila represses angiogenesis and tumor growth of lung cancer via strengthening let-7d-5p and targeting NAP1L1. Bioengineered. 2022 Mar 1;13(3):6698–710.

40. Xiaohua Zhu, Xie Y, Huang W, Chen Z, Guo S. NAP1L1 promotes tumor proliferation through HDGF/C-JUN signaling in ovarian cancer. BMC Cancer. 2022 Dec;22(1):339.

41. Zhai W, Ma J, Zhu R, Xu C, Zhang J, Chen Y, et al. MiR-532-5p suppresses renal cancer cell proliferation by disrupting the ETS1-mediated positive feedback loop with the KRAS-NAP1L1/P-ERK axis. Br J Cancer. 2018 Aug;119(5):591–604.

42. Qi Y, Wang M, Jiang Q. PABPC1——mRNA stability, protein translation and tumorigenesis. Front Oncol. 2022 Dec 1;12:1025291.

43. Chi C, Hou W, Zhang Y, Chen J, Shen Z, Chen Y, et al. PDHB-AS suppresses cervical cancer progression and cisplatin resistance via inhibition on Wnt/β-catenin pathway. Cell Death Dis. 2023 Feb 7;14(2):90.

44. Moon SW, Son HJ, Mo HY, Yoo NJ, Lee SH. Somatic Mutation of NLRP Genes in Gastric and Colonic Cancers. Pathol Oncol Res. 2021 Apr 16;27:607385.

45. Kolb R, Liu GH, Janowski AM, Sutterwala FS, Zhang W. Inflammasomes in cancer: a double-edged sword. Protein Cell. 2014 Jan;5(1):12–20.

46. Li K, Qiu J, Pan J, Pan JP. Pyroptosis and Its Role in Cervical Cancer. Cancers. 2022 Nov 23;14(23):5764.

Declarations

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by Mario Penna Institutional Review Board #1.583.784. The patients/participants provided their written informed consent to participate in this study.

Consent for publication

Not applicable.

Availability of data and materials

The data used in this article is available at the NCBI's BioProject site under the accession number: PRJNA705088.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Author's contributions

ACFF accomplished the bioinformatic analyses and drafted the manuscript; TCA and MSG participated in the study design and writing of the manuscript; ABMN contributed to the bioinformatics analyses; LRA and PLLB contributed to data processing *via* machine learning; MSG coordinated the project.

Acknowledgements

We thank all the patients for their contribution to this study. We also thank the research group from the Mario Penna Institute for obtaining and supplying the data utilized in this study

Normas da revista (BMC Medical Genomics)

Preparing your manuscript

The information below details the section headings that you should include in your manuscript and what information should be within each section.

Please note that your manuscript must include a 'Declarations' section including all of the subheadings (please see below for more information).

Title page

The title page should:

- present a title that includes, if appropriate, the study design e.g.:
    - "A versus B in the treatment of C: a randomized controlled trial", "X is a risk factor for Y: a case control study", "What is the impact of factor X on subject Y: A systematic review"
    - or for non-clinical or non-research studies a description of what the article reports
- list the full names and institutional addresses for all authors
    - if a collaboration group should be listed as an author, please list the Group name as an author. If you would like the names of the individual members of the Group to be searchable through their individual PubMed records, please include this information in the “Acknowledgements” section in accordance with the instructions below
    - Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript.
- indicate the corresponding author

Abstract

The Abstract should not exceed 350 words. Please minimize the use of abbreviations and do not cite references in the abstract. Reports of randomized controlled trials should follow the CONSORT extension for abstracts. The abstract must include the following separate sections:

- **Background:** the context and purpose of the study
- **Methods:** how the study was performed and statistical tests used
- **Results:** the main findings
- **Conclusions:** brief summary and potential implications
- **Trial registration:** If your article reports the results of a health care intervention on human participants, it must be registered in an appropriate registry and the

registration number and date of registration should be stated in this section. If it was not registered prospectively (before enrollment of the first participant), you should include the words 'retrospectively registered'. See our editorial policies for more information on trial registration

Keywords

Three to ten keywords representing the main content of the article.

Background

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary or its contribution to the field.

Methods

The methods section should include:

- the aim, design and setting of the study
- the characteristics of participants or description of materials
- a clear description of all processes, interventions and comparisons. Generic drug names should generally be used. When proprietary brands are used in research, include the brand names in parentheses
- the type of statistical analysis used, including a power calculation if appropriate

Results

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

Discussion

This section should discuss the implications of the findings in context of existing research and highlight limitations of the study.

Conclusions

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study reported.

List of abbreviations

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations should be provided.

Declarations

All manuscripts must contain the following sections under the heading 'Declarations':

- Ethics approval and consent to participate
- Consent for publication
- Availability of data and materials
- Competing interests
- Funding
- Authors' contributions
- Acknowledgements
- Authors' information (optional)

Please see below for details on the information to be included in these sections.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Ethics approval and consent to participate

Manuscripts reporting studies involving human participants, human data or human tissue must:

- include a statement on ethics approval and consent (even where the need for approval was waived)
- include the name of the ethics committee that approved the study and the committee's reference number if appropriate

Studies involving animals must include a statement on ethics approval and for experimental studies involving client-owned animals, authors must also include a statement on informed consent from the client or owner.

See our editorial policies for more information.

If your manuscript does not report on or involve the use of any animal or human data or tissue, please state "Not applicable" in this section.

Consent for publication

If your manuscript contains any individual person's data in any form (including any individual details, images or videos), consent for publication must be obtained from that person, or in the case of children, their parent or legal guardian. All presentations of case reports must have consent for publication.

You can use your institutional consent form or our consent form if you prefer. You should not send the form to us on submission, but we may request to see a copy at any stage (including after publication).

See our editorial policies for more information on consent for publication.

If your manuscript does not contain data from any individual person, please state "Not applicable" in this section.

Availability of data and materials

All manuscripts must include an 'Availability of data and materials' statement. Data availability statements should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported in the article. We recognise it is not always possible to share research data publicly, for instance when individual privacy could be compromised, and in such instances data availability should still be stated in the manuscript along with any conditions for access.

Authors are also encouraged to preserve search strings on searchRxiv https://searchrxiv.org/, an archive to support researchers to report, store and share their searches consistently and to enable them to review and re-use existing searches. searchRxiv enables researchers to obtain a digital object identifier (DOI) for their search, allowing it to be cited.

Data availability statements can take one of the following forms (or a combination of more than one if required for multiple datasets):

- The datasets generated and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS]
- The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article [and its supplementary information files].
- The datasets generated and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
- Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

- The data that support the findings of this study are available from [third party name] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [third party name].
- Not applicable. If your manuscript does not contain any data, please state 'Not applicable' in this section.

More examples of template data availability statements, which include examples of openly available and restricted access datasets, are available here.

BioMed Central strongly encourages the citation of any publicly available data on which the conclusions of the paper rely in the manuscript. Data citations should include a persistent identifier (such as a DOI) and should ideally be included in the reference list. Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style. Dataset identifiers including DOIs should be expressed as full URLs. For example:

Hao Z, AghaKouchak A, Nakhjiri N, Farahmand A. Global integrated drought monitoring and prediction system (GIDMaPS) data sets. figshare. 2014. http://dx.doi.org/10.6084/m9.figshare.853801

With the corresponding text in the Availability of data and materials statement:

The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].[Reference number]

If you wish to co-submit a data note describing your data to be published in *BMC Research Notes*, you can do so by visiting our submission portal. Data notes support open data and help authors to comply with funder policies on data sharing. Co-published data notes will be linked to the research article the data support (example).

Competing interests

All financial and non-financial competing interests must be declared in this section.

See our editorial policies for a full explanation of competing interests. If you are unsure whether you or any of your co-authors have a competing interest please contact the editorial office.

Please use the authors initials to refer to each authors' competing interests in this section.

If you do not have any competing interests, please state "The authors declare that they have no competing interests" in this section.

Funding

All sources of funding for the research reported should be declared. The role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

Authors' contributions

The individual contributions of authors to the manuscript should be specified in this section. Guidance and criteria for authorship can be found in our editorial policies.

Please use initials to refer to each author's contribution in this section, for example: "FC analyzed and interpreted the patient data regarding the hematological disease and the transplant. RH performed the histological examination of the kidney, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript."

Acknowledgements

Please acknowledge anyone who contributed towards the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements section.

See our editorial policies for a full explanation of acknowledgements and authorship criteria.

If you do not have anyone to acknowledge, please write "Not applicable" in this section.

Group authorship (for manuscripts involving a collaboration group): if you would like the names of the individual members of a collaboration Group to be searchable through their individual PubMed records, please ensure that the title of the collaboration Group is included on the title page and in the submission system and also include collaborating author names as the last paragraph of the “Acknowledgements” section. Please add authors in the format First Name, Middle initial(s) (optional), Last Name. You can add institution or country information for each author if you wish, but this should be consistent across all authors.

Please note that individual names may not be present in the PubMed record at the time a published article is initially included in PubMed as it takes PubMed additional time to code this information.

Authors' information

This section is optional.

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

References

Examples of the Vancouver reference style are shown below.

See our editorial policies for author guidance on good citation practice

Web links and URLs: All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database. http://tumor.informatics.jax.org/mtbwi/index.do. Accessed 20 May 2013. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

Example reference style:

Article within a journal

Smith JJ. The world of science. Am J Sci. 1999;36:234-5.

Article within a journal (no page numbers)

Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, et al. Meat consumption and mortality - results from the European Prospective Investigation into Cancer and Nutrition. BMC Medicine. 2013;11:63.

Article within a journal by DOI

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. Dig J Mol Med. 2000; doi:10.1007/s801090000086.

Article within a journal supplement

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. Blood 1979;59 Suppl 1:26-32.

Book chapter, or an article within a book

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH, Danielli JF, Jeon KW, editors. International review of cytology. London: Academic; 1980. p. 251-306.

OnlineFirst chapter in a series (without a volume designation but with a DOI)

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. Top Curr Chem. 2007. doi:10.1007/128_2006_108.

Complete book, authored

Blenkinsopp A, Paxton P. Symptoms in the pharmacy: a guide to the management of common illness. 3rd ed. Oxford: Blackwell Science; 1998.

Online document

Doe J. Title of subordinate document. In: The dictionary of substances and their effects. Royal Society of Chemistry. 1999. http://www.rsc.org/dose/title of subordinate document. Accessed 15 Jan 1999.

Online database

Healthwise Knowledgebase. US Pharmacopeia, Rockville. 1998. http://www.healthwise.org. Accessed 21 Sept 1998.

Supplementary material/private homepage

Doe J. Title of supplementary material. 2000. http://www.privatehomepage.com. Accessed 22 Feb 2000.

University site

Doe, J: Title of preprint. http://www.uni-heidelberg.de/mydata.html (1999). Accessed 25 Dec 1999.

FTP site

Doe, J: Trivial HTTP, RFC2169. ftp://ftp.isi.edu/in-notes/rfc2169.txt (1999). Accessed 12 Nov 1999.

Organization site

ISSN International Centre: The ISSN register. http://www.issn.org (2006). Accessed 20 Feb 2007.

Dataset with persistent identifier

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome data from sweet and grain sorghum (Sorghum bicolor). GigaScience Database. 2011. http://dx.doi.org/10.5524/100012.

Figures, tables and additional files

See General formatting guidelines for information on how to format figures, tables and additional files.

Anexos

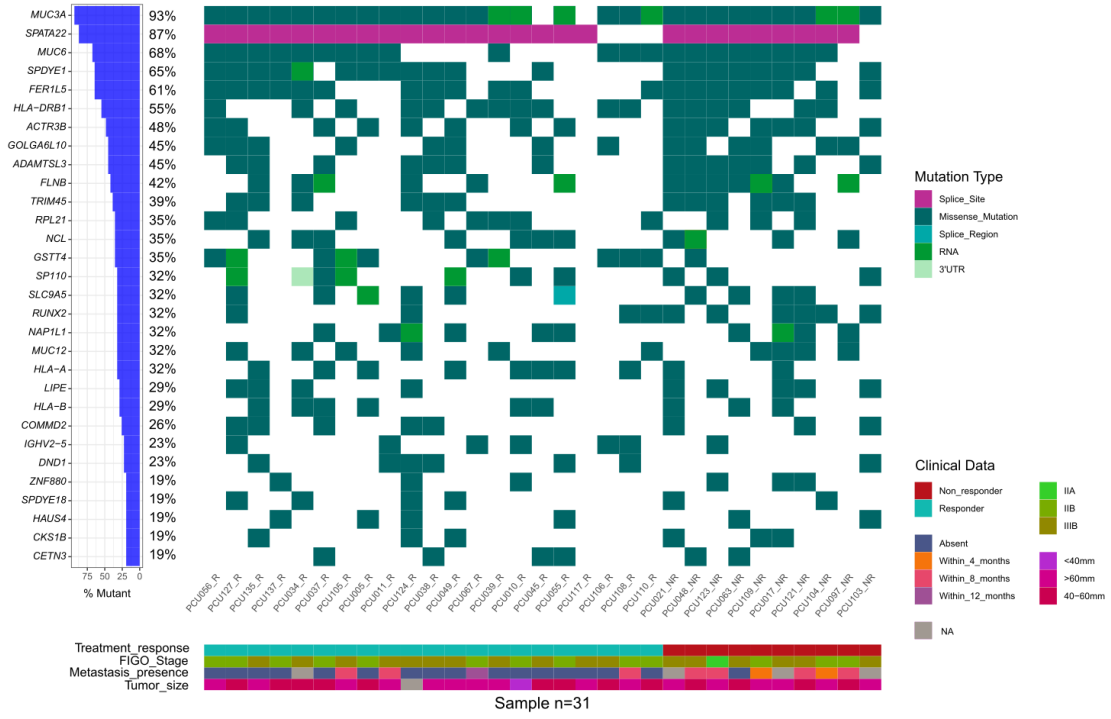


Figure 1: Distribution of frequently mutated genes in each patient. The bar plot on the left represents the percentage of patients harboring mutations in each gene. Patient clinical

data is displayed under their respective columns. RNA mutation type refers to variants located in non-coding transcripts from that gene. If multiple mutations were found in a gene in a single sample, only one is shown in the order as listed in the color legend from top to bottom.

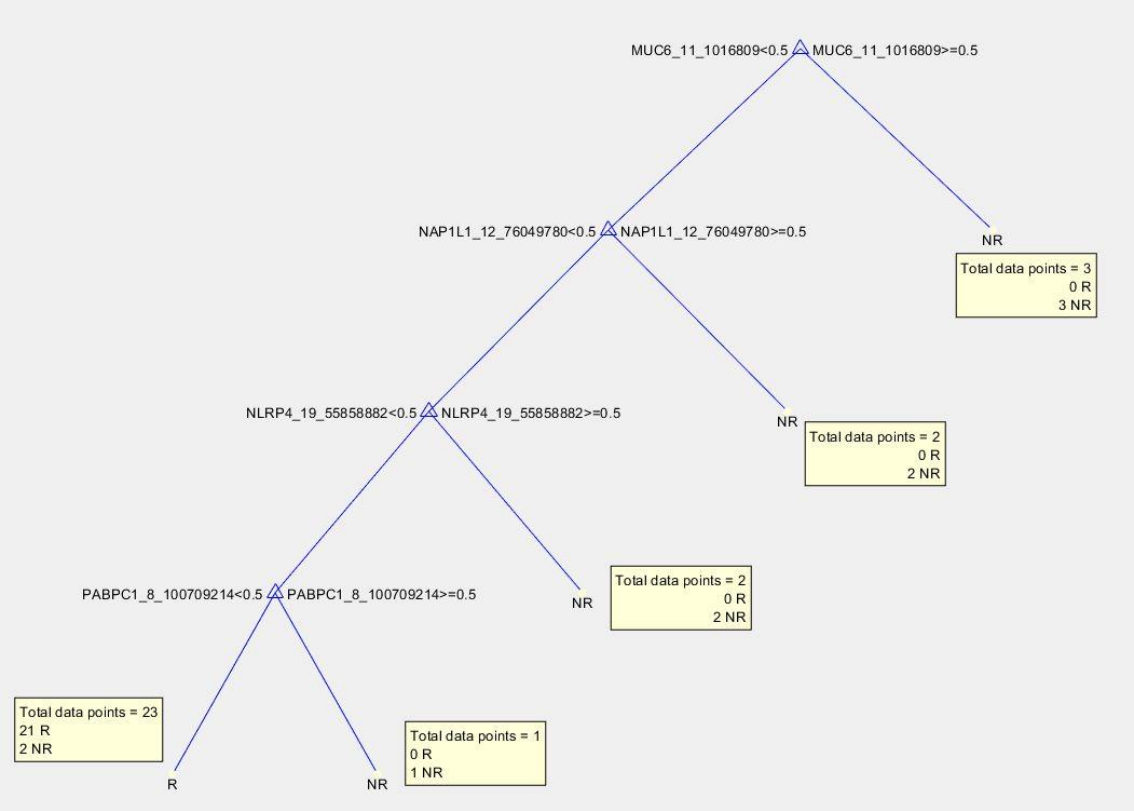
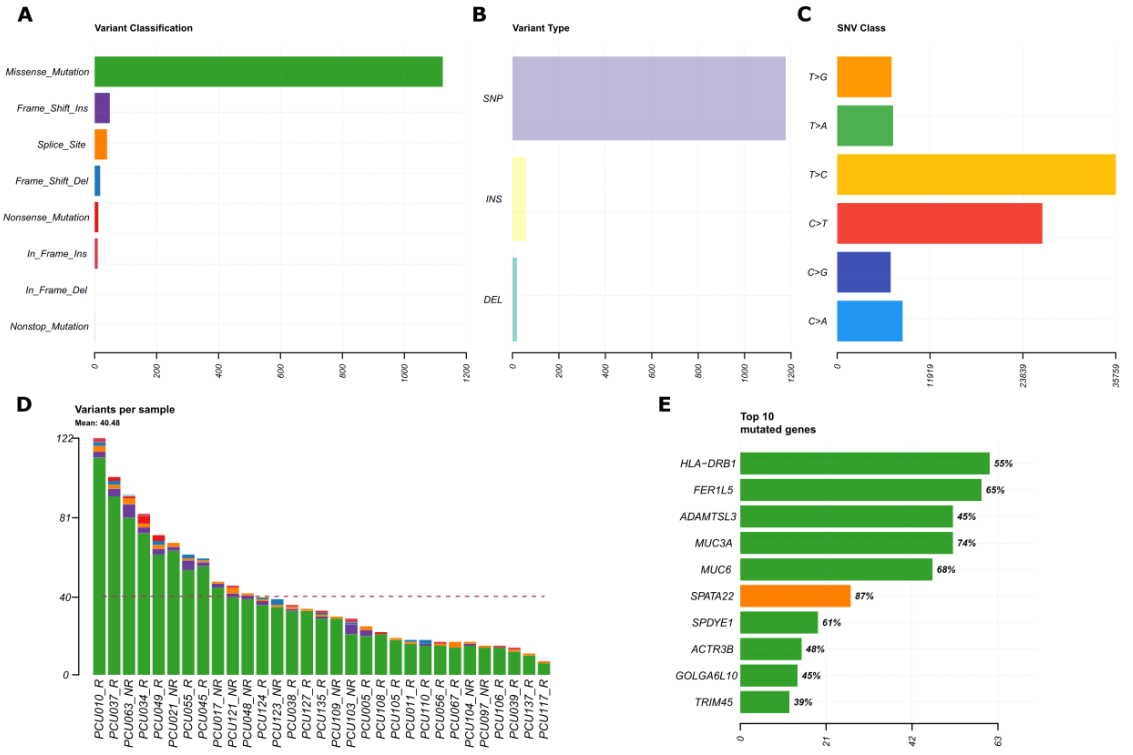
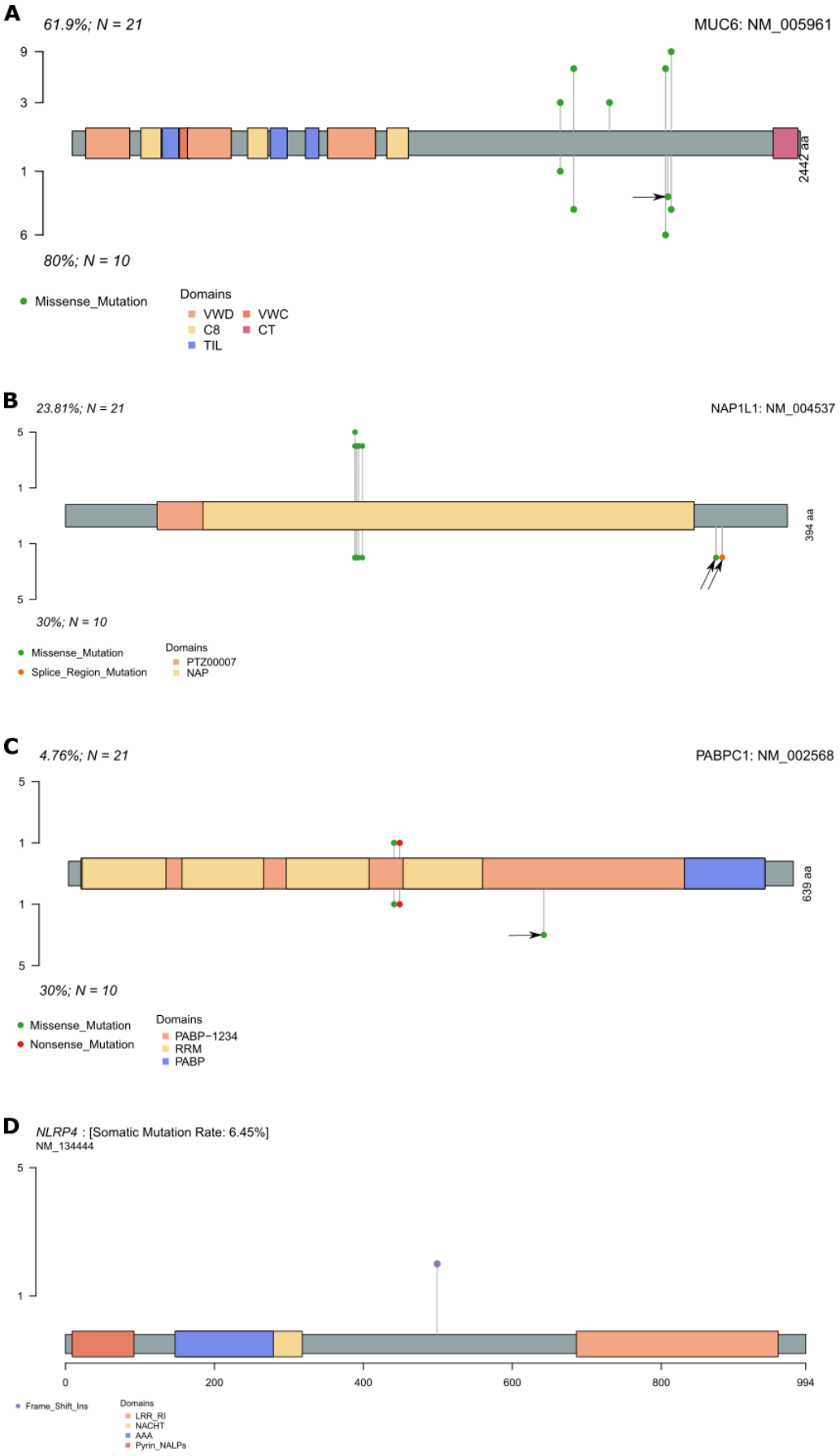


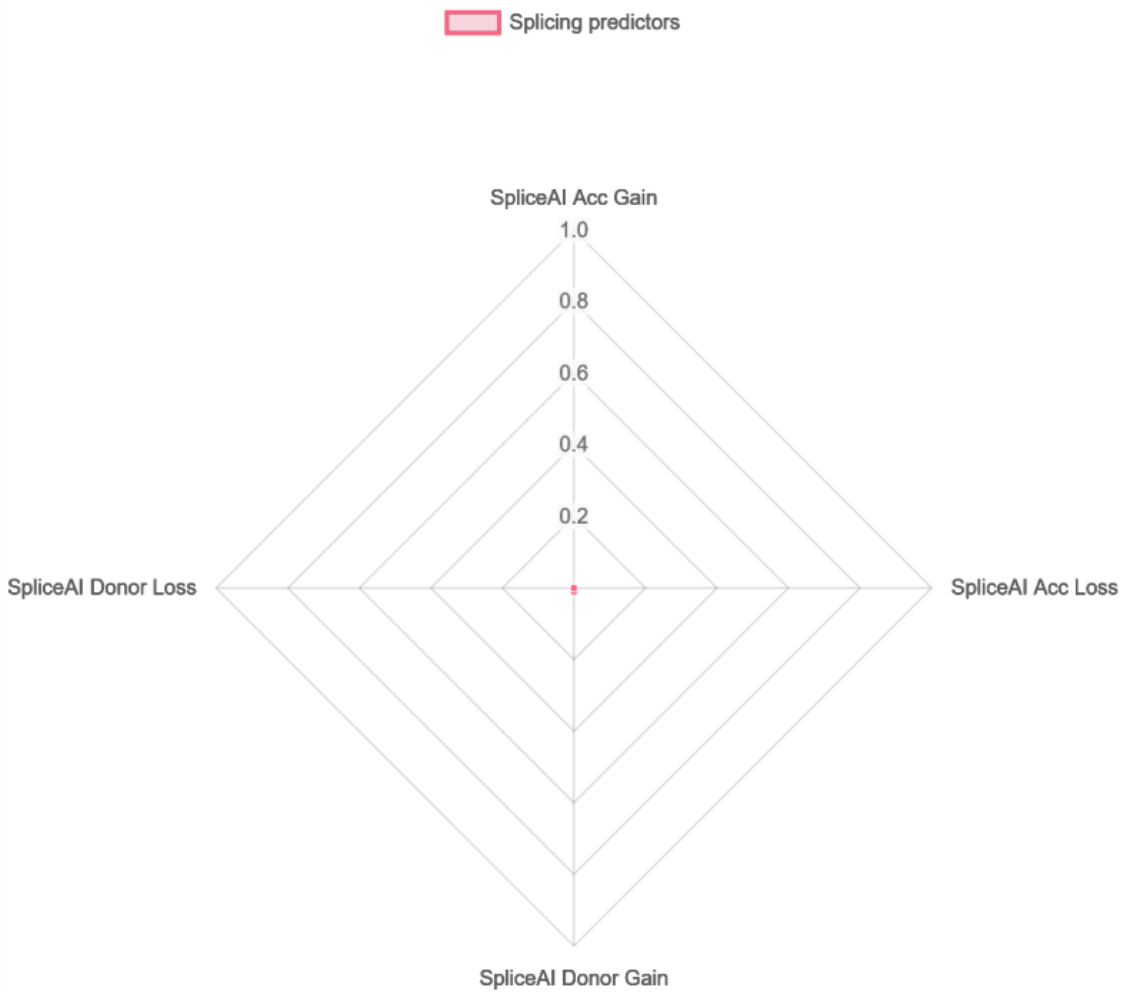
Figure 2: Decision Tree representation of the classification model using full training approach and variants related to non-response as attributes. The tree structure illustrates the hierarchical decision-making process based on key features, leading to the classification of variants into distinct categories.



Supplementary Figure 1: Summary of variant data from 31 patients. (A) Variant classification displaying total number of mutations identifies of each kind; (B) Number of SNPs, insertions and deletions indentified; (C) Distribution of base substitutions among SNPs; (D) Number of coding gene variants per sample; (E) Top 10 mutated genes. D and E follow the same color scheme presented in A to differentiate variant classes.


A
61.9%; N = 21
MUC6: NM_005961
9
3
1
6
2442 aa
80%; N = 10
Missense_Mutation
Domains
VWD
VWC
C8
CT
TIL
B
23.81%; N = 21
NAP1L1: NM_004537
5
1
1
5
394 aa
30%; N = 10
Missense_Mutation
Splice_Region_Mutation
Domains
PTZ00007
NAP
C
4.76%; N = 21
PABPC1: NM_002568
5
1
1
5
639 aa
30%; N = 10
Missense_Mutation
Nonsense_Mutation
Domains
PABP−1234
RRM
PABP
D
NLRP4 : [Somatic Mutation Rate: 6.45%]
NM_134444
5
1
0
200
400
600
800
994
Frame_Shift_Ins
Domains
LRR_RI
NACHT
AAA
Pyrin_NALPs

Supplementary Figure 2: Mutations identified across the protein structures of MUC6(A), NAP1L1(B), PABPC1(C) and NLRP4(D) proteins. Arrows indicate the variants from **Table 1** when multiple variants are located on the same gene. Percentages relate to number of R (N = 21) and NR (N = 10) patients harboring mutations in the gene. In A, B and C, where both R and NR patients harbored mutations in those genes, the top side represents mutations in R patients, while the underside represents NR patients.



Supplementary Figure 3: Splicing radar displaying predicted impact from NAP1L1(NM_004537.7):c.1062T>C.