RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 05/06/2026.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Instituto de Biociências
Unesp Botucatu

Programa de Pós-Graduação em Ciências Biológicas (Genética)

KRAS driver mutations and DNA methylation profiling in pancreatic cancer: from association with the immune cell microenvironment to theranostic biomarkers

BARBARA MITSUYASU BARBOSA

BOTUCATU - SP

2024

Universidade Estadual Paulista "Júlio de Mesquita Filho"
Instituto de Biociências de Botucatu
Programa de Pós-Graduação em Ciências Biológicas (Genética)

KRAS driver mutations and DNA methylation profiling in pancreatic cancer: from association with the immune cell microenvironment to theranostic biomarkers

Discente: Barbara Mitsuyasu Barbosa
Orientadora: Profa. Dra. Cláudia Aparecida Rainho
Coorientador: Prof. Dr. Roberto da Silva Gomes

Tese apresentada ao Instituto de Biociências, Campus de Botucatu, UNESP, como requisito para a obtenção do título de Doutor em Ciências Biológicas (Genética).

BOTUCATU - SP
2024

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Barbosa, Barbara Mitsuyasu.
KRAS driver mutations and DNA methylation profiling in pancreatic cancer : from association with the immune-cell microenvironment to theranostic biomarkers / Barbara Mitsuyasu Barbosa. - Botucatu, 2024

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Cláudia Aparecida Rainho
Coorientador: Roberto da Silva Gomes
Capes: 20200005

1. Molecular Docking Simulation. 2. Precision Medicine. 3. Tumor Microenvironment. 4. Computer simulation.

Palavras-chave: Deconvolution; In silico; Molecular docking; Precision medicine; Tumor microenvironment.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

ATA DA DEFESA EM SESSÃO FECHADA DA TESE DE DOUTORADO DE BARBARA MITSUYASU BARBOSA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA), DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 05 dias do mês de junho do ano de 2024, às 14:00 horas, no(a) por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de BARBARA MITSUYASU BARBOSA, intitulada **KRAS driver mutations and DNA methylation profiling in pancreatic cancer: from association with the immune-cell microenvironment to theranostic biomarkers**. A Comissão Examinadora foi constituida pelos seguintes membros: Profa. Dra. CLAUDIA APARECIDA RAINHO (Orientador(a) - Participação Virtual) do(a) Departamento de Ciências Químicas e Biológicas / Instituto de Biociências de Botucatu UNESP, Profa. Dra. DAISY MARIA FAVERO SALVADORI (Participação Virtual) do(a) Departamento de Patologia / Faculdade de Medicina de Botucatu - Unesp, Profa. Dra. ESTER SILVEIRA RAMOS (Participação Virtual) do(a) Departamento de Genética / Faculdade de Medicina - USP - Ribeirão Preto, Prof. Dr. EDUARDO BORGES DE MELO (Participação Virtual) do(a) Departamento de Ciências Farmacêuticas / Centro de Ciências Médicas e Farmacêuticas - Universidade Estadual do Oeste do Paraná - UNIOESTE, Prof. Dr. MARCELO RIZZATTI LUIZON (Participação Virtual) do(a) Departamento de Genética, Ecologia e Evolução / Instituto de Ciências Biológicas - Universidade Federal de Minas Gerais (ICB-UFMG). Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final:_ _ _ _ _ _ _ _ _ _ _ _ _ _ APROVADA _ _ _ _ _ _ _ _ _ _ _ _ _ _ . Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Profa. Dra. CLAUDIA APARECIDA RAINHO

Documento assinado digitalmente
gov.br
CLAUDIA APARECIDA RAINHO
Data: 07/06/2024 13:17:20-0300
Verifique em https://validar.iti.gov.br

Agradecimentos

À minha família, em especial à minha mãe, **Tereza Mitie Mitsuyasu**, e à minha tia, **Célia Tomie Sato**, a quem devo quem sou. E à minha irmã, **Michele Mitsuyasu Barbosa**, e à minha prima, **Tatiana Yurie Sato**, que sempre estiveram ao meu lado.

À minha orientadora, **Profa. Dra. Cláudia Aparecida Rainho**, pela excelente orientação e dedicação. Obrigada por todo o suporte oferecido para que eu pudesse mostrar minha melhor versão. Além disso, obrigada por se importar não apenas com a minha formação acadêmica, mas também com meu bem-estar.

Ao meu coorientador, **Prof. Dr. Roberto da Silva Gomes**, pela excelente orientação e por me receber em seu laboratório na *North Dakota State University* (NDSU), USA. E também à **sua família**, que foi uma rede de apoio e amigos essenciais para tornar a experiência de morar fora do Brasil excelente.

A todos os meus amigos que torceram por mim e se mostraram animados com minha trajetória acadêmica, em especial **Mariana Gazoli Barbosa** e **Leonardo Cesar de Lima Inocêncio**.

Aos meus colegas de pós-graduação, em especial à equipe do Laboratório de Epigenética: **João Henrique Maia Assumpção**, **Diogo Pessuto Callegari**, **Naiade Calanca**, **Fernanda Aparecida dos Santos France**, **Debora Kazumi Maeda** e **Mai Ono**, com quem discuti ciência e compartilhei os bons e maus momentos desses últimos anos.

À **Alana Cristina Martins** e sua filha **Maria**, que me acolheram em sua casa nos Estados Unidos por um ano. Morar com vocês foi uma experiência única. A Alana foi minha parceira e quem me acompanhava debaixo de sol ou neve, em qualquer dia, para que eu pudesse aproveitar ao máximo a oportunidade do intercâmbio. E a Maria, que, apesar de jovem, tem uma inteligência e grande sensibilidade; ouvir você cantando no banho trazia alegria ao meu dia.

À comunidade brasileira de Fargo, North Dakota, USA, que me fez sentir acolhida e me proporcionou uma sensação de pertencimento, mesmo estando tão longe de nossa terra natal.

Ao programa de Pós-Graduação em Ciências Biológicas (Genética) por todo o auxílio prestado.

À todos os funcionários e docentes do Departamento de Ciências Químicas e Biológicas - Setor de Genética do Instituto de Biociências (UNESP, Botucatu, Brasil) e do *Pharmaceutical Sciences Department* (NDSU, Fargo, EUA), cujo trabalho facilitou o desenvolvimento da minha pesquisa.

À **CAPES** pelo auxílio financeiro concedido para a execução deste trabalho. E ao programa CAPES-PrInt UNESP (processo 88887.310463/2018-00, mobilidade 88887.569794/2020-00) pela bolsa de intercâmbio.

Resumo

O câncer de pâncreas é uma das neoplasias mais letais em todo o mundo, frequentemente resultando em desfechos desfavoráveis e falha do tratamento. Isso se deve principalmente à ausência de estratégias de rastreamento, dificuldades na detecção precoce do tumor devido à falta de sinais ou sintomas específicos, e ao comportamento agressivo da doença. O sucesso do tratamento depende principalmente da intervenção cirúrgica, que nem sempre é viável. Além disso, o desafio é acentuado pela baixa contagem de células tumorais e intensa desmoplasia, que são características desta doença. As mutações condutoras (do inglês*: driver mutations*) no gene *KRAS* é prevalente, ocorre nos estágios iniciais do câncer de pâncreas e são alvo de terapias direcionadas. A mutação *missense* c.34G>T (rs121913530) acarreta na substituição de glicina por cisteína na proteína codificada KRAS p.G12C, sendo esta correlacionada com pior prognóstico e resistência à terapia em alguns cânceres humanos, incluindo o câncer de pulmão, mas pouco estudada em cânceres pancreáticos. Neste contexto, os principais objetivos deste estudo foram: 1) investigar o impacto das variantes patogênicas de *KRAS* no microambiente tumoral (MAT) e 2) planejar pequenas moléculas que seletivamente se liguem à variante patogênica p.G12C do gene *KRAS*. A composição das células imunes no MAT do câncer de pâncreas foi predita pela deconvolução de dados de metilação de DNA usando uma coorte recuperada de bancos de dados públicos (TCGA Pancreatic Cancer ou TCGA-PAAD). A avaliação de três conjugados de droga-sonda de infravermelho foi realizada no software PyRx (*Virtual Screening software for Computational Drug Discovery*) para prever *in silico* a melhor orientação e conformação dos compostos de teste ligados à KRAS p.G12C (código PDB: 6OIM). A análise de metilação do DNA revelou perfis distintos de acordo com o grau de pureza da amostra tumoral e a presença de mutações em *KRAS*. A deconvolução dos dados de metilação do DNA para predição das frações de células imunes revelou três padrões de MAT: hipoinflamado, enriquecido por células mieloides e enriquecido por células linfoides. Os dados predições *in silico* foram realizadas para determinar a melhor orientação e conformação dos três compostos ligados à KRAS p.G12C (PDB ID: 6OIM) utilizando o *software* de *docking* AutoDock Vina por meio da interface PyRx. Conforme evidenciado pela de energia de ligação, nossos resultados indicam que os compostos testados têm excelente afinidade pela KRAS p.G12C. Logo, nossa pesquisa oferece abordagens promissoras para aprimorar o manejo do câncer de pâncreas por meio da estratificação de pacientes e do desenvolvimento de novos teranósticos.

Abstract

Pancreatic cancer is one of the most lethal malignancies worldwide, often leading to therapeutic failure and unfavorable outcomes. This is primarily due to the absence of screening strategies, challenges in early tumor detection because of the lack of specific signs or symptoms, and the aggressive behavior of the disease. Successful treatment heavily relies on surgical intervention, which may not always be feasible. Additionally, the challenge is exacerbated by the characteristic low tumor cell count and intense desmoplasia. Driver mutations in the KRAS gene are common, occurring in the early stages of pancreatic cancer, and serving as targets for specialized therapies. The missense mutation c.34G>T (rs121913530) leads to the substitution of glycine with cysteine in the encoded KRAS protein (p.G12C). This specific mutation is associated with poorer prognosis and resistance to therapy in various human cancers, including lung cancer, but remains understudied in pancreatic cancers. In this context, the main objectives of this study were: 1) investigate the impact of pathogenic KRAS variants on the tumor microenvironment (TME), and 2) design small molecules that selectively bind to the pathogenic p.G12C variant of the *KRAS* gene. The immune cell composition of the pancreatic TME was predicted through the deconvolution of DNA methylation data using a cohort retrieved from public databases (TCGA Pancreatic Cancer or TCGA-PAAD). The evaluation of three infrared drug-probe conjugates was performed using PyRx software (Virtual Screening software for Computational Drug Discovery) to *predict in silico* the best orientation and conformation of the test compounds bound to KRAS p.G12C (PDB ID: 6OIM). DNA methylation analysis revealed distinct profiles based on the purity degree of the tumor sample and the presence of *KRAS* mutations. Deconvolution of DNA methylation data for predicting immune cell fractions revealed three TME patterns: hypo-inflamed, myeloid-enriched, and lymphoid-enriched. *In silico* predictions were conducted to determine the best orientation and conformation of these compounds bound to KRAS p.G12C (PDB code: 6OIM) utilizing the AutoDock Vina docking software through the PyRx interface. The binding energy data indicated that the tested compounds have excellent affinity for KRAS p.G12C. Therefore, our research offers promising approaches to enhance pancreatic cancer management through patient stratification and the development of novel theranostics.

Table of Contents

1. Literature review

1.1. Cancer overview

Cancer is considered a major public health problem. According to data from the World Health Organization, cancer is the leading cause of death worldwide[1]. A total of 20 million new cases and 9.7 million deaths from the disease were estimated globally in 2022 [2].

The incidence and mortality of cancer have been increasing in recent decades, and this trend is expected to be maintained. According to Global Cancer Observatory (GLOBOCAN; https://gco.iarc.fr/), the cancer burden is expected to increase 77% by 2050 resulting in 35 million new cases worldwide due to rising risk factors associated with globalization, the aging of the population, and the growing economy [2].

Cancer is the denomination used to represent a wide range of diseases characterized by the uncontrolled proliferation of abnormal cells which can infiltrate and destroy normal body tissue, and often have the ability to metastasize, i.e., spread from the primary site where the neoplasm first formed to distant anatomical location in the body [3,4].

Cancer can be classified into three main classes according to the tissue from which the tumor originates: 1. Sarcomas (originated in mesenchymal tissue such as bone, muscle, or connective; or the nervous system); 2. Carcinomas (the most common type of cancer, it originates in the epithelial tissue such as epithelial skin tissue or the tissue that lines internal organs); 3. Hematopoietic and lymphoid malignant neoplasms (originated in the bone marrow, lymphatic system, or peripheral blood). Tumors can be further classified by site/organ of origin,

histological characteristics, malignancy degree, chromosomal aneuploidy, gene mutations, and abnormalities in gene expression and epigenetic marks [3].

In 2000, Hanahan and Weinberg outlined the common aspects amongst cancer types at the cellular phenotype level to propose *"The Hallmarks of Cancer"*, a set of six capabilities necessary for normal human cells and tissues to become malignant: proliferative signaling, growth suppressors evasion, tissue invasion and metastasis, limitless replicative potential, inducing angiogenesis, and evading apoptosis [5]. The hallmarks have been frequently updated to incorporate new information generated by recent research. In 2011, four new attributes were added to the list: two new potential hallmarks (deregulated metabolism, and evasion of the immune system); and two enabling characteristics (genome instability and mutation, and tumor-promoting inflammation). It was proposed that the acquisition of enabling characteristics would facilitate the hallmarks capabilities [6,7]. Recently, the list were update to include two new emerging hallmarks (unlocking phenotypic plasticity, and cellular senescence) and two new enabling characteristics (nonmutation epigenetic reprogramming, and polymorphic variability in microbiomes) [6,7].

In general, a tumor is the result of natural selection acting on somatic mutations in a cell population within an individual being, i.e., a single multicellular organism, over generations of cells. The cell's ability to proliferate is determined by the relative rates of cell division and death. In a normal state, cells multiply through cell division and when they become old or damaged, they undergo the cell death process and are replaced by new ones. Both processes are sophisticatedly controlled in accordance with the needs of each specific tissue to prevent unwanted cells from multiplying. Any cell that does not pass this control

triggers apoptosis. When the control mechanisms fail, the abnormal and/or damaged cells proliferate, and tumors can be formed [4].

Although a tumor can develop from a single mutant cell, cells in a tumor are not identical clones because they acquire consecutive and cumulative genetic (somatic acquired mutations) and epigenetic changes that can lead to hallmark capabilities, resulting in complex and heterogeneous disease during cancer development and progression [4,6,8]. The characteristics acquired by tumor cells as a result of this process include physiological, morphological and/or biochemical changes in cell signaling and proliferation, genomic instability, inhibition of tumor suppressors, resistance to cell death, replicative immortality, and angiogenesis – resulting in a dynamic tissue composed of different cell types with functions and distinct interactions that is unique to each individual [7,9]. Thus, tumor cells within the same tumor can contain different molecular signatures and this heterogeneity can affect the tumor's sensibility to treatment[8].

Although mutations are key occurrences in cancer development, different mutations have different functional impact on carcinogenesis. Random mutations with no reoccurrence in specific cancer types are called passenger mutations, they appear as a result of cancer development rather than directly influence the development or progression of the disease. Mutations found at high frequency in the same or multiple types of cancer are presumed to be part of cancer development and/or progression and are called driver mutations [3]. These mutations occur in genes related to the acquisition of the hallmark capabilities, which involve pathways related to cell growth and proliferation, apoptosis inhibition, immune response evasion, invasion, and metastasis processes [10,11].

Based on the effect drive mutations have on oncogenesis, driver genes usually are classified into oncogenes and tumor suppressor genes. Oncogenes are derived from their normal counterparts called proto-oncogenes. A mutated version of a proto-oncogene is an activated oncogene due to a gain-of-function mutation, which can be even a single mutation at one allele, resulting in an enhanced phenotype of unregulated cell division. In a normal state, tumor suppressor genes limit cell division as a control mechanism, cancer cells have loss-of-function mutations that disable this control. Tumor suppressor genes usually require mutation in two alleles to result in loss of protein expression [3,4].

1.1.1. Cancer and precision medicine

There is no single treatment that is effective for all cancers due to the heterogeneous profile of the disease. Nowadays, treatment decision making to choose which treatment is the most suitable for a cancer patient (i.e., surgical, neoadjuvant or adjuvant chemo and/or radiotherapy) is according to the organ of origin, clinical staging, and histopathological analysis. Commonly, cancer treatment combines multimodal therapies: surgery followed by adjuvant chemotherapy and radiation therapy is a common treatment approach. The conventional protocols often consist of cytotoxic and non-specific chemotherapy in a one-size-fits-all approach. Precision medicine is a new promising field of oncology that focuses on understanding the individual variability of cancer, using information about the molecular profile of the tumor to decide the best therapy for each individual [9,10,12,13].

Precision medicine focuses on tumor-specific alterations responsible for malignant transformation and/or tumor progression that can be accurately targeted. This molecular target should have absolute or relative sparing of normal

cells and a higher selectivity to tumor cells, resulting in a better outcome and fewer adverse effects than traditional cancer treatments. The targeted treatment according to the genetic profile of the disease is proposed to be the future of oncotherapy and it is only possible due to the advances in tumor biology and molecular stratification of cancer subtypes [10,12,14].

The success of precision medicine relies on the identification of targets on which cancer cells are completely dependent, which are usually driver genes. If cancer cells are completely dependent on that specific tumor alteration, a lethal or cytostatic effect will be observed when its function is compromised [14]. The target can be named a biomarker if it has diagnostic, prognostic, predictive, and/or pharmacogenomic attributes [10,12].

Precision medicine in cancer research is a promising field, in a one-year-period (08/2020 - 07/2021), the FDA approved 10 new anticancer therapeutics targeting genetic mutations or against specific molecules (Table 1) [15].

Table 1. New FDA-approved anticancer therapeutics from August 1st, 2020-July 31st, 2021.

APPROVED INDICATION	TRADE NAME (GENERIC NAME)
Angiogenesis Inhibitors	
Certain type of kidney cancer	Fotivda® (Tivozanib)
Cell-signaling Inhibitors	
Certain type of lung cancer	Rybrevant® (amivantamab-vmjw)
Certain type of non-Hodkin lymphoma	Xalkori® (crizotinib)
Certain types of lung and thyroid cancers	Graveto® (pralsetinib)
Certain type of lung cancer	Tepmetko® (tepotinib)
Certain types of non-Hodkin lymphoma	Ukoniq® (umbralisib)
Certain type of lung cancer	Lumakras® (sotorasib)
Certain type of leukemia	Ayvakit® (avapritinib)
Bile duct cancer	Truseltiq® (infigratinib)
Cell Cytoskeleton-modifying Agents	
Multiple myeloma	Blenrep® (belantamab mafodotin-blmf)
DNA-damaging Agents	
Certain type of gastrointestinal cancers	Enhertu® (fam-trastuzumab deruxtecan-nxki)
Certain type of bladder cancer	Trodelvy® (sacituzumab govitecan-hziy)
Certain type of non-Hodkin lymphoma	Zynlonta® (loncastuximab tesirine-lpyl)
Multiple myeloma	Pepaxto® (melphalan flufenamide)
Epigenome-modifying Agents	
Certain type of leukemia	Onureg® (azacytidine)
Hormones/Antihormones	
Certain type of prostate cancer	Orgovyx® (relugolix)
Immunotherapeutics	
Certain types of skin and lung cancers	Libtayo® (cemiplimab-rwlc)
Certain type of non-Hodkin lymphoma	Breyanzi® (lisocabtagene maraleucel)
Certain type of non-Hodkin lymphoma	Yescarta® (axicabtagene ciloleucel)
Mesothelioma	Yervoy® (ipilimumab) and Opdivo® (nivolumab)
Neuroblastoma	Danyelza® (naxitamab-gqgk)
Multiple myeloma	Abecma® (idecabtagene vicleucel)
Certain type of endometrial cancer	Jemperli® (dostarlimab-gxly)
Gastroesophageal junction cancers	Opdivo® (nivolumab)
Certain type of breast cancer	Margenza® (margetuximab-cmkb)
Certain types of breast, gastric, and gastroesophageal junction cancers	Keytruda® (pembrolizumab)
Imaging Agents	
Prostate cancer	Ga 68 PSMA-11® (gallium 68 PSMA-11)
Prostate cancer	Pylarify® (piflufolastat F-18)
Certain type of neuroendocrine tumor	Detectnet® (copper Cu 64 dotatate)
Companion Diagnostic Tests	
Certain type of lung cancer	Guardant 360 CDx® (N/A)
Certain types of breast, lung, ovarian, and prostate cancers	FoundationOne Liquid CDx® (N/A)
Surgery Guiding Devices	
Osteoid osteoma in the extremities	Sonalleve MR-HIFU system® (N/A)
Artificial intelligence-guided assessment for liver cancer	Hepatica® (N/A)

Table adapted from American Association for Cancer Research (AACR) Cancer Progress Report 2021 [15].

1.2. Pancreatic cancer

Pancreatic cancer originates from exocrine and neuroendocrine cell types. Neuroendocrine pancreatic tumors are less common and have a better prognosis than pancreatic adenocarcinoma (PAAD), the most frequent form which corresponds to 80-90% of total cases. GLOBOCAN estimates an increase in its incidence and mortality rates [2]. In 2022, pancreatic cancer was estimated to be responsible for 467,409 deaths worldwide, placing it among the most lethal malignancies [2].

The disease is characterized by its clinical aggressiveness, advanced staging at diagnosis, and resistance to the available treatment modalities. The determining factors for therapeutic failure and unfavorable outcomes of pancreatic cancer are the absence of screening strategies, difficulties in tumor early detection due to the lack of specific signs/symptoms, and the aggressive behavior of the disease [16]. The morphological and molecular intra- and inter-tumor heterogeneity accompany the disease progression. Histologically, pancreatic cancer shows a prominent desmoplastic reaction with dense fibrotic stroma and low tumor cell count (5-20% of cancer cells), which makes it difficult to interpret molecular data obtained from tumor tissue biopsies and data based on high-level performance analysis used to determine genomic, transcriptomic, epigenomic, proteomic, among other profiles [17]. This desmoplastic profile could also impose a problem for drug delivery. Despite the challenges imposed by these intrinsic biological characteristics, many efforts have been made to identify molecular signatures, including screening for recurrent genetic mutations that can be used to better understand the molecular basis of the disease and to develop strategies for early detection and new therapeutic approaches [17].

There are no well-established treatment strategies based on the molecular profile of pancreatic cancer and the only therapeutic modality capable of offering chances of cure is a surgical approach, which is possible only in a minority of cases due to its late diagnosis [18-20]. In addition, morbidities associated with pancreatic cancer, such as fatigue and pain, significantly affect the quality of life and tolerance to the effects of chemoradiotherapy by patients. Data from the Surveillance, Epidemiology, and End Results (SEER) program demonstrated that pancreatic cancer early detection has a positive impact on prognosis [18].

However, early detection of the disease is still a challenge with the currently available technology. Full-body computed tomography is used for high-risk patients, which in addition to hardly identifying masses smaller than 2 centimeters and presenting a high incidence of false positives, it is also costly and exposes the patient to radiation [18,19].

The Pancreatic Intraepithelial Neoplasia (PanIN) classification is a standardized system for describing precancerous lesions in the pancreatic ducts, categorizing them based on the degree of dysplasia observed in epithelial cells. It includes low-grade lesions (PanIN-1A, with minimal cellular abnormalities and basally located nuclei; PanIN-1B, with more complex architecture but minimal atypia; and PanIN-2, showing moderate atypia with nuclear abnormalities and increased cell division) and high-grade lesions (PanIN-3, characterized by severe atypia, significant nuclear abnormalities, loss of cell polarity, and high mitotic activity, indicating carcinoma in situ and the most advanced precursor to invasive cancer) [21]. Ideally, pancreatic cancer detected and treated in the PanIN 3 stage (carcinoma *in situ*) would maximize the chance of cure and decrease the need of adjuvant treatments [19].

In this scenario, the integration of molecular profile data at different “omics” levels provided by The Cancer Genome Atlas Project (TCGA, available at https://www.cancer.gov/tcga) allowed the identification of specific molecular signatures for pancreatic cancer subtyping (based on differential expression of transcripts including coding and non-coding such as microRNAs and long non-coding RNAs, as well epigenomics data, such as DNA methylation profiling). It should be highlighted that analysis of the mutational profile showed that the *KRAS* gene is mutated in ~93% samples with a subset of pancreatic tumors showing multiple mutations in this gene in addition to the evidence of biallelic mutations [17]. Also, 60% of cancers negative for *KRAS* mutation had a mutation in a RAS-MAPK signaling pathway gene, such as *RREB1* (*ras responsive element binding protein 1*) [17,20].

The mutational activation of *KRAS* occurs in the early stages of pancreatic cancer, before its progression to invasive carcinoma; therefore, *KRAS* mutations can be considered biomarkers for early detection of the disease [19].

1.3. Cancer epigenetics

Epigenetics is the study of modifications to DNA or its packaging that do not alter the DNA sequence and can be passed from parent to daughter cell during cell division or from one generation to the next [22].

Epigenetic modifications encompass a range of processes, including histone post-translational modifications, DNA modifications such as DNA methylation, noncoding RNAs, and chromatin remodeling. The epigenetic marks alter DNA accessibility and chromatin structure, as those observed in hetero- and euchromatin, thereby modulating gene regulation [23,24]. Epigenetic mechanisms are crucial in essential cellular processes during normal development: de novo

DNA methylation is implicated in cell differentiation, leading to stable tissue- and cell-specific methylome and transcriptome profiles. Therefore, epigenetics plays a pivotal role in establishing and maintaining cell identity [24-26].

DNA methylation, the predominant epigenetic modification, is the addition of a methyl group (CH_3) by DNA methyltransferase enzymes to the 5' position of cytosine residues predominantly in CpG dinucleotides associated with promoters (promoter-associated CpG islands). This regulatory process influences gene expression by modulating transcription factor binding and accessibility to regulatory regions within the DNA [27]. The methylation patterns observed in replicates of the same cell type from different individuals are remarkably similar [28].

However, cancer cells display distinct methylation patterns compared to non-cancer cells, featuring global hypomethylation alongside targeted hypermethylation of specific genes important for the malignant phenotype [28,29]. These DNA methylation alterations progress gradually during cancer initiation and progression [30]. The influence of the tissue of origin on the methylation pattern observed in cancer, coupled with the characteristic patterns of DNA methylation across the genome in different cell types, results in a heterogeneous methylation pattern in cancer cells [28,29]. The DNA methylation profile of a tumor not only offers insights into the current state of the cancer cells but also aids in the identification of its cell type of origin [31].

Previous studies have explored the potential of DNA methylation deconvolution in different tumor samples, such as glial/glioneuronal [32], hepatocellular [33], and prostate cancers [34]. Notably, a recent study identified immunologically hot and cold subsets of lung cancer by utilizing bioinformatic

tools for immune cell deconvolution based on DNA methylation data [35]. This study investigated concurrent factors influencing the immune microenvironment, including the presence of *KRAS* or *TP53* mutations. *KRAS* mutations have been extensively studied as potential prognostic and therapeutic targets in various human cancers. The driver mutations were associated with poor overall survival and specific immune cell populations [35]. To our knowledge, no studies have explored the potential of DNA methylation deconvolution in PAAD.

1.4. Bioinformatics and tumor heterogeneity

High-throughput technologies have yielded cancer data comprising DNA, RNA, protein, and epigenomic information. Although the unprecedented volume and complexity of these data from large tumor cohorts pose significant challenges for scientists, they also present opportunities for reanalysis to generate new knowledge about this disease [36].

Bioinformatics is important for the management, analysis, and interpretation of extensive high-throughput data in cancer research. By employing bioinformatics methodologies, new insights into the molecular factors underlying cancer, such as tumor-specific DNA alterations, pathways, and metabolic processes driving tumor growth and metastasis can be uncovered. Moreover, bioinformatics facilitates predicting outcomes such as cancer onset and progression, thereby aiding in the development of personalized treatment approaches [37].

High-throughput biomedical data, deposited in public platforms, enables researchers worldwide to reanalyze and integrate diverse datasets, facilitating the discovery of differential patterns associated with clinical and prognostics parameters in human disease [38]. The Cancer Genome Atlas Project (TCGA,

available at https://www.cancer.gov/tcga), is an example of such a platform. The integration of profile data at different “omics” levels provided by TCGA allowed the molecular identification of specific molecular signatures for pancreatic cancer subtyping. It should be highlighted that analysis of the mutational profile showed that the *KRAS* gene is mutated in approximately 93% of samples, with a subset of pancreatic tumors showing multiple mutations in this gene in addition to evidence of biallelic mutations [17]. In addition, 60% of cancers negative for *KRAS* mutation had a mutation in a RAS-MAPK signaling pathway gene, such as *RREB1* (*ras responsive element binding protein 1*) [17,20]. These data indicate the role of *KRAS* as a hub/key gene specifically associated with PAAD.

A tumor biopsy encompasses tumor cells, infiltrating non-tumor cells, and adjacent normal cells. This intrinsic heterogeneity poses a significant challenge, impeding the robustness and reproducibility of bioinformatic analyses for biomarker identification utilizing multi-omics profiles[36,39]. To address the limitations of bulk analysis, single-cell approaches have been developed to study intratumoral heterogeneity [40,41]. The single-cell RNA sequencing (scRNA-seq) approach has become a leading method for investigating cellular diversity, allowing scientists to explore not only tumors but also normal tissues. This method has facilitated the discovery of diverse cell types and states within healthy tissues, leading to substantial progress in our understanding of tumor physiology and pathophysiology [42].

While single-cell analysis provides unparalleled resolution for investigating tumor heterogeneity, it is constrained by technical limitations. The large datasets produced by single-cell profiling require advanced computational methods and substantial resources to integrate, analyze, and interpret the complex patterns of

cellular heterogeneity. Consequently, the technical and computational challenges associated with this technique render it not viable for the study of large patient cohorts and fixed tumor samples [40,41]. However, the single-cell profiles obtained can serve as reference to dissect *in silico* bulk-tumor data through deconvolution analysis [43].

Deconvolution is a computational approach that aims to infer the cellular composition of a bulk tissue sample by leveraging single-cell reference data. Harnessing the detailed cellular signatures generated through single-cell profiling, deconvolution algorithms can quantify the relative proportions of different cell types present within a bulk tumor sample [44]. These algorithms employ diverse mathematical approaches and can be applied to various types of data, including genomic (e.g., CIBERSORT [45] and xCELL [46]), transcriptomic (e.g., EPIC [47]), and epigenomic (e.g., MethylCIBERSORT [48] and HiTIMED [49]) datasets.

References

1. Prior IA, Lewis PD, Mattos C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Research*. 2012-05-15 2012;72(10):2457-2467. doi:10.1158/0008-5472.can-11-2612

2. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*. 2019-01-08 2019;47(D1):D941-D947. doi:10.1093/nar/gky1015

3. Wilson CY, Tolias P. Recent advances in cancer drug discovery targeting RAS. *Drug Discovery Today*. 2016-12-01 2016;21(12):1915-1919. doi:10.1016/j.drudis.2016.08.002

4. Liu P, Wang Y, Li X. Targeting the untargetable KRAS in cancer therapy. *Acta Pharmaceutica Sinica B*. 2019-09-01 2019;9(5):871-879. doi:10.1016/j.apsb.2019.03.002

5. Moore AR, Rosenberg SC, Mccormick F, Malek S. RAS-targeted therapies: is the undruggable drugged? *Nature Reviews Drug Discovery*. 2020-08-15 2020;19(8):533-552. doi:10.1038/s41573-020-0068-6

6. Yabar CS, Winter JM. Pancreatic Cancer. *Gastroenterology Clinics of North America*. 2016-09-01 2016;45(3):429-445. doi:10.1016/j.gtc.2016.04.003

7. Neoptolemos JP, Kleeff J, Michl P, Costello E, Greenhalf W, Palmer DH. Therapeutic developments in pancreatic cancer: current and future perspectives. *Nature Reviews Gastroenterology & Hepatology*. 2018-06-01 2018;15(6):333-348. doi:10.1038/s41575-018-0005-x

8. Lanfredini S, Thapa A, O'Neill E. RAS in pancreatic cancer. *Biochemical Society Transactions*. 2019-08-30 2019;47(4):961-972. doi:10.1042/bst20170521

9. Pujar S, O'Leary NA, Farrell CM, et al. Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding

regions supported by expert curation. *Nucleic Acids Research*. 2018-01-04 2018;46(D1):D221-D228. doi:10.1093/nar/gkx1031

10. Harvey JJ. An Unidentified Virus which causes the Rapid Production of Tumours in Mice. *Nature*. 1964/12/01 1964;204(4963):1104-1105. doi:10.1038/2041104b0

11. Kirsten WH, Mayer LA. Morphologic responses to a murine erythroblastosis virus. *J Natl Cancer Inst*. Aug 1967;39(2):311-35.

12. Scolnick EM, Rands E, Williams D, Parks WP. Studies on the Nucleic Acid Sequences of Kirsten Sarcoma Virus: a Model for Formation of a Mammalian RNA-Containing Sarcoma Virus. *Journal of Virology*. 1973-09-01 1973;12(3):458-463. doi:10.1128/jvi.12.3.458-463.1973

13. Tsuchida N, Uesugi S. Structure and Functions of the Kirsten Murine Sarcoma Virus Genome: Molecular Cloning of Biologically Active Kirsten Murine Sarcoma Virus DNA. *Journal of Virology*. 1981-05-01 1981;38(2):720-727. doi:10.1128/jvi.38.2.720-727.1981

14. Der CJ, Krontiris TG, Cooper GM. Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. *Proceedings of the National Academy of Sciences*. 1982-06-01 1982;79(11):3637-3640. doi:10.1073/pnas.79.11.3637

15. Chang EH, Gonda MA, Ellis RW, Scolnick EM, Lowy DR. Human genome contains four genes homologous to transforming genes of Harvey and Kirsten murine sarcoma viruses. *Proceedings of the National Academy of Sciences*. 1982-08-01 1982;79(16):4848-4852. doi:10.1073/pnas.79.16.4848

16. Der CJ, Cooper GM. Altered gene products are associated with activation of cellular rasK genes in human lung and colon carcinomas. *Cell*. Jan 1983;32(1):201-8. doi:10.1016/0092-8674(83)90510-x

17. McCoy MS, Toole JJ, Cunningham JM, Chang EH, Lowy DR, Weinberg RA. Characterization of a human colon/lung carcinoma oncogene. *Nature*. Mar 3 1983;302(5903):79-81. doi:10.1038/302079a0

18. McGrath JP, Capon DJ, Smith DH, et al. Structure and organization of the human Ki-ras proto-oncogene and a related processed pseudogene. *Nature*. Aug 11-17 1983;304(5926):501-6. doi:10.1038/304501a0

19. Popescu NC, Amsbaugh SC, DiPaolo JA, Tronick SR, Aaronson SA, Swan DC. Chromosomal localization of three human ras genes by in situ molecular hybridization. *Somat Cell Mol Genet*. Mar 1985;11(2):149-55. doi:10.1007/bf01534703

20. Forrester K, Almoguera C, Han K, Grizzle WE, Perucho M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature*. 1987/05/01 1987;327(6120):298-303. doi:10.1038/327298a0

21. Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell*. May 20 1988;53(4):549-54. doi:10.1016/0092-8674(88)90571-5

22. Smit VTHBM, Boot AJM, Smits AMM, Fleuren GJ, Cornelisse CJ, Bos JL. KRAS codon 12 mutations occur very frequently in pancreatic adenocarcinomas. *Nucleic Acids Research*. 1988-01-01 1988;16(16):7773-7782. doi:10.1093/nar/16.16.7773

23. Shirasawa S, Furuse M, Yokoyama N, Sasazuki T. Altered Growth of Human Colon Cancer Cell Lines Disrupted at Activated Ki-ras. *Science*. 1993/04/02 1993;260(5104):85-88. doi:10.1126/science.8465203

24. Koera K, Nakamura K, Nakao K, et al. K-Ras is essential for the development of the mouse embryo. *Oncogene*. 1997-09-04 1997;15(10):1151-1159. doi:10.1038/sj.onc.1201284

25. Johnson L, Mercer K, Greenbaum D, et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature*. 2001-04-26 2001;410(6832):1111-1116. doi:10.1038/35074129

26. Carta C, Pantaleoni F, Bocchinfuso G, et al. Germline missense mutations affecting KRAS Isoform B are associated with a severe Noonan syndrome phenotype. *Am J Hum Genet*. Jul 2006;79(1):129-35. doi:10.1086/504394

27. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature*. 2013-11-01 2013;503(7477):548-551. doi:10.1038/nature12796

28. Canon J, Rex K, Saiki AY, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature*. 2019-11-07 2019;575(7781):217-223. doi:10.1038/s41586-019-1694-1

29. (FDA) USFaDA. FDA Approves First Targeted Therapy for Lung Cancer Mutation Previously Considered Resistant to Drug Therapy. *Press Announcements*. 2021. https://www.fda.gov/news-events/press-announcements/fda-approves-first-targeted-therapy-lung-cancer-mutation-previously-considered-resistant-drug

30. Jänne PA, Riely GJ, Gadgeel SM, et al. Adagrasib in Non-Small-Cell Lung Cancer Harboring a KRAS(G12C) Mutation. *N Engl J Med*. Jul 14 2022;387(2):120-131. doi:10.1056/NEJMoa2204619

31. Ostrem JML, Shokat KM. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. *Nature Reviews Drug Discovery*. 2016/11/01 2016;15(11):771-785. doi:10.1038/nrd.2016.139

32. Chuang H-C, Huang P-H, Kulp SK, Chen C-S. Pharmacological strategies to target oncogenic KRAS signaling in pancreatic cancer. *Pharmacological Research*. 2017-03-01 2017;117:370-376. doi:10.1016/j.phrs.2017.01.006

33. Christensen JG, Olson P, Briere T, Wiel C, Bergo MO. Targeting Krasg12c-mutant cancer with a mutation-specific inhibitor. *Journal of Internal Medicine*. 2020-08-01 2020;288(2):183-191. doi:10.1111/joim.13057

34. Zhang SS, Nagasaka M. Spotlight on Sotorasib (AMG 510) for KRAS (G12C) Positive Non-Small Cell Lung Cancer. *Lung Cancer (Auckl)*. 2021;12:115-122. doi:10.2147/lctt.s334623

35. Lyons DM, Lauring AS. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. *Molecular Biology and Evolution*. 2017-12-01 2017;34(12):3205-3215. doi:10.1093/molbev/msx251

36. Zako T, Hyodo H, Tsuji K, et al. Development of Near Infrared-Fluorescent Nanophosphors and Applications for Cancer Diagnosis and Therapy. *Journal of Nanomaterials*. 2010-01-01 2010;2010:1-7. doi:10.1155/2010/491471

37. Yang X, Shi C, Tong R, et al. Near IR Heptamethine Cyanine Dye–Mediated Cancer Imaging. *Clinical Cancer Research*. 2010-05-15 2010;16(10):2833-2844. doi:10.1158/1078-0432.ccr-10-0059

38. Shi C, Wu JB, Pan D. Review on near-infrared heptamethine cyanine dyes as theranostic agents for tumor imaging, targeting, and photodynamic therapy. *Journal of Biomedical Optics*. 2016-05-10 2016;21(5):050901. doi:10.1117/1.jbo.21.5.050901

39. Ebaston TM, Rozovsky A, Zaporozhets A, et al. Peptide-Driven Targeted Drug-Delivery System Comprising Turn-On Near-Infrared Fluorescent Xanthene–Cyanine Reporter for Real-Time Monitoring of Drug Release. *ChemMedChem*. 2019-10-04 2019;14(19):1727-1734. doi:10.1002/cmdc.201900464

40. Xiang B, Jia X-L, Qi J-L, et al. Enhancing siRNA-based cancer therapy using a new pH-responsive activatable cell-penetrating peptide-modified liposomal system. *International Journal of Nanomedicine*. 2017-03-01 2017;Volume 12:2385-2405. doi:10.2147/ijn.s129574

41. Raphael BJ, Hruban RH, Aguirre AJ, et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell*. 2017-08-01 2017;32(2):185-203.e13. doi:10.1016/j.ccell.2017.07.007

42. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*. 2013-10-01 2013;45(10):1113-1120. doi:10.1038/ng.2764

43. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*. 2009-01-01 2009:NA-NA. doi:10.1002/jcc.21334

44. Pathria RK, Beale PD. 16 - Computer Simulations. In: Pathria RK, Beale PD, eds. *Statistical Mechanics (Third Edition)*. Academic Press; 2011:637-652.

45. Jin S, Dang HT, Haug GC, et al. Visible Light-Induced Borylation of C–O, C–N, and C–X Bonds. *Journal of the American Chemical Society*. 2020-01-22 2020;142(3):1603-1613. doi:10.1021/jacs.9b12519

46. Hinkes SPA, Klein CDP. Virtues of Volatility: A Facile Transesterification Approach to Boronic Acids. *Organic Letters*. 2019-05-03 2019;21(9):3048-3052. doi:10.1021/acs.orglett.9b00584

47. Lovering F, Bikker J, Humblet C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *Journal of Medicinal Chemistry*. 2009-11-12 2009;52(21):6752-6756. doi:10.1021/jm901241e

48. Berman HM. The Protein Data Bank. *Nucleic Acids Research*. 2000-01-01 2000;28(1):235-242. doi:10.1093/nar/28.1.235

49. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021-05-01 2021;71(3):209-249. doi:10.3322/caac.21660

50. Nyíri K, Koppány G, Vértessy BG. Structure-based inhibitor design of mutant RAS proteins—a paradigm shift. *Cancer and Metastasis Reviews*. 2020-12-01 2020;39(4):1091-1105. doi:10.1007/s10555-020-09914-6

51. Pantsar T. The current understanding of KRAS protein structure and dynamics. *Computational and structural biotechnology journal*. 2019;18:189-198. doi:10.1016/j.csbj.2019.12.004

52. Bos JL, Rehmann H, Wittinghofer A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell*. 2007-06-01 2007;129(5):865-877. doi:10.1016/j.cell.2007.05.018

53. Vasan N, Boyer JL, Herbst RS. A RAS Renaissance: Emerging Targeted Therapies for KRAS-Mutated Non–Small Cell Lung Cancer. *Clinical Cancer*

Research. 2014-08-01 2014;20(15):3921-3930. doi:10.1158/1078-0432.ccr-13-1762

54. Medicine BMNLo, (US). ClinicalTrials.gov. Identifier: NCT03600883. https://clinicaltrials.gov/ct2/show/NCT03600883

55. Naim N, Moukheiber S, Daou S, Kourie HR. KRAS-G12C covalent inhibitors: A game changer in the scene of cancer therapies. *Critical Reviews in Oncology/Hematology*. 2021/12/01/ 2021;168:103524.

7. Conclusions

Chapter 1

- Two distinct methylation profiles were identified within pancreatic cancer samples. The group characterized by higher tumor purity demonstrated a higher frequency of *KRAS* mutations and was associated with poorer overall survival outcomes, indicating a potential link between methylation patterns, genetic mutations, and patient prognosis. The presence of *KRAS* mutations in pancreatic cancer was also correlated with positive epigenetic age acceleration, suggesting that *KRAS* mutation status may be a significant marker for assessing biological aging within tumor cells.

- We identified three distinct immune profiles within the tumor microenvironment through DNA methylation deconvolution: hypo-inflamed, myeloid-enriched, and lymphoid-enriched. This classification was further validated through gene expression data, confirming the suitability of DNA methylation-based deconvolution approaches to stratify patients with pancreatic cancer according to TIME composition.

Chapter 2

- The TCGA-PAAD dataset has a high prevalence of *KRAS* mutations. Although most KRAS mutations affect codon 12, the allele KRAS p.G12C occurs at a low frequency in pancreatic cancer.
- Structural analysis of the KRAS protein indicated that mutations at codons 12, 13, and 61 alter the surrounding surface of the GDP/GTP binding pocket, potentially creating new binding sites. These findings highlight the necessity for mutation-specific drug design, as each mutation presents unique surface characteristics that must be targeted for effective therapeutic intervention.
- The compounds designed in this study demonstrated excellent binding affinity to the KRAS p.G12C. This suggests that these compounds hold promise both as diagnostic biomarkers and as therapeutic targets in the context of pancreatic cancer drug development.