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.

2. Objectives

2.1. Main objectives

To study the relationship between *KRAS* driver mutations, DNA methylation profiles, and the immune cell microenvironment in pancreatic cancer, while concurrently developing a theoretical framework for a new pH-sensitive NIR fluorescent conjugate probe linked to a hydrazine moiety to selectively prevent tumor cell growth and increase the efficacy of early imaging diagnosis of pancreatic cancer.

2.2. Specific objectives

Chapter 1

Reanalyze data from The Cancer Genome Atlas Pancreatic Adenocarcinoma (TCGA-PAAD) dataset to achieve the following specific objectives:

- Characterize global DNA methylation patterns in pancreatic cancer;
- Investigate the relationship between DNA methylation age and *KRAS* driver mutations;
- Employ DNA methylation-based tumor deconvolution to infer the composition of immune cells in the tumor microenvironment and identify patterns based on the immune cell enrichment;
- Utilize gene expression data to validate deconvolution PAAD subtypes derived from DNA methylation data through co-expression network analysis.

C*hapter* 2

- Analyze the distribution frequency of the pathogenic variants of *the KRAS* gene in the TCGA-PAAD cohort;
- Identify and assess variations in the conformation and accessibility of binding pockets of allele-specific *KRAS*-mutated proteins;
- Assess the binding affinity and specificity of the probe-drug conjugates theoretically designed to target the site on KRAS p.G12C by molecular docking.

3. References

1. (WHO) WHO. Cancer Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/cancer

2. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. May-Jun 2024;74(3):229-263. doi:10.3322/caac.21834

3. Nussbaum R, McInnes R, Willard H. *Thompson & Thompson Genetics in Medicine*. 8th ed. Elsevier; 2015.

4. Strachan T, Read A. *Human Molecular Genetics*. 5th ed. Garland Science; 2018.

5. Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell*. 2000-01-01 2000;100(1):57-70. doi:10.1016/s0092-8674(00)81683-9

6. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery*. 2022-01-01 2022;12(1):31-46. doi:10.1158/2159-8290.cd-21-1059

7. Hanahan D, Weinberg A, Robert. Hallmarks of Cancer: The Next Generation. *Cell*. 2011-03-01 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013

8. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*. 2018/02/01 2018;15(2):81-94. doi:10.1038/nrclinonc.2017.166

9. Shin SH, Bode AM, Dong Z. Precision medicine: the foundation of future cancer therapeutics. *npj Precision Oncology*. 2017-12-01 2017;1(1)doi:10.1038/s41698-017-0016-z

10. Zugazagoitia J, Guedes C, Ponce S, Ferrer I, Molina-Pinelo S, Paz-Ares L. Current Challenges in Cancer Treatment. *Clinical Therapeutics*. 2016-07-01 2016;38(7):1551-1566. doi:10.1016/j.clinthera.2016.03.026

11. Novikov NM, Zolotaryova SY, Gautreau AM, Denisov EV. Mutational drivers of cancer cell migration and invasion. *British Journal of Cancer*. 2021-01-05 2021;124(1):102-114. doi:10.1038/s41416-020-01149-0

12. Zhang Q, Fu Q, Bai X, Liang T. Molecular Profiling–Based Precision Medicine in Cancer: A Review of Current Evidence and Challenges. Review. *Frontiers in Oncology*. 2020;10

13. Collins H, Calvo S, Greenberg K, Forman Neall L, Morrison S. Information Needs in the Precision Medicine Era: How Genetics Home Reference Can Help. *Interact J Med Res*. 2016/04/27 2016;5(2):e13. doi:10.2196/ijmr.5199

14. De Bono JS, Ashworth A. Translating cancer research into targeted therapeutics. *Nature*. 2010-09-01 2010;467(7315):543-549. doi:10.1038/nature09339

15. (AACR) AAfCR. DISCOVERY SCIENCE DRIVING CLINICAL BREAKTHROUGHS. https://cancerprogressreport.aacr.org/progress/cpr21-contents/cpr21-discovery-science-driving-clinical-breakthroughs/

16. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World Journal of Oncology*. 2019-01-01 2019;10(1):10-27. doi:10.14740/wjon1166

17. 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

18. Grossberg AJ, Chu LC, Deig CR, et al. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. *CA: A Cancer Journal for Clinicians*. 2020-09-01 2020;70(5):375-403. doi:10.3322/caac.21626

19. 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

20. 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

21. Basturk O, Hong SM, Wood LD, et al. A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas. *Am J Surg Pathol*. Dec 2015;39(12):1730-41. doi:10.1097/pas.0000000000000533

22. Weinhold B. Epigenetics: the science of change. *Environ Health Perspect*. 2006:A160-7. vol. 3.

23. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet*. Aug 2016;17(8):487-500. doi:10.1038/nrg.2016.59

24. Golbabapour S, Abdulla MA, Hajrezaei M. A concise review on epigenetic regulation: insight into molecular mechanisms. *Int J Mol Sci*. 2011;12(12):8661-94. doi:10.3390/ijms12128661

25. Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews Genetics*. 2017/11/01 2017;18(11):643-658. doi:10.1038/nrg.2017.57

26. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. Mar 2013;14(3):204-20. doi:10.1038/nrg3354

27. Lea AJ, Vockley CM, Johnston RA, et al. Genome-wide quantification of the effects of DNA methylation on human gene regulation. *Elife*. Dec 21 2018;7doi:10.7554/eLife.37513

28. Loyfer N, Magenheim J, Peretz A, et al. A DNA methylation atlas of normal human cell types. *Nature*. 2023/01/01 2023;613(7943):355-364. doi:10.1038/s41586-022-05580-6

29. Nishiyama A, Nakanishi M. Navigating the DNA methylation landscape of cancer. *Trends Genet*. Nov 2021;37(11):1012-1027. doi:10.1016/j.tig.2021.05.002

30. Baxter E, Windloch K, Gannon F, Lee JS. Epigenetic regulation in cancer progression. *Cell Biosci*. 2014;4:45. doi:10.1186/2045-3701-4-45

31. Moran S, Martínez-Cardús A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol*. Oct 2016;17(10):1386-1395. doi:10.1016/s1470-2045(16)30297-2

32. Singh O, Pratt D, Aldape K. Immune cell deconvolution of bulk DNA methylation data reveals an association with methylation class, key somatic alterations, and cell state in glial/glioneuronal tumors. *Acta Neuropathologica Communications*. 2021/09/08 2021;9(1):148. doi:10.1186/s40478-021-01249-9

33. Jain SS, McNamara ME, Varghese RS, Ressom HW. Deconvolution of immune cell composition and biological age of hepatocellular carcinoma using DNA methylation. *Methods*. Oct 2023;218:125-132. doi:10.1016/j.ymeth.2023.08.004

34. Reynolds SR, Zhang Z, Salas LA, Christensen BC. Tumor microenvironment deconvolution identifies cell-type-independent aberrant DNA methylation and gene expression in prostate cancer. *Clin Epigenetics*. Jan 3 2024;16(1):5. doi:10.1186/s13148-023-01609-3

35. Guidry K, Vasudevaraja V, Labbe K, et al. DNA Methylation Profiling Identifies Subgroups of Lung Adenocarcinoma with Distinct Immune Cell Composition, DNA Methylation Age, and Clinical Outcome. *Clinical Cancer Research*. 2022;28(17):3824-3835. doi:10.1158/1078-0432.CCR-22-0391

36. Song Q, Merajver SD, Li JZ. Cancer classification in the genomic era: five contemporary problems. *Hum Genomics*. Oct 19 2015;9:27. doi:10.1186/s40246-015-0049-8

37. Brenner C. Applications of Bioinformatics in Cancer. *Cancers (Basel)*. 2019. vol. 11.

38. Kannan L, Ramos M, Re A, et al. Public data and open source tools for multi-assay genomic investigation of disease. *Briefings in Bioinformatics*. 2016;17(4):603-615. doi:10.1093/bib/bbv080

39. Lee D, Park Y, Kim S. Towards multi-omics characterization of tumor heterogeneity: a comprehensive review of statistical and machine learning approaches. *Briefings in Bioinformatics*. 2021;22(3):bbaa188. doi:10.1093/bib/bbaa188

40. Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biology*. 2018/12/03 2018;19(1):211. doi:10.1186/s13059-018-1593-z

41. Saadatpour A, Lai S, Guo G, Yuan GC. Single-Cell Analysis in Cancer Genomics. *Trends Genet*. Oct 2015;31(10):576-586. doi:10.1016/j.tig.2015.07.003

42. Tamaddon M, Azimzadeh M, Gifani P, Tavangar SM. Single-cell transcriptome analysis for cancer and biology of the pancreas: A review on recent progress. *Front Genet*. 2023;14:1029758. doi:10.3389/fgene.2023.1029758

43. Schelker M, Feau S, Du J, et al. Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat Commun*. Dec 11 2017;8(1):2032. doi:10.1038/s41467-017-02289-3

44. Sutton GJ, Poppe D, Simmons RK, et al. Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nature Communications*. 2022/03/15 2022;13(1):1358. doi:10.1038/s41467-022-28655-4

45. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. May 2015;12(5):453-7. doi:10.1038/nmeth.3337

46. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*. 2017/11/15 2017;18(1):220. doi:10.1186/s13059-017-1349-1

47. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. Nov 13 2017;6doi:10.7554/eLife.26476

48. Chakravarthy A, Furness A, Joshi K, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun*. Aug 13 2018;9(1):3220. doi:10.1038/s41467-018-05570-1

49. Zhang Z, Wiencke JK, Kelsey KT, Koestler DC, Christensen BC, Salas LA. HiTIMED: hierarchical tumor immune microenvironment epigenetic

deconvolution for accurate cell type resolution in the tumor microenvironment using tumor-type-specific DNA methylation data. *Journal of Translational Medicine*. 2022-11-08 2022;20(1)doi:10.1186/s12967-022-03736-6

CHAPTER 1

Unveiling the heterogeneity of pancreatic cancer: bulk DNA methylation data analysis reveals association with somatic KRAS mutations and distinct patterns of immune cell microenvironment composition

Manuscript prepared under the **Epigenomics** (*Print ISSN: 1750-1911; Online ISSN: 1750-192X*) journal submission guideline.

Title

Unveiling the heterogeneity of pancreatic cancer: bulk DNA methylation data analysis reveals association with somatic KRAS mutations and distinct patterns of immune cell microenvironment composition

Barbara Mitsuyasu Barbosa[1], Roberto da Silva Gomes[2], Claudia Aparecida Rainho[1,*]
[1]Department of Chemical and Biological Sciences, Institute of Biosciences, São Paulo State University (UNESP), Botucatu, Sao Paulo, 18618-689, Brazil
[2]Department of Pharmaceutical Sciences, North Dakota State University, Fargo, North Dakota, 58102, USA

*Author for correspondence: claudia.rainho@unesp.br

Abstract

Background: Despite significant progress in the knowledge of pancreatic cancer pathogenesis and progression, and advances in therapeutic strategies, this malignancy remains one of the deadliest tumors. Pancreatic tumors are characterized by low tumor cell counts, intense desmoplasia, and complex interactions among various cell types. Unveiling the tumor immune microenvironment (TIME) composition and its interactions are crucial for developing new treatment strategies and a better understanding of this disease. **Methods:** We integrated genome-wide DNA methylation, KRAS driver mutations, and transcriptome data from the TCGA Pancreatic Cancer (TCGA-PAAD) cohort to identify distinct subgroups based on immune cell composition. Bioinformatic tools were used for DNA methylation profiling, DNA methylation age, hierarchical tumor deconvolution of cell types, and Weighted Gene Correlation Network Analysis (WGCNA). **Results:** Global DNA methylation analysis revealed two distinct profiles in pancreatic cancer. Notably, group 2 (n=95 tumor samples) exhibited significantly higher tumor purity ($p<0.0001$) and a greater frequency of *KRAS* mutations ($p<0.0001$) compared to group 1 (n=87 tumor samples). The hierarchical DNA methylation-based tumor deconvolution identified three immune profiles: the first cluster, characterized by a lower abundance of non-tumoral cells, was designated hypo-inflamed (immune deserted), cluster 2 displayed a predominance of myeloid cells (myeloid-enriched), and cluster 3 enriched with lymphoid cells, notably T cells (lymphoid-enriched). WGCNA associated specific gene modules with predicted immune groups. KEGG and Reactome enrichment of WGCNA modules provided robust support for the proposed classification. **Conclusions:** This integrative approach holds considerable promise for stratifying pancreatic cancer patients and uncovering potential prognostic biomarkers and therapeutic targets to enhance the outcomes of this challenging disease.

Keywords: tumor immune microenvironment, tumor purity, DNA methylation age, DNA methylation-based deconvolution, weighted co-expression network, cancer subtypes

Introduction

The tumor microenvironment (TME) comprises a complex and heterogenic network containing tumor cells, stromal cells, immune cells, and the extracellular matrix. Interactions among these components regulate key aspects of tumor biology, including proliferation, invasion, and metastasis [1, 2]. Furthermore, genomic and transcriptomic analyses have revealed distinct TME signatures across different cancer types, underscoring the tissue-specific nature of TME interactions and their implications for cancer progression and therapeutic response [3-8]. Stromal cells, such as cancer-associated fibroblasts (CAFs), can modulate TME composition and communicate with immune cells to influence tumor development. Studies have shown that CAFs can not only promote but also inhibit tumors, adding to the complexity of TME [9, 10]. The immune microenvironment is a critical component of the TME and influences both anti-tumor immunity and immune evasion strategies. It modulates treatment response and patient outcomes, with immunotherapies demonstrating remarkable efficacy in certain types of cancers by harnessing the host immune response against malignant cells [5, 8, 11, 12]. TME heterogeneity is also influenced by the diverse nature of driver mutations within tumors, which are responsible for initiating and maintaining neoplastic growth [13, 14].

In this context, pancreatic cancer poses an even greater challenge. Difficulties in early detection due to the lack of specific signs/symptoms, absence of screening strategies, and aggressive behavior of pancreatic cancer lead to a disease characterized by clinical aggressiveness, advanced staging at diagnosis, and resistance to the available treatment modalities, resulting in therapeutic failure and unfavorable outcome [15]. Morphological and molecular intra- and inter-tumor heterogeneity of pancreatic cancer accompany disease progression. Histologically, this disease 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, and proteomic profiles. The desmoplastic profile within the TME of pancreatic cancer could also pose a problem for drug delivery [16, 17]. Although pancreatic TME usually has an immunosuppressive profile, it varies considerably across different subtypes and undergoes dynamic changes during disease progression [18]. This heterogeneous immune microenvironment was also observed by Li et al., who used an in vivo approach to investigate factors associated with tumor immune heterogeneity and immunotherapy sensitivity. These authors co-inject pancreatic cell clones into immunocompetent mice and identify distinct tumor immune microenvironment (TIME): the non-T-cell-inflamed phenotype was dominant and both quantitative and qualitative features of intratumoral CD8+ T cells were implicated in immunotherapy responses [12].

The integration of profile data at different “omics” levels provided by The Cancer Genome Atlas Project (TCGA, available at https://www.cancer.gov/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 [16]. 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*) [16, 19]. *KRAS* mutations are the most common in pancreatic cancer. Mutational activation of

KRAS occurs in the early stages of pancreatic cancer, before its progression to invasive carcinoma [20, 21]. Apart from its well-established oncogenic functions, emerging evidence suggests that *KRAS* mutations collaborate to the immunosuppressive phenotype of the TME in pancreatic cancer [22]. Recently, DNA methylation profiling has unveiled distinct subgroups within lung adenocarcinoma. These subgroups exhibit correlations with the presence of a set of oncogenic drivers, including *KRAS*, and distinct immune cell populations [23].

Evaluation of TME conventionally requires image analysis by a pathologist to identify cell subpopulations. In recent years, novel computational methodologies have been developed alongside genomic advancements to determine tumor composition. Tumor deconvolution uses genomic data to infer the complex composition of tumors into their constituent cellular elements. These methods deconvolve the tumor sample using various types of genomic information, such as microarray [24], transcriptomic [25, 26], and methylation data [27]. In this scenario, we reanalyzed publicly available data from The Cancer Genome Atlas (TCGA) to better understand the heterogeneity of pancreatic cancer. To provide new insights into the relationship between the pancreatic cancer methylome and TME, we first investigated the influence of the purity of the tumor sample on DNA methylation clustering and DNA methylation age. Hierarchical tumor immune microenvironment deconvolution based on DNA methylation was performed to predict the cell type composition of the TME following PAM (partition around medoids) clustering to classify tumor samples based on predicted tumor cell fractions. Finally, we compared the TIME patterns with specific expression of gene modules and co-regulated networks.

Materials & Methods

A flowchart summarizing the study design and methodology employed for the identification of tumor-infiltrating immune cells from methylation data and the construction of a weighted gene correlation network from RNA-Seq is illustrated in Figure 1.

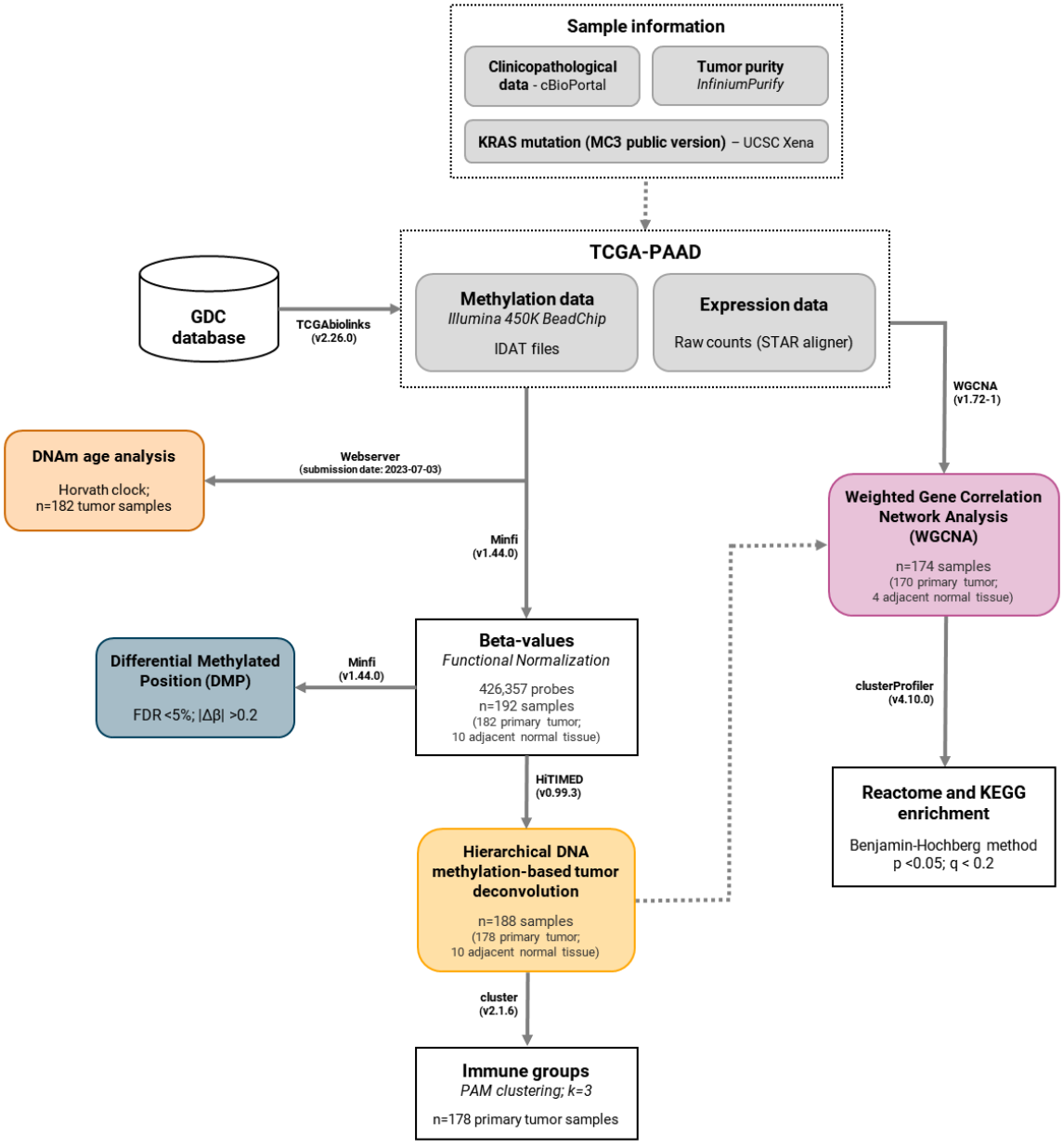


Figure 1. Flowchart summarizing the study design and the methodology employed for the identification of tumor-infiltrating immune cells from methylation data and the construction of a weighted gene correlation network from RNA-Seq in pancreatic cancer.

DNA methylation profiling in pancreatic cancer

Data acquisition and pre-processing

DNA methylation data from the TCGA Pancreatic Cancer (TCGA-PAAD) cohort, sourced from The Cancer Genome Atlas Research Network (https://www.cancer.gov/tcga), was selected for this study. These data were based on the Infinium Human Methylation 450K BeadChip platform. Raw Intensity Data (IDAT) files were retrieved from the Genomic Data Commons (GDC) database with TCGAbiolinks R package (v2.26.0) [28]. This cohort comprised 184 primary pancreatic tumors, 1 metastatic tumor, and 10 adjacent normal tissue samples. Data were preprocessed using the Minfi Bioconductor R package (v1.44.0) [29] and functional normalization was applied as recommended [30]. Following functional normalization, one tumoral sample was excluded because it failed to pass quality control (QC) parameters. Subsequently, cross-reactive probes that can potentially hybridize to multiple genomic locations due to sequence similarity [31]; probes located on the X or Y chromosome, and polymorphic CpGs that overlapped Single Nucleotide Polymorphisms (SNPs) with a minor allele frequency (MAF<0.05) based on dbSNP 137, using the annotation file for the 450K array in UCSC hg19 assembly (GRCh37 build reference sequence) [32] were filtered. In addition, probes with poor quality signals (detection p-value<0.05) and low bead count (<3) in more than 5% of the samples were also removed. DNA methylation values, described as beta values, were then calculated for 192 tissue samples, comprising 182 pancreatic tumors and 10 normal samples (Supplementary Table 1).

In parallel, clinicopathological information was retrieved from cBioPortal (https://www.cbioportal.org/). The MC3 public version mutation data [33] for *KRAS* mutation in tumor samples was obtained from the University of California at Santa Cruz

(UCSC) Xena repository (https://xena.ucsc.edu/). The tumor purity of TCGA-PAAD samples was estimated by *InfiniumPurify* [34].

Differentially methylated positions analysis

Differentially methylated CpG positions (DMPs) between tumor and normal samples were detected by using the Minfi R package (v1.44.0) [29]. Probes meeting the criteria of a False Discovery Rate (FDR) of less than 5% and $|\Delta\beta|>0.2$ were considered statistically significant. Significant probes were then annotated using the manifest file [35]. The probes resulting from the DMP analysis were annotated to UCSC Reference Sequencing (RefSeq) genes using the annotation file for the 450K array in UCSC hg19 assembly [32]. DMPs were categorized based on their CpG content and neighborhood context (shore, shelf, open sea, and island) and functional genomic distribution (3'UTR, body, 1st exon, 5'UTR, TSS1500, and TSS200).

DNA methylation age analysis

DNA methylation (DNAm) age for each tumor sample within the cohort was determined using the Horvath clock, a multi-tissue algorithm that employs a weighted average of methylation levels at specific CpG sites throughout the genome to predict epigenetic age directly from IDAT files [36]. Epigenetic age acceleration (EAA) was subsequently defined based on the residual obtained by regressing DNAm age on chronological age. A positive EAA (>0) indicates an acceleration of epigenetic aging, whereas a negative EAA (<0) indicates a deceleration relative to chronological aging (Supplementary Figure 1A).

Hierarchical DNA methylation-based tumor deconvolution

The cell type proportions within each sample were determined using the HiTIMED R package (v.0.99.3) [37]. This analysis utilized beta values from the TCGA-PAAD dataset (Supplementary Table 1) using a 6-layer-tumor-type-specific model. A total of seventeen cell types, representing the three major components of the tumor microenvironment were identified: tumor; angiogenic (endothelial, epithelial, and stromal), and immune (basophil, eosinophil, neutrophil, monocyte, dendritic, Natural Killer (NK), B naïve, B memory, CD4T naïve, CD4T memory, T regulatory, CD8T naïve, and CD8Tmemory) cells [37]. The cluster R package [38] was utilized for conducting Partitioning Around Medoids (PAM) clustering to identify TIME patterns based on the obtained specific cell fractions.

Gene expression analysis using a network-based approach

Data acquisition and pre-processing

The TCGA-PAAD gene expression quantification cohort was obtained from the Genomic Data Commons (GDC) database with TCGAbiolinks R package (v2.26.0) [28] as raw counts aligned to the human genome with STAR RNA-seq aligner [39]. The dataset comprises the same 182 samples (178 primary solid tumor, and 4 adjacent normal issue) and 60,660 distinct transcripts.

The data was preprocessed according to the Weighted Correlation Network Analysis (WGCNA) recommendations [40]. Initially, low-count genes (<15) in more than 75% of samples were excluded. Subsequently, the raw counts were normalized by the median ratio method followed by variance stabilizing transformations (VST) using the DESeq2 R package (v1.41.10) [41]. Quality Control (QC) employing the *goodSamplesGenes*

function from the WGCNA R package (v1.72-1) [40] deemed all samples and genes acceptable. Outlier samples were identified using two distinct methodologies: Hierarchical Clustering and Principal Component Analysis (PCA) and removed.

Weighted Gene Correlation Network Analysis (WGCNA)

The WGCNA R package (v.1.72-1) [40] *blockwiseModules* function was utilized to construct a weighted co-expression network where genes are represented as nodes, and the connectivity between them is established based on pairwise correlations derived from their expression profiles [42]. Aiming to obtain a scale-free topology, a power of 14 was chosen for the soft-threshold to create an adjacency matrix. Genes were grouped in different modules according to their similar expression patterns using a topological overlap matrix (TOM) of dissimilarity measures and the branches of a dendrogram derived from average linkage hierarchical clustering. Then, identified modules were correlated with sample traits (sample type and deconvolution-derived immune groups) using Pearson correlation coefficients. Key driver genes in interesting modules were found by intramodular connectivity.

Enrichment analysis of WGCNA modules

Entrez gene identifiers (Entrez ID) for the UCSC RefSeq genes of each module were retrieved with the biomaRt R package (v2.58.0) [43, 44] using an annotation database [45]. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathway enrichment analysis were performed for selected WGCNA modules with ClusterProfiler

R package (v.4.10.0) [46]. KEGG and Reactome enrichment were analyzed using the Benjamini-Hochberg (BH) method to adjust p-values (p-value<0.05; q-value<0.2).

Statistical analysis

EAA was calculated based on the residual obtained through linear regression of DNAm age on chronological age. Fisher's exact tests were conducted to assess the association between DNA methylation-based groups and the clinical and molecular characteristics of the samples. ANOVA with Tukey's post-hoc multiple comparisons test was employed to assess the differences in HiTIMED-estimated cell subpopulations among PAM clustering-derived groups. Overall survival of PAM-clusters was calculated using Mantel-Cox test. The statistical analysis was performed using GraphPad Prism software (v10.2.1) (*:p≤0.05; **:p≤0.01; ***:p≤0.001; ****:p≤0.0001).

Results

Global DNA methylation profiling identifies different patterns in pancreatic cancer

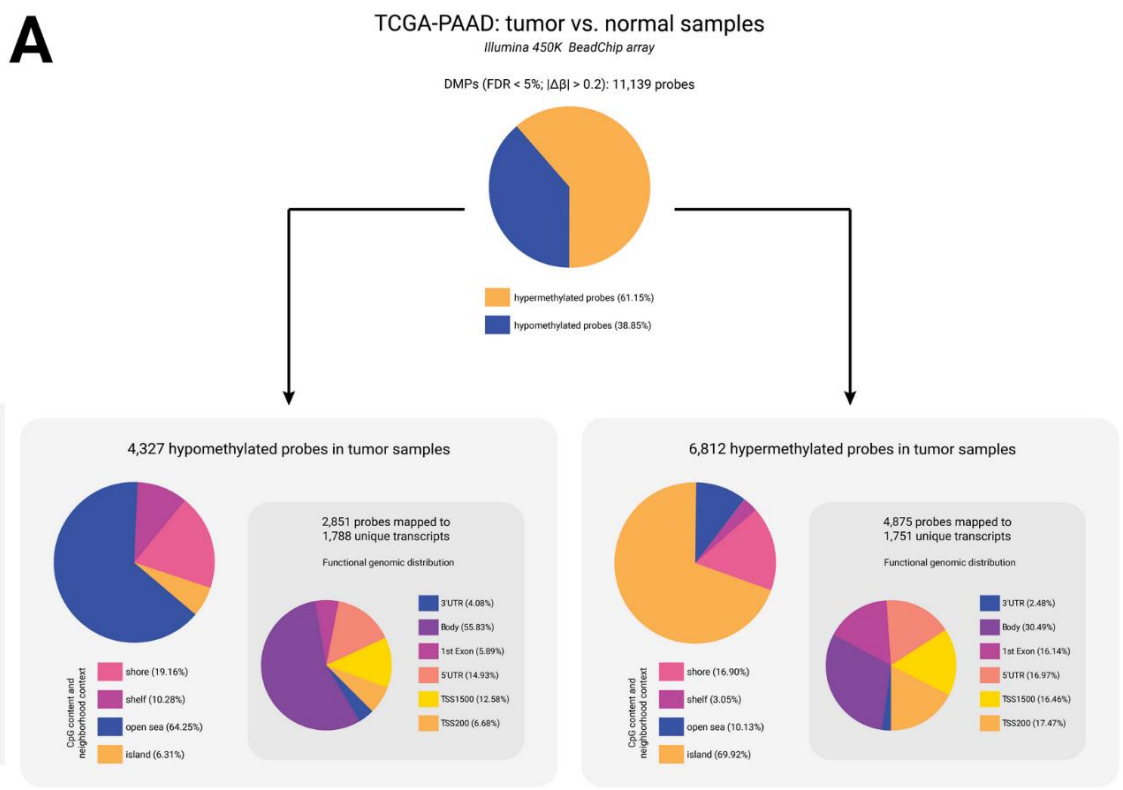
DMP analysis (FDR<5%; |Δβ|>0.2) unveiled different methylation patterns in pancreatic cancer when compared to their normal counterparts, identifying a total of 11,139 differently methylated probes. These probes were annotated to 3,328 unique UCSC gene symbols. Figure 2A shows the distribution of hypo- and hypermethylated probes according to their CpG content and neighborhood context, and functional genome distribution. Hypomethylated probes (n=4,327) were concentrated in open sea (64.25%) and gene body (55.83%) regions, while hypermethylated probes (n=6,812) were mapped in CpG islands (69.92%) and promoter regions (33.93%) (Supplementary Table 2).

Unsupervised hierarchical clustering of DMPs revealed two major groups with differential distribution of molecular and clinical parameters [47]. Group 1 consists of 87 tumor and 9 normal samples, while Group 2 predominantly comprises tumor samples, with only 1 normal sample among the 96 total samples. These methylation profiles were associated with distinct molecular and tumor tissue features. Group 2 displayed higher tumor purity (p<0.0001) and a higher frequency of *KRAS* mutations (p<0.0001) compared to Group 1 (Figure 2B). *KRAS* mutation status was also associated with positive epigenetic age acceleration (p=0.0128; Supplementary Figure 1B). Furthermore, overall survival outcomes differed significantly between the clusters: Group 2 has a lower survival rate (p=0.0046) (Figure 2B; Table 1).

Table 1. Demographic, clinical, and tumor sample features of TCGA-PAAD cohort used to retrieve DNA methylation data.

		Group 1		Group 2		
Variables	**Samples total**	N	(%)	N	(%)	*p*-value*
Gender	182	87	*(100%)*	*95*	*(100%)*	0,6553
Female		41	*(47.13%)*	41	*(43.16%)*	
Male		46	*(52.87%)*	54	*(56.84%)*	
Age (years)		87	*(100%)*	*95*	*(100%)*	0,6362
<50		11	*(12.64%)*	*9*	*(9.47%)*	
>50		76	*(87.36%)*	*86*	*(90.53%)*	
Overall Survival	182	*87*	*(100%)*	*95*	*(100%)*	**0,0046**
Alive		50	(57.5%)	34	(35.8%)	
Dead		37	(42.5%)	61	(64.2%)	
KRAS mutation	173	*80*	*(100%)*	*93*	*(100%)*	**<0,0001**
Presence		30	(37.5%)	84	(90.3%)	
Absence		50	(62.5%)	9	(9.7%)	
Tumor purity	182	87	*(100%)*	*95*	*(100%)*	**<0,0001**
Low		46	(52.9%)	2	(2.1%)	
Medium-low		27	(31.0%)	12	(12.6%)	
Medium-high		13	(14.9%)	37	(38.9%)	
High		1	(1.1%)	44	(46.3%)	

* Fisher's exact test


A
TCGA-PAAD: tumor vs. normal samples
Illumina 450K BeadChip array
DMPs (FDR < 5%; |Δβ| > 0.2): 11,139 probes
hypermethylated probes (61.15%)
hypomethylated probes (38.85%)
4,327 hypomethylated probes in tumor samples
CpG content and neighborhood context
shore (19.16%)
shelf (10.28%)
open sea (64.25%)
island (6.31%)
2,851 probes mapped to 1,788 unique transcripts
Functional genomic distribution
3'UTR (4.08%)
Body (55.83%)
1st Exon (5.89%)
5'UTR (14.93%)
TSS1500 (12.58%)
TSS200 (6.68%)
6,812 hypermethylated probes in tumor samples
CpG content and neighborhood context
shore (16.90%)
shelf (3.05%)
open sea (10.13%)
island (69.92%)
4,875 probes mapped to 1,751 unique transcripts
Functional genomic distribution
3'UTR (2.48%)
Body (30.49%)
1st Exon (16.14%)
5'UTR (16.97%)
TSS1500 (16.46%)
TSS200 (17.47%)

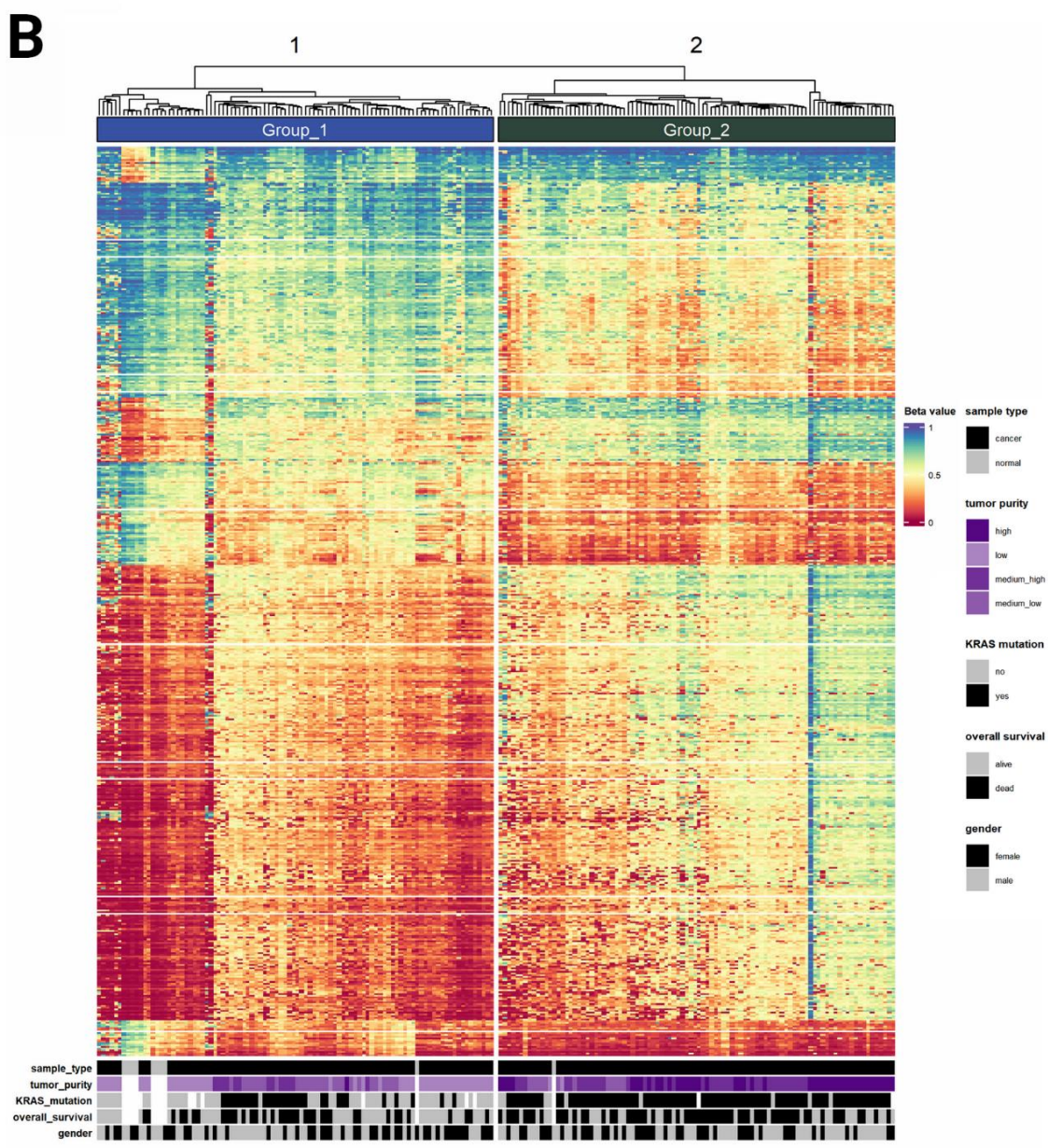

B
1
2
Group_1
Group_2
Beta value
1
0.5
0
sample type
cancer
normal
tumor purity
high
low
medium_high
medium_low
KRAS mutation
no
yes
overall survival
alive
dead
gender
female
male
sample_type
tumor_purity
KRAS_mutation
overall_survival
gender

Figure 2. Differentially methylated positions (DMPs) in pancreatic cancer between tumor and normal samples (FDR<5%; $|\Delta\beta|>0.2$). **A.** Frequency distribution of DMPs according to their CpG content and neighborhood context (shore, shelf, open sea, and island) and functional genomic distribution (3'UTR, body, 1st exon, 5'UTR, TSS1500, and TSS200). **B.** Heatmap of DNA methylation profile of DMPs. Sample's clinicopathological and molecular features are shown at the bottom.

Hierarchical DNA methylation-based tumor deconvolution reveals distinct patterns of tumor immune microenvironment in pancreatic cancer

Figure 3A shows heterogeneous patterns of estimated cell population composition across 178 TCGA-PAAD tumor samples. The hierarchical distribution of HiTIMED cell fractions is also depicted in the figure (Supplementary Table 3). PAM clustering distinguishes three clusters characterized by their distinct cellular compositions. Cluster 1 exhibits a higher proportion of tumor cells in comparison to clusters 2 and 3 (Figure 3B). While the proportion of tumor cells is comparable between clusters 2 and 3, their composition of non-tumoral cells differs (Figure 3B and 3C). Our analysis unveiled distinct distributions of angiogenic components across the identified clusters. Specifically, cluster 2 exhibited higher levels of epithelial and stromal cells compared to clusters 1 and 3. Conversely, cluster 1 displayed elevated proportions of endothelial cells compared to clusters 2 and 3. Regarding immune components, we identified distinct enrichment of myeloid and lymphoid cell populations within PAM-clusters. Cluster 2 exhibited higher levels of basophils, eosinophils, neutrophils, monocytes, and dendritic cells compared to clusters 1 and 3 within the myeloid cell subset. Additionally, dendritic cell proportions differed between clusters 1 and 3. Regarding lymphoid cells, cluster 1 displayed lower levels of natural killer and T regulatory cells compared to clusters 2 and 3. Furthermore, cluster 3 demonstrated elevated proportions of B memory, CD4T memory, and CD8T naïve cells compared to clusters 1 and 2. CD8T memory cells exhibited a similar pattern

to dendritic cells, with cluster 3 having higher levels than clusters 1 and 2, while cluster 2 also displayed higher levels than cluster 1. Notably, no significant differences were observed in B naïve and CD4T naïve cell populations among the three clusters (Figure 3C). Thus, the three identified PAM-clusters were named according to their proportion of immune cell fractions. Cluster 1, characterized by a lower abundance of non-tumoral cells, was designated hypo-inflamed (immune deserted). Cluster 2 displayed a predominance of myeloid cells and was labeled myeloid-enriched. Cluster 3, enriched with lymphoid cells, notably T cells, was named lymphoid-enriched. Interestingly, no difference (p-value=0.7749) was observed in the overall survival when comparing the three PAM-clusters despite their different cell composition profiles (Figure 3D).

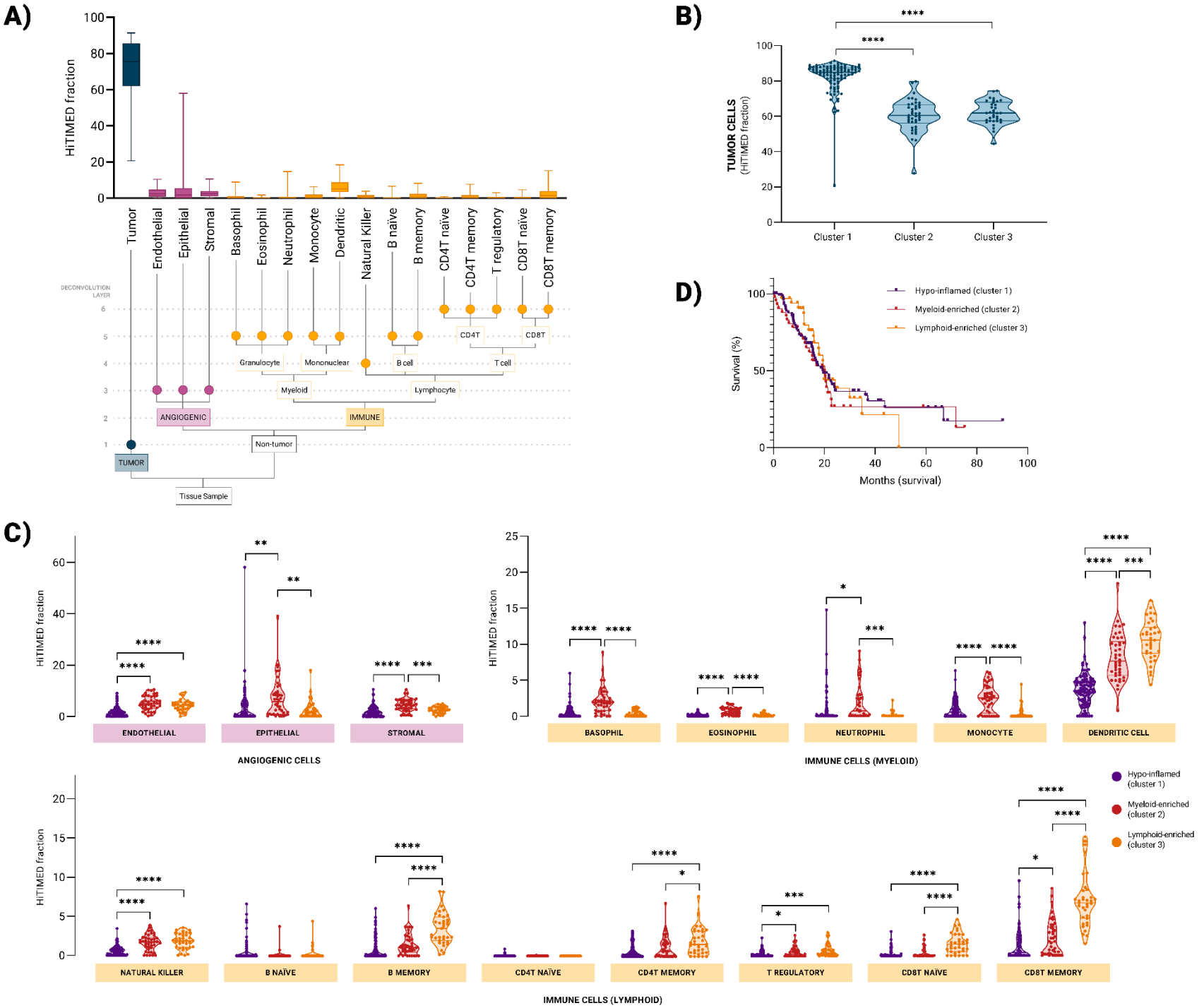


Figure 3. HiTIMED DNA methylation-based tumor deconvolution. **A.** Distribution of cell composition in TCGA-PAAD tumor samples (n=178). The structure of the pancreatic cancer hierarchical model used to deconvolve tumor samples is represented below the graph showing the six hierarchical layers. **B.** Tumor cells distribution of the three clusters. Clusters were distinguished using HiTIMED cell fractions PAM clustering. **C.** Comparison of non-tumor cell populations in PAM clustering-derived groups of TCGA-PAAD. ANOVA with Tukey's post-hoc multiple comparisons test was employed to assess the differences between clusters (*:p≤0.05; **:p≤0.01; ***:p≤0.001; ****:p≤0.0001). **D.** Overall survival of pancreatic cancer patients stratified according to PAM-clusters (Mantel-Cox test; p-value: 0.7749).

Weighted Correlation Network Analysis (WGCNA) uncovers group of genes associated with immune groups in pancreatic cancer

In our analysis of the TCGA-PAAD RNA-seq dataset, WGCNA analysis revealed the presence of 14 modules, representing consensus clusters (Figure 4A). The correlation between each module and immune groups is depicted in Figure 4B. Subsequent enrichment analyses were conducted to elucidate the biological pathways and functions associated with these co-expressed gene modules. Notably, our investigation identified module (ME) 10 as pivotal for the three predicted immune phenotypes, showing a negative correlation with the immune deserted cluster, and a positive correlation with the immune-enriched clusters. Furthermore, KEGG enrichment analysis has elucidated that the network comprises genes closely associated with pancreatic function such as pancreatic secretion, protein digestion and absorption, and maturity-onset diabetes of the young (Figure 4C).

Genes in ME1 are negatively correlated with the hypo-inflamed and positively correlated with the lymphoid-enriched clusters. ME6 displayed a negative correlation with the myeloid-enriched and a positive correlation with the lymphoid-enriched clusters. Similarly, ME11 exhibited a negative correlation with the hypo-inflamed and a positive correlation with the lymphoid-enriched, while ME12 demonstrated a negative correlation with the hypo-inflamed and a positive correlation with the myeloid-enriched (Figure 4B).

Further Reactome enrichment analysis revealed distinct biological processes and pathways associated with each network. ME1 was enriched in processes related to immunoregulation pathways such as immunoregulatory interactions between a lymphoid and a non-lymphoid cell, neutrophil degranulation, and translocation of ZAP-70 to

immunological synapse (Figure 4D). ME6 and ME11 are enriched in signaling and regulatory pathways, including antigen processing: ubiquitination & proteasome degradation, transcriptional regulation of white adipocyte differentiation, and PPARA activate gene expression for ME6; and classical antibody-mediated complement activation, FCGR activation, initial triggering of complement, role of phospholipids in phagocytosis, and regulation of complement cascade for ME11 (Figures 4E and 4F). Finally, ME12 exhibited enrichment in processes related to cellular signaling and cytoskeletal regulation pathways such as cell-extracellular matrix interactions, and RHO GTPases activate CIT, ROCKs, PAKs, and PKNs (Figure 4G).

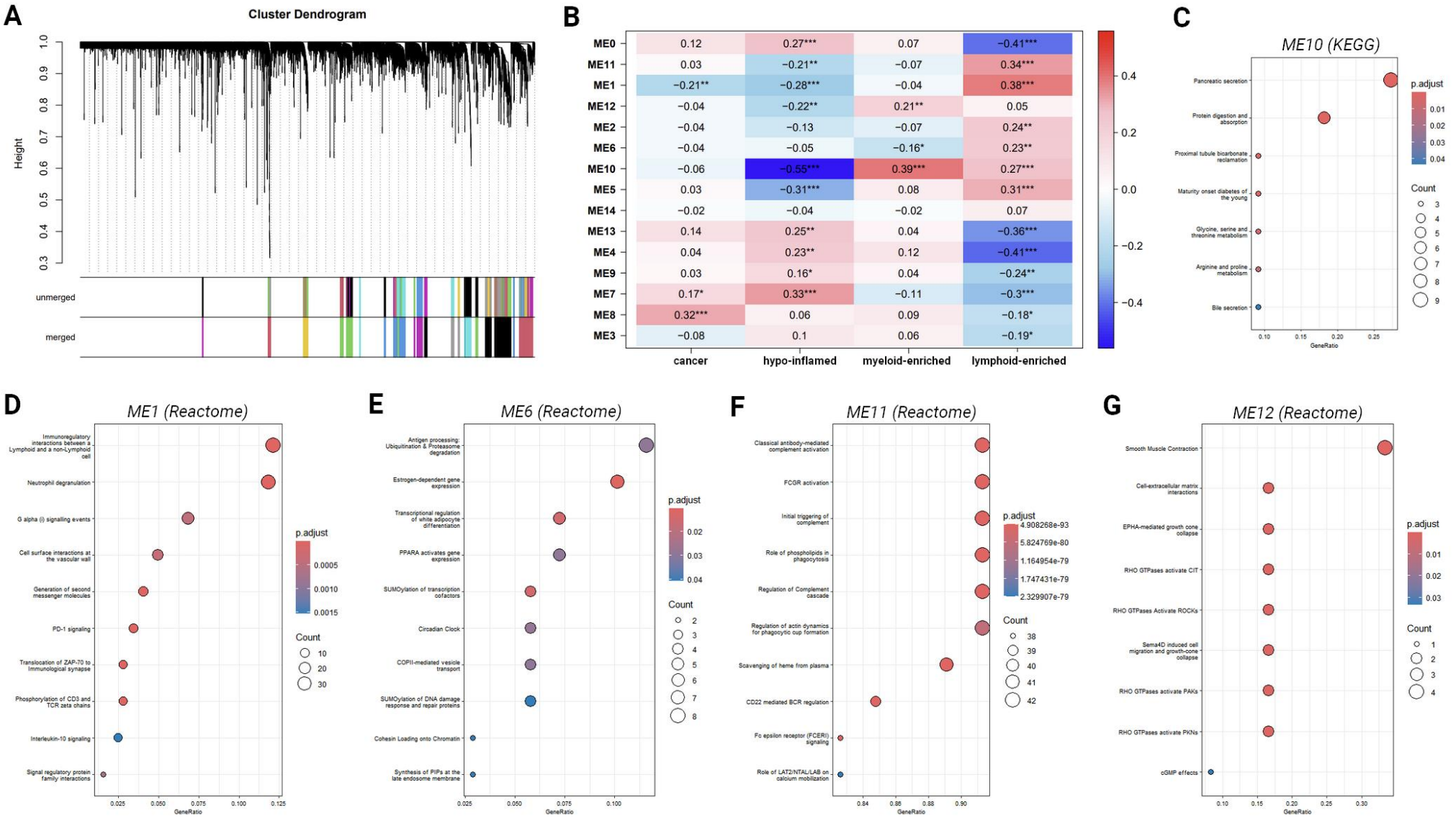


Figure 4. Visual representation of Weighted Gene Correlation Network Analysis (WGCNA) obtained from the TCGA-PAAD RNA-seq dataset. **A.** Cluster dendrogram based on topological overlap distances in gene expression patterns of primary solid tumor and adjacent normal samples associated with pancreatic cancer. Each color displayed in the bar below the dendrogram represents a module, which signifies a cluster of genes. Genes displayed in white indicate a lack of similar expression patterns resulting in their exclusion from any specific module. Clusters within the unmerged bar represent the diverse patterns of expression initially identified. Unmerged clusters were combined according to their similarity, resulting in the final clusters represented in the merged bar. **B.** Correlation analysis of Module Eigengene with traits – disease status (cancer vs. normal samples) or immune tumor groups (hypo-inflamed; myeloid; and lymphoid): the correlation values between module Eigengene values and traits are depicted numerically within the cells. The color gradient of the cells corresponds to the correlation values, as illustrated by the scale bar on the right. Asterisks within the cells signify the level of statistical significance. **C.** Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Analysis for module 10; **D.** Reactome Pathway Enrichment Analysis of ME1; **E.** ME6; **F.** ME11; and **G.** ME12 genes.

Discussion

Unraveling intratumor heterogeneity at the molecular and cellular levels is crucial for a better understanding of pancreatic cancer biology as well as the impact of the complex interactions in outcomes and therapy responses of this tumor type, which is known for having an immunosuppressive tumor microenvironment (TME). The molecular and histological landscape of the disease varies widely within individual tumors and the diverse cell populations present in TME influence prognosis and clinical outcomes [18, 48]. At the molecular level, distinct tumor cell subpopulations can be identified within a single tumor because of a myriad of factors, such as clonal evolution, and acquired genetic and epigenetic alterations [49, 50]. Furthermore, tumor tissues also comprise non-tumor cells, including stromal, vascular, and immune components, thus conferring cellular heterogeneity [18]. The heterogeneity of epithelial cancer cells observed in precursor lesions of pancreatic cancer intensifies as the disease advances toward invasive carcinoma [51]. Additionally, non-immune cell populations, including the immune cells. also display intra- and inter-tumor heterogeneity distribution in pancreatic cancer [52].

Herein, we adopt a multidimensional methodology to explore this tumor heterogeneity by exploring DNA methylation profiling, DNA methylation age, RNA-seq, and mutational data to characterize and classify 178 TCGA-PAAD samples based on their predicted TME composition. We identified three tumor microenvironments (TMEs) with distinct immune profiles. In addition, we observed an association between oncogenic mutations in KRAS and positive epigenetic age acceleration.

Traditional bulk analysis techniques, encompassing genomic, transcriptomic, and epigenomic approaches, have provided valuable insights into pancreatic cancer biology. The unsupervised clustering of DMPs (FDR<5%; $|\Delta\beta|>0.2$) revealed two different methylation profiles. We observed a correlation between tumor purity and methylation profile, with tumors exhibiting lower purity displaying distinct methylation patterns compared to those with higher purity. Intriguingly, the group with the highest tumor purity also exhibited a higher frequency of *KRAS* mutations, a driver mutation commonly associated with pancreatic cancer (Figure 2B; Table 1). This association suggests a potential link between tumor purity and the detection of *KRAS* mutations. It is plausible that in samples with lower purity, the abundance of non-tumor cells may dilute the presence of tumor-derived genetic material, potentially impacting the detection sensitivity for *KRAS* mutations. Consequently, the observed enrichment of *KRAS* mutations in samples with higher tumor purity may reflect a greater representation of tumor cells capable of harboring this driver mutation.

However, traditional bulk methods often do not consider the diversity within the TME. The advent of single-cell and spatial analysis technologies enabled the characterization of distinct cell populations and their spatial organization within the tumor [48, 49]. Moreover, these technologies have facilitated the identification of gene expression patterns within specific cell populations that are associated with treatment responses in pancreatic cancer [50].

Integrative approaches provide valuable insights into the prognostic significance of TIME profiles in PAAD, holding promise for enhancing our understanding of the microenvironment and its impact on pancreatic cancer treatment [51-53]. A recent study

integrated data from whole tissue slides analyzed by pathologists with next-generation sequencing. While no distinct mutation pattern was identified between short- and long-term survival in PAAD, differences in the TIME composition were noted. Notably, macrophage infiltration was associated with poorer overall survival, whereas patients with long-term survival showed increased CD4+ T cell infiltration and predominantly inactivated stromal profiles [54].

KRAS driver mutations have also been implicated in shaping the epigenetic landscape of tumors, albeit not on a global scale of DNA methylation. A comprehensive analysis involving 47 cell lines underscored the significant impact of tumor cell type on DNA methylation profiling, overshadowing the influence of *KRAS* driver mutations. However, investigations into 11 pancreatic cancer isogenic cell line pairs revealed that *KRAS* knockdown induced cell-line-specific alterations in DNA methylation, suggesting a role for *KRAS* in mediating epigenetic regulation within the tumor microenvironment [55].

In a recent paper, lung cancer samples were categorized into hot or cold immune phenotypes using deconvolution results obtained with MethylCIBERSORT. Consistent with our results, lower DNA methylation age was also correlated with the presence of oncogenic drivers in lung cancer [23].

Our findings using cell-type deconvolution of DNA methylation data also reveal distinct patterns of methylation in pancreatic cancer, and we were able to identify three clusters showing distinct TIME composition. These methylation patterns revealed heterogeneous cellular compositions reflective of the tumor microenvironment. We observed pancreatic tumors with low immune infiltration (hypo-inflamed) and tumors exhibiting different immune infiltrates (myeloid-enriched and lymphoid-enriched),

characterized by distinct immune lineage cell populations. Remarkably, the cellular subtypes predicted by deconvolution closely align with the characterization of the tumor immune microenvironment using cell-type specific markers and immunohistochemistry imaging [56]. This comprehensive analysis highlights the inherent heterogeneity of immune profiles within pancreatic tumors, which would be laborious and time-consuming to fully capture through routine clinical assessment by a pathologist. The deconvolution approach provides a powerful, high-throughput alternative to directly profile the complex immune landscape of these tumors. Nonetheless, it is imperative to acknowledge the limitations inherent in our study, notably the absence of matched histopathological analysis to confirm the predominant TIME pattern on the same PAAD tissues because the study was a reanalysis of a publicly available dataset.

Weighted Correlation Network Analysis (WGCNA) uncovered distinct gene modules associated with immune subtypes in pancreatic cancer, revealing the underlying tumor-immune interactions. Of particular significance are ME1, ME6, ME10, ME11, and ME12, which exhibit distinct gene expression patterns correlating with different immune profiles. Remarkably, genes within ME10 demonstrate associations with all three immune groups predicted in our study. KEGG enrichment analysis elucidates their relevance to pancreatic function (Figure 4C). This observation supports the validity of our analysis, underscoring its appropriateness given the utilization of pancreatic cancer samples.

Cluster 2 is enriched with cells of myeloid lineage (myeloid-enriched), which are the main hematopoietic cells in the human body. The Reactome pathways enriched in module 12 may be involved in the cytoskeletal rearrangements necessary for myeloid cell migration and phagocytosis. For instance, cell-extracellular matrix (ECM) interactions are

crucial for myeloid cells such as macrophages and neutrophils. These interactions influence the migration, adhesion, and pro-inflammatory activities of myeloid cells within the tumor by remodeling the ECM [57]. The RHO family of small GTPases are key regulators of the actin cytoskeleton in myeloid cells. Macrophages can promote angiogenesis within tumors by secreting factors such as Sema4D and VEGF, which activate RHO GTPases [58, 59]. Activation of these RHO GTPases and their downstream effectors like ROCKs, PAKs, and PKNs triggers cytoskeletal rearrangements that drive myeloid cell migration and infiltration into the tumor [60, 61].

Cluster 3 displayed a predominance of lymphoid cells (lymphoid-enriched), notable CD4T memory, CD8T naïve, and CD8T memory cells (Figure 3C). Upon activation by antigens, naïve T cells become effector cells that target and destroy cancer cells as part of the immediate immune response. After tumor cells are cleared, memory T cells remain vigilant within the immune system, ready to proliferate and respond more efficiently if the same antigen reappears [62]. The enrichment terms in modules positively correlated with the lymphoid-enriched cluster (ME1, ME6, and ME11) support this, as immunoregulatory interactions between lymphoid and non-lymphoid cells are crucial for the activation and function of T cells, which interact with antigen-presenting cells [62]. Processes such as the translocation of ZAP-70 to the immunological synapse and subsequent phosphorylation events of CD3 and TCR zeta chains are key events in T cell activation and signaling pathways [63]. The PD-1 signaling pathway, represented by programmed cell death protein 1 (PD-1), is a fundamental inhibitory receptor universally expressed by activated T cells [64, 65]., Pathways essential for antigen processing and presentation alongside CD22-mediated B cell receptor (BCR) regulation are pivotal for adaptive

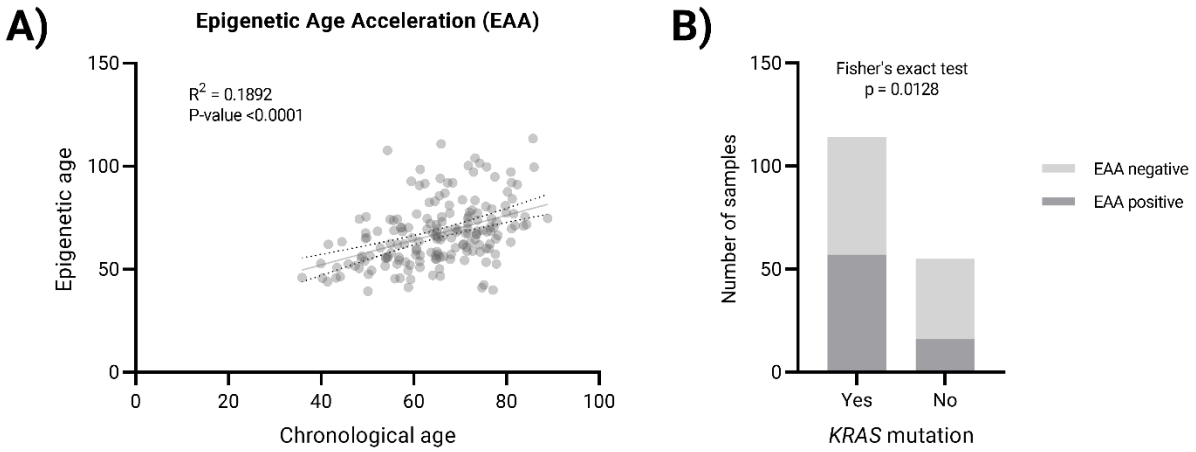
immunity mediated by lymphoid lineage cells [66, 67]. Additionally, classical antibody-mediated complement activation, initial triggering of complement, and regulation of the complement cascade support the action of effector T cells by facilitating tumor cell destruction and promoting inflammation to recruit additional immune cells [68].

Conclusions

In summary, our study provides significant insights into the heterogeneity of TME in pancreatic cancer. By identifying distinct DNA methylation patterns, we categorized the pancreatic cancer samples into three distinct TIME patterns, which are similar to the proposed histological-based PAAD stratification. This classification approach offers promising prospects for the prognostic classification of patients with pancreatic cancer based on their TIME patterns. The identification of gene co-expression networks enriched with immune-related processes serves as a robust validation of our results of immune cell deconvolution of bulk DNA methylation analysis. To fully elucidate the functional significance of these networks in modulating tumor immunity and influencing clinical outcomes, further investigations are required. This integrative approach holds considerable promise not only for the stratification of pancreatic cancer patients but also for unveiling potential prognostic biomarkers and therapeutic epi-targets to enhance the outcomes of this challenging disease.

Supplementary Material

Supplementary figure



Supplementary Figure 1. A. Epigenetic age acceleration (EAA) is defined based on the residuals from the linear regression of DNAm age on chronological age. **B.** *KRAS* mutation status and EAA exhibit a significant relationship (Fisher's exact test; p=0.0128).

Supplementary tables

Supplementary Table 1. Beta values of The Cancer Genome Atlas (TCGA) Pancreatic Cancer (PAAD) cohort. DNA methylation data was processed by Functional Normalization utilizing the Minfi Bioconductor R package.

Available at https://docs.google.com/spreadsheets/d/1p3uGr5oAR6l88YZ9fHMQYprgNsFwGkBy/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

Supplementary Table 2. A. List of 4,327 hypomethylated (FDR < 5%; $\Delta\beta$ <-0.2) and **B.** 6,812 hypermethylated probes (FDR < 5%; $\Delta\beta$ > 0.2) in tumor samples of The Cancer Genome Atlas (TCGA) Pancreatic Cancer (PAAD) cohort.

Available at https://docs.google.com/spreadsheets/d/1EBDks-kixp9DIevL9v1Z9PHhvbYpJtlA/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

Supplementary Table 3. DNA methylation-based tumor deconvolution of cell types in The Cancer Genome Atlas (TCGA) Pancreatic Cancer (PAAD) cohort. The deconvolution results are expressed as proportions of cell types, where samples highlighted in orange were unable to undergo further cell type stratification. PAM (Partitioning Around Medoids) clusterization was employed to differentiate the phenotype of three clusters based on the obtained cell fractions. Tumor purity information was obtained from *InfiniumPurify*. *KRAS* status was retrieved from *MC3 public version mutation data*.

Available at https://docs.google.com/spreadsheets/d/1h_TBenYqTRMopGusvOYGakTTRvHcg9S3/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

Author's contributions

BMB, RSG, and CAR conceived and designed the study, interpreted the results, and wrote and revised the manuscript. BMB conducted the bioinformatic analysis. All authors read and approved the final version of the manuscript.

Competing interest disclosure

The authors reported no potential conflict of interest.

Financial disclosure

This study was supported by Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), Finance Code 001. BMB was granted a scholarship from the CAPES within the scope of the Program CAPES-PrInt (process number 88887.310463/2018-00, mobility number 88887.569794/2020-00).

Data availability statement

The results published here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. All data generated or analyzed during this study are included in this published article and its supplementary files.

References

1. Quail, D.F. and J.A. Joyce, Microenvironmental regulation of tumor progression and metastasis. Nat Med, 2013. 19(11): p. 1423-37.

2. Bożyk, A., et al., Tumor Microenvironment-A Short Review of Cellular and Interaction Diversity. Biology (Basel), 2022. 11(6).

3. Iglesia, M.D., et al., Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. JNCI: Journal of the National Cancer Institute, 2016. 108(11): p. djw144.

4. Gentles, A.J., et al., The prognostic landscape of genes and infiltrating immune cells across human cancers. Nature Medicine, 2015. 21(8): p. 938-945.

5. Li, B., et al., Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol, 2016. 17(1): p. 174.

6. Conway, J.W., et al., The effect of organ-specific tumor microenvironments on response patterns to immunotherapy. Front Immunol, 2022. 13: p. 1030147.

7. Maman, S. and I.P. Witz, A history of exploring cancer in context. Nature Reviews Cancer, 2018. 18(6): p. 359-376.

8. Xue, R., et al., Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature, 2022. 612(7938): p. 141-147.

9. Mun, J.Y., et al., Dual Relationship Between Stromal Cells and Immune Cells in the Tumor Microenvironment. Front Immunol, 2022. 13: p. 864739.

10. Zhao, Y., et al., Stromal cells in the tumor microenvironment: accomplices of tumor progression? Cell Death & Disease, 2023. 14(9): p. 587.

11. Wang, L., et al., Hot and cold tumors: Immunological features and the therapeutic strategies. MedComm (2020), 2023. 4(5): p. e343.

12. Li, J., et al., Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy. Immunity, 2018. 49(1): p. 178-193.e7.

13. Sun, X.-x. and Q. Yu, Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. Acta Pharmacologica Sinica, 2015. 36(10): p. 1219-1227.

14. Reiter, J.G., et al., An analysis of genetic heterogeneity in untreated cancers. Nat Rev Cancer, 2019. 19(11): p. 639-650.

15. Rawla, P., T. Sunkara, and V. Gaduputi, Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. World Journal of Oncology, 2019. 10(1): p. 10-27.

16. Raphael, B.J., et al., Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell, 2017. 32(2): p. 185-203.e13.

17. Ho, W.J., E.M. Jaffee, and L. Zheng, The tumour microenvironment in pancreatic cancer — clinical challenges and opportunities. Nature Reviews Clinical Oncology, 2020. 17(9): p. 527-540.

18. Falcomatà, C., et al., Context-Specific Determinants of the Immunosuppressive Tumor Microenvironment in Pancreatic Cancer. Cancer Discov, 2023. 13(2): p. 278-297.

19. Neoptolemos, J.P., et al., Therapeutic developments in pancreatic cancer: current and future perspectives. Nature Reviews Gastroenterology & Hepatology, 2018. 15(6): p. 333-348.

20. Yabar, C.S. and J.M. Winter, Pancreatic Cancer. Gastroenterology Clinics of North America, 2016. 45(3): p. 429-445.

21. Lanfredini, S., A. Thapa, and E. O'Neill, RAS in pancreatic cancer. Biochemical Society Transactions, 2019. 47(4): p. 961-972.

22. Bear, A.S., R.H. Vonderheide, and M.H. O'Hara, Challenges and Opportunities for Pancreatic Cancer Immunotherapy. Cancer Cell, 2020. 38(6): p. 788-802.

23. Guidry, K., et al., DNA Methylation Profiling Identifies Subgroups of Lung Adenocarcinoma with Distinct Immune Cell Composition, DNA Methylation Age, and Clinical Outcome. Clinical Cancer Research, 2022. 28(17): p. 3824-3835.

24. Liebner, D.A., K. Huang, and J.D. Parvin, MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. Bioinformatics, 2014. 30(5): p. 682-9.

25. Im, Y. and Y. Kim, A Comprehensive Overview of RNA Deconvolution Methods and Their Application. Mol Cells, 2023. 46(2): p. 99-105.

26. Chen, B., et al., Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol, 2018. 1711: p. 243-259.

27. Chakravarthy, A., et al., Pan-cancer deconvolution of tumour composition using DNA methylation. Nat Commun, 2018. 9(1): p. 3220.

28. Colaprico, A., et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research, 2015. 44(8): p. e71-e71.

29. Aryee, M.J., et al., Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 2014. 30(10): p. 1363-1369.

30. Fortin, J.-P., et al., Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biology, 2014. 15(11).

31. Chen, Y.-A., et al., Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics, 2013. 8(2): p. 203-209.

32. Hansen, K.D., IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package, 2021. version 0.6.1.

33. Ellrott, K., et al., Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Systems, 2018. 6(3): p. 271-281.e7.

34. Qin, Y., et al., InfiniumPurify: An R package for estimating and accounting for tumor purity in cancer methylation research. Genes & Diseases, 2018. 5(1): p. 43-45.

35. Hansen, K.D. and M. Aryee, IlluminaHumanMethylation450kmanifest: Annotation for Illumina's 450k methylation arrays. R package, 2012. version 0.4.0.

36. Horvath, S., DNA methylation age of human tissues and cell types. Genome Biology, 2013. 14(10): p. R115.

37. Zhang, Z., et al., HiTIMED: hierarchical tumor immune microenvironment epigenetic deconvolution for accurate cell type resolution in the tumor microenvironment using tumor-type-specific DNA methylation data. Journal of Translational Medicine, 2022. 20(1).

38. Maechler, M., et al., cluster: Cluster Analysis Basics and Extensions. R package, 2022. version 2.1.4.

39. Dobin, A., et al., STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 2013. 29(1): p. 15-21.

40. Langfelder, P. and S. Horvath, WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 2008. 9(1): p. 559.

41. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 2014. 15(12): p. 550.

42. Zhang, B. and S. Horvath, A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol, 2005. 4: p. Article17.

43. Durinck, S., et al., BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics, 2005. 21(16): p. 3439-40.

44. Durinck, S., et al., Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc, 2009. 4(8): p. 1184-91.

45. Carlson, M., et al., org.Hs.eg.db: Genome wide annotation for Human. R package, 2019. version 3.8.2.

46. Wu, T., et al., clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation, 2021. 2(3): p. 100141.

47. Gu, Z., Complex heatmap visualization. iMeta, 2022. 1(3).

48. Liu, X., et al., Multi-omics analysis of intra-tumoural and inter-tumoural heterogeneity in pancreatic ductal adenocarcinoma. Clinical and Translational Medicine, 2022. 12(1): p. e670.

49. Ma, Y. and X. Zhou, Spatially informed cell-type deconvolution for spatial transcriptomics. Nat Biotechnol, 2022. 40(9): p. 1349-1359.

50. Hwang, W.L., et al., Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. Nat Genet, 2022. 54(8): p. 1178-1191.

51. Grünwald, B.T., et al., Spatially confined sub-tumor microenvironments in pancreatic cancer. Cell, 2021. 184(22): p. 5577-5592.e18.

52. Cui Zhou, D., et al., Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. Nat Genet, 2022. 54(9): p. 1390-1405.

53. Wang, J., et al., Multiomic analysis for optimization of combined focal and immunotherapy protocols in murine pancreatic cancer. Theranostics, 2022. 12(18): p. 7884-7902.

54. Sadozai, H., et al., Distinct Stromal and Immune Features Collectively Contribute to Long-Term Survival in Pancreatic Cancer. Front Immunol, 2021. 12: p. 643529.

55. Tew, B.Y., et al., Genome-wide DNA methylation analysis of KRAS mutant cell lines. Sci Rep, 2020. 10(1): p. 10149.

56. Sherman, M.H. and G.L. Beatty, Tumor Microenvironment in Pancreatic Cancer Pathogenesis and Therapeutic Resistance. Annu Rev Pathol, 2023. 18: p. 123-148.

57. Yuan, Z., et al., Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. Mol Cancer, 2023. 22(1): p. 48.

58. El Baba, N., et al., The Role of Rho GTPases in VEGF Signaling in Cancer Cells. Anal Cell Pathol (Amst), 2020. 2020: p. 2097214.

59. Sierra, J.R., et al., Tumor angiogenesis and progression are enhanced by Sema4D produced by tumor-associated macrophages. J Exp Med, 2008. 205(7): p. 1673-85.

60. Spiering, D. and L. Hodgson, Dynamics of the Rho-family small GTPases in actin regulation and motility. Cell Adh Migr, 2011. 5(2): p. 170-80.

61. Ridley, A.J., Rho GTPase signalling in cell migration. Curr Opin Cell Biol, 2015. 36: p. 103-12.

62. Brummelman, J., K. Pilipow, and E. Lugli, The Single-Cell Phenotypic Identity of Human CD8(+) and CD4(+) T Cells. Int Rev Cell Mol Biol, 2018. 341: p. 63-124.

63. Rossy, J., et al., The integration of signaling and the spatial organization of the T cell synapse. Front Immunol, 2012. 3: p. 352.

64. Sharpe, A.H. and K.E. Pauken, The diverse functions of the PD1 inhibitory pathway. Nature Reviews Immunology, 2018. 18(3): p. 153-167.

65. Riley, J.L., PD-1 signaling in primary T cells. Immunol Rev, 2009. 229(1): p. 114-25.

66. Clark, E.A. and N.V. Giltiay, CD22: A Regulator of Innate and Adaptive B Cell Responses and Autoimmunity. Front Immunol, 2018. 9: p. 2235.

67. Pishesha, N., T.J. Harmand, and H.L. Ploegh, A guide to antigen processing and presentation. Nat Rev Immunol, 2022. 22(12): p. 751-764.

68. Sarma, J.V. and P.A. Ward, The complement system. Cell Tissue Res, 2011. 343(1): p. 227-35.

CHAPTER 2

Investigation of KRASG12C as a diagnostic biomarker and therapeutic target in pancreatic cancer

Introduction

RAS family

The *RAS* (derived from 'RAt Sarcoma' virus) family is a subset of the *RAS* small GTPase superfamily and encodes well-known proto-oncogenes to date. The genes *KRAS* (*Kirsten rat homologous viral proto-oncogene sarcoma*), *HRAS* (*Harvey rat viral oncogene homolog sarcoma*) and *NRAS* (*neuroblastoma rat viral oncogene homolog sarcoma*) are the *RAS* family members in humans [1]. According to the Catalogue of Somatic Mutations in Cancer (COSMIC), activating mutations of *RAS* genes occurs in a significant number of all human cancers [2]: KRAS (49,476 samples with mutation of 276,752 unique samples), HRAS (2,498 samples with mutation of 109,830 unique samples) and NRAS (8,171 samples with mutation of 158,534 unique samples) (COSMIC v96, released 31-MAY-22; available at https://cancer.sanger.ac.uk/cosmic - last access in 31-JUL-22).

RAS genes are ubiquitously expressed, and they encode GTPases proteins that control pathways responsible for cell proliferation and survival. The abnormal activity of RAS proteins in cancer cells is linked to gain-of-function mutations in codons 12, 13 or 61 [1,3-5].

The RAS protein (21 kDa) is regulated by guanidine nucleotide exchange factors (GEF), which activate the protein by exchanging guanosine diphosphate (GDP) for guanosine triphosphate (GTP), and by GTPase activating proteins (GAP), which deactivate the protein through GTP hydrolysis [1,3]. The activated RAS protein (RAS-GTP) recruits rapidly accelerated fibrosarcoma (RAF) kinases to the cell membrane leading to phosphorylation and activation of mitogen-activated protein kinase (MEK) and extracellular-signal-regulated kinase (ERK)

(BRAF/MAP-K pathway). RAS-GTP is also a mediator of the p110 catalytic subunit activation of phosphatidylinositol 3-kinase (PI3K), stimulating the protein kinase B (PKB) kinase activity via PIP3 (PI3K/mTOR pathway; mTOR: mammalian target of rapamycin). The BRAF/MAP-K and PI3K/mTOR signaling pathways regulate several processes associated with cancer, including cell proliferation, growth, and survival [6-8].

The mutation makes RAS insensitive to GTP hydrolysis, resulting in constitutively activated proteins, which stimulate the disordered proliferation of cancer cells even in the absence of growth factors [1,3].

KRAS

The *RAS* genes are amongst the most studied and well-characterized cancer-related genes due to numerous studies being conducted since their discovery in the 1960s. The most common mutated RAS member in human cancers is *KRAS* (*KRAS*: 20%; *NRAS*: 8%; *HRAS*: 3.3%). This gene is located in 12p12.1 and has 4 alternative splicing transcripts that encodes two isoforms (Table 1) [3,9].

Table 1. List of *KRAS* alternative splicing variants and its encoded protein.

TRANSCRIPT			PROTEIN		
Name	**Transcript ID**	**Length**	**Name**	**UniProt ID**	**Length**
KRAS-201	ENST00000256078.10	5,430 bp	KRAS4A	P01116-1	189 aa
KRAS-202	ENST00000311936.8	5,306 bp	KRAS4B	P01116-2	188 aa
KRAS-205	ENST00000685328.1	5,287 bp	KRAS4B	P01116-2	188 aa
KRAS-210	ENST00000688940.1	3,630 bp	KRAS4B	P01116-2	188 aa

Legend: bp=base pairs; and aa=amino acids.

Timeline: highlight in *KRAS* discovery and research

- 1964: Harvey discovered that preparation of murine leukemia virus obtained from leukemic rats was able to induce sarcomas in newborn rodents [10];

- 1967: Ki-MSV (Kirsten Murine Sarcoma Virus) was obtained by serially passaging murine leukemia viruses through rats [11];
- 1973: Ki-MSV was identified as a recombinant virus that contains rat DNA sequences [12];
- 1977: Ki-MSV was characterized by molecular cloning in bacteria [13];
- 1982a: *KRAS* was discovered to be an oncogene in human cancer due to a homologous gene to Ki-MSV being identified in mouse 3T3 fibroblast cells transformed by DNA isolated from lung carcinoma cells [14];
- 1982b: clones corresponding to the human *KRAS* gene were isolated from human placental and embryonic cDNA libraries. Two *KRAS* isoforms were identified: *KRAS1*, 0.9kb homologous to viral *KRAS*, and *KRAS2*, 0.3kb homologous to viral *KRAS* [15];
- 1983a: KRAS mutants p.G12C and p.G12V were cloned from human lung and colon carcinomas [16];
- 1983b: *KRAS* was isolated from SW840 (human colon adenocarcinoma cell line) and characterized as *KRAS2* [17], which was identified in 1982 [15]. *KRAS2* was amplified in several tumor cell lines [17];
- 1983c: *KRAS1* and *KRAS2* were cloned and identified respectively as pseudogene and functional gene. It was found that *KRAS2* encodes a protein with 188 residues (21.66kD), which differs from the viral gene by 6 amino acids. A comparison between both isoforms revealed that *KRAS1* lacks several introns, suggesting it is a pseudogene derived from processed KRAS2 mRNA. Alternative splicing produced the two variants (isoform A and B), which differ in the C-terminal region [18];

- 1985: *KRAS2* was mapped to chromosome 12p12.1-p11.1 by *in situ* hybridization [19];
- 1987: high incidence of *KRAS* mutation at codon 12 was detected in primary colon cancer cells [20];
- 1988: high frequency of *KRAS* mutation at codon 12 was described in human carcinomas of the exocrine pancreas [21] and pancreatic adenocarcinomas [22];
- 1993: mutated *KRAS* was found to be important in colorectal tumorigenesis through altered cell differentiation and cell growth [23];
- 1997: *KRAS* was found to be essential for normal embryonic development of mice [24];
- 2001: the first Genetically Engineered Models of mice with KRAS mutant were developed for KRAS p.G12D in lung cancer [25];
- 2006: *KRAS* isoforms were described as result of alternative splicing of exon 5, whereas exon 6 encodes the 3′-UTR (*KRASA*) and the C-terminal region (*KRASB*). The C-terminal regions of both isoforms are subjected to posttranslational modifications (PTM), these modifications can have functional effect resulting in alternative trafficking pathways and protein localization [26];
- 2013: the first inhibitor of a KRAS mutation (p.G12C) was discovered [27];
- 2019: AMG 510 (Sotorasib) became the first drug to successfully target a *KRAS* mutation (p.G12C) in patients with the chemical compound showing antitumor effect in clinical trials [28];
- 2021: The U.S. Food and Drug Administration (FDA) approved Sotorasib for the treatment of KRAS p.G12C lung cancer under the trade name Lumakras® [29].

- 2022: The FDA approved Adagrasib for the treatment of KRAS p.G12C lung cancer under the trade name Krazati® [30].

KRAS driver mutations as therapeutic target

The development of effective RAS inhibitors depends on preventing the interaction between RAS and its downstream effectors. This can be achieved through different mechanisms, such as reducing the proportion of RAS in its GTP-bound state, disrupting RAS-GTP-effector interactions, stabilizing non-productive protein complexes, or reducing the population of RAS at the membrane [31].

The transversion g.34G>T in *KRAS* (rs121913530) results in the substitution of glycine for cysteine at position 12 of the KRAS protein, the *KRAS* p.G12C is associated with poorer prognosis and therapy resistance in lung cancer [32,33]. Covalent inhibitors targeting this specific mutation have been developed as the nucleophilicity of the cysteine thiol group (-SH) in the mutated protein allows the incorporation of an electrophilic fraction to the molecule, a chemical characteristic that can be exploited to develop inhibitors targeting the cysteine [32]. Sotorasib is an FDA-approved covalent inhibitor of KRASG12C. It binds to the KRAS protein in its GDP-bound conformation, thereby inhibiting mutant KRAS-driven signaling through the prevention of its activation [34]. Adagrasib, the second KRAS inhibitor approved by the FDA, also has the same mechanism of action as Sotorasib [30].

The substitution rate of transitions is higher than transversions in general. Nonsynonymous transversions are unlikely to conserve the amino acid's biochemical properties, thus it is believed that this selection disfavors them [35].

Near-infrared (NIR) imaging and fluorescent probe

The early detection of pancreatic cancer is still a challenge with the currently available technology. The NIR imaging process can bring some improvements such as high spatial resolution, real-time imaging with low acquisition time, highly sensitive and inexpensive process, which makes its use very promising for the diagnosis of pancreatic cancer [36].

The NIR imaging utilizes fluorescent probes, one example of these probes is the heptamethine cyanine. This near-infrared dye has both imaging and targeting properties: it has preferential accumulation and retention in different cancer cell types with no accumulation in non-tumor cells (both *in vitro* and *in vivo*) [37,38]; it does not cause cytotoxicity or systemic toxicity in mice when administered in a dose range appropriate for imaging; it is low auto-fluorescent, but when internalized by a tumor cell (lower pH than non-cancerous cells) it induces a strong fluorescent signal after being chemically modified to improve the stability of the molecule. The compound also has an IR absorption ranging between 700-800 nm, which prevents water absorbance interference, resulting in an improvement of the imaging exam precision [37,38].

Turn-on probes have been conjugated with anticancer drugs to improve tumor detection and deliver the drug to the target tissues [39]. The NIR probe and the candidate drug can be linked with pH-sensitive hydrazine. This substance remains intact during systemic circulation (pH 7.3-7.5), but undergoes hydrolysis when it reaches a slightly acidic pH, therefore when the conjugate is internalized by the tumor cell, the drug is released from the conjugate [40]. This approach improves tumor detection and drug delivery, two of the biggest challenges in pancreatic cancer treatment [41].

Here, we introduce a novel series of pH-sensitive near-infrared (NIR) compounds designed for early detection of pancreatic ductal adenocarcinoma (PDAC), targeting KRAS p.G12C. We examined the effects of mutations on the protein surface and employed molecular docking techniques to evaluate the potential binding affinity of the new compounds to the KRAS p.G12C.

Materials & Methods

Frequency distribution of *KRAS* in human cancer

The frequency distribution of *KRAS* in human cancer was based upon data generated by The Cancer Genome Atlas (TCGA) Research Network (https://www.cancer.gov/tcga) from human samples obtained in a clinical setting. The web server UCSC Xena (https://xenabrowser.net/; last access on 23-MAY-22) was used to obtain the dataset TCGA Pan-cancer (PANCAN) [42], which comprises different tabular sheets (.tsv) with diverse clinical, genomic, and phenotypic information (Supplementary material S1-S3).

For this analysis, we selected relevant genomic (somatic mutation in *KRAS*; Supplementary material S2) and phenotypic (type of tumor; Supplementary material S1) information and merged them based on a unique sample identifier (TCGA ID) (Supplementary material S4). This data was categorized using two criteria for PANCAN analysis (Supplementary materials S5-S6): 1. Mutated codon and amino acid change; 2. Tumor type. For the Pancreatic Cancer frequency distribution (Supplementary materials S7-S8), The Pancreatic Ductal Adenocarcinoma data were extracted from the PANCAN tumor type dataset and further categorized by KRAS mutation.

KRAS protein crystallized structure comparison

Computational details

The structure comparison was performed on the 2-in-1 Laptop SAMSUNG Galaxy Book Pro, with Intel® Core™ i7-1165G7 processor, 4.7 GHz, Intel® Iris® Xe Graphics Card, RAM 16 GB, under the Windows® 11 operating system.

Comparison of the structural pocket in KRAS variants

Not all KRAS variants have crystallized structures available. All the protein structures of the available variants were obtained from RSCB PDB (https://www.rcsb.org/) in PDB format (Table 2). When more than one structure was available for that specific variant, the structure in complex with GDP was chosen. If there was no structure in complex with GDP, the protein in complex with GTP or an analog was chosen. Although there is a crystal available for KRAS p.Q61L on RSCB PDB, the only available PDB (PBD ID: 4WA7) does not have the residue 61, thus the pocket on this mutation was not analyzed as the mutated residue is important for the structure of the protein.

The location where a known KRAS p.G12C inhibitor (AMG 510; Sotorasib) is covalently bound to in the KRAS p.G12C was used as a reference for the comparison of the active site pocket. Protein surfaces were highlighted according to the mutated residue, and ligands are shown in stick format. The structures were compared in BIOVIA Discovery Studio (https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/).

Table 2. List of PDB IDs used for pocket comparison of available KRAS protein variants.

KRAS MUTATION	PDB ID
Wild type	5W22
G12C	6OIM*
G12D	5US4
G12R	4QL3
G12V	7C40
G12A	5VP7
G13D	4TQA
Q61H	3GFT#
Q61A	5VQ1

Unless stated otherwise, all proteins are complexed with GDP. *Legend:* ***=protein in complex with GDP and AMG 510;* #*=protein in complex with GNP (GTP analog).*

Docking molecular

Computational details

The virtual protocol was performed on the DELL® Workstation computer, with Intel® Xeon E5-1660 processor, 3.3 GHz, 4 CPUs, NVIDIA® GeForce RTX 2060 graphics card, RAM 8 GB, under the Windows® 10 operating system.

Ligand preparation

The 3 compounds were built in ChemDraw suite (https://www.perkinelmer.com/product/chemdraw-professional-chemdrawpro) from their 2D structures. All geometries were virtually constructed and optimized in molecular mechanics (MM+) and semiempirical *Austin Model 1* (AM1) methods using HyperChem 7 (http://www.hyper.com).

The output files from HyperChem 7 were converted using Open Babel (http://openbabel.org/wiki/Main_Page) to .mol2 format, which is the required input file format for the PyRx software (https://pyrx.sourceforge.io). The PyRx software was utilized to run the docking experiment using AutoDock Vina to predict *in silico* the best orientation and conformation of the compounds to the KRAS proteins.

AutoDock Vina utilizes the stochastic (random) algorithm to optimize random conformations with the lowest energy using a rigid receptor and flexible ligand approach, i.e., the program tries to bind the ligand to the receptor randomly altering its position and orientation considering the torsions of the molecule. The stochastic method used by AutoDock Vina for local optimization is the Broyden-Fletcher-Goldfarb-Shanno (BFGS), which is an efficient quasi-Newton method [43]. The predicted structure of ligands is evaluated by the Metropolis acceptance criteria, which is based on the difference between the docking score of the initial and optimized conformations to determine if the new conformation is acceptable. Then, the accepted conformation will be considered as the initial conformation for the next optimization in a serial manner [43,44].

Target selection and molecular docking simulations

The proteins utilized for the docking simulations were sourced from the KRAS crystallized structures listed in Table 2, acquired via the RCSB PDB database. Each protein PDB file underwent preparation in BIOVIA Discovery Studio, wherein co-crystallized water molecules and ligands were removed. Subsequently, the ligand AMG510 bound to KRASG12C (PDB ID: 6OIM) was extracted and saved in a separate PDB file, designated for use as input file in PyRx.

The design and validation of the molecular docking methodology were performed using the redocking of AMG 510 and GDP into the KRAS p.G12C structure. This approach was chosen to assess the capability of reproducing the observed binding modes present in complexes with a known crystal structure. The successful redocking of AMG 510 and GDP, demonstrating similar

interactions to those observed in the experimental complex, served as validation for our methodology.

The docking simulations were made using the following parameters: AMG 510 (Sotorasib) (Figure 1A), GDP (Figure 1B), and each one of the 3 compounds (described in detail in the *Drug design* section; Figures 1C-E) as a binding energy reference; a grid box with dimensions of 25 Angstroms (Å) and several repetitions by exhaustion equal to 40, which was centralized based on the reference's coordinates. The positioning and sizing of the grid box were selected to cover the binding site of interest while covering the important amino acids for the pocket interaction.

Based on these parameters, AutoDock Vina infers the possible pharmacological interactions and allows the most promising compounds to be classified according to the studied biological activity, using the empirical score function for this classification [43]. The docking simulations were performed for each inhibitor into the corresponding target using the structures of the protein site (.pdb) and the ligands (.mol2) in PyRx. Then, the binding affinity energy was assessed, and a detailed analysis of the participating amino acids was conducted using BIOVIA Discovery Studio.

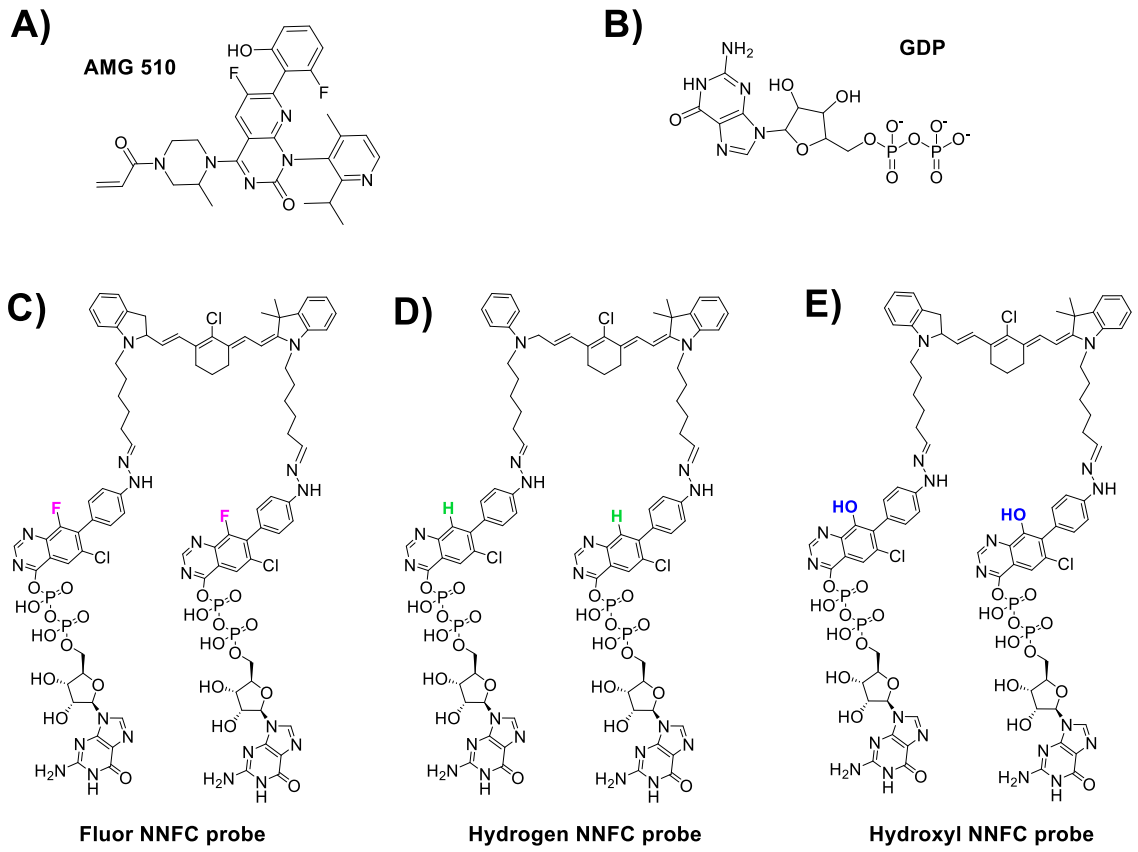


Figure 1. Chemical structure of molecules used in the docking simulations. **A.** AMG 510 (Sotorasib); **B.** GDP (guanosine diphosphate); **C.** Fluor NNFC probe; **D.** Hydrogen NNFC probe; **E.** Hydroxyl NNFC probe**.** *Legend: GDP=guanosine diphosphate; NNFC=nucleotide-near-infrared fluorescent conjugate.*

Drug design

Probe-drug conjugate theoretical development

The fluorescent NIR probe was based on the commercially available heptamethine cyanine MHI-148, which has an IR absorption ranging between 700-800 nm that prevents water absorbance interference in the imaging exam results [37,38]. The NIR probe and the candidate drug were linked with pH-sensitive hydrazine, which undergoes hydrolysis when internalized by the tumor cell, resulting in the drug release from the conjugate [40]. We propose an approach to develop theoretical frameworks for the creation of three distinct nucleotide-NIR fluorescent conjugate (NNFC) probes.

This approach uses a modified GDP as a candidate drug. This drug candidate can bind with the mutated protein KRAS with a similar affinity of GDP. The three drug candidates differ based on 3 different radical substitutions (H, F, or OH) (Figure 1C-D).

The commercially available multi-functionalized benzoic acid is converted to the mono-phosphate linker through cyclization with hydrazine, and phosphorylation followed by oxidation with MCPBA and de-protection (Figure 2A). The product from A reacts with guanosine monophosphate to produce the functionalized guanosine di-phosphate (Figure 2B). After a metal-free visible-light-induced photocatalytic borylation of C–O bonds catalyzed by phenothiazine, the synthesized boronic acids are used to finally provide the multi-functionalized GDP by a monophasic transesterification [45,46] (Figures 2C and D).

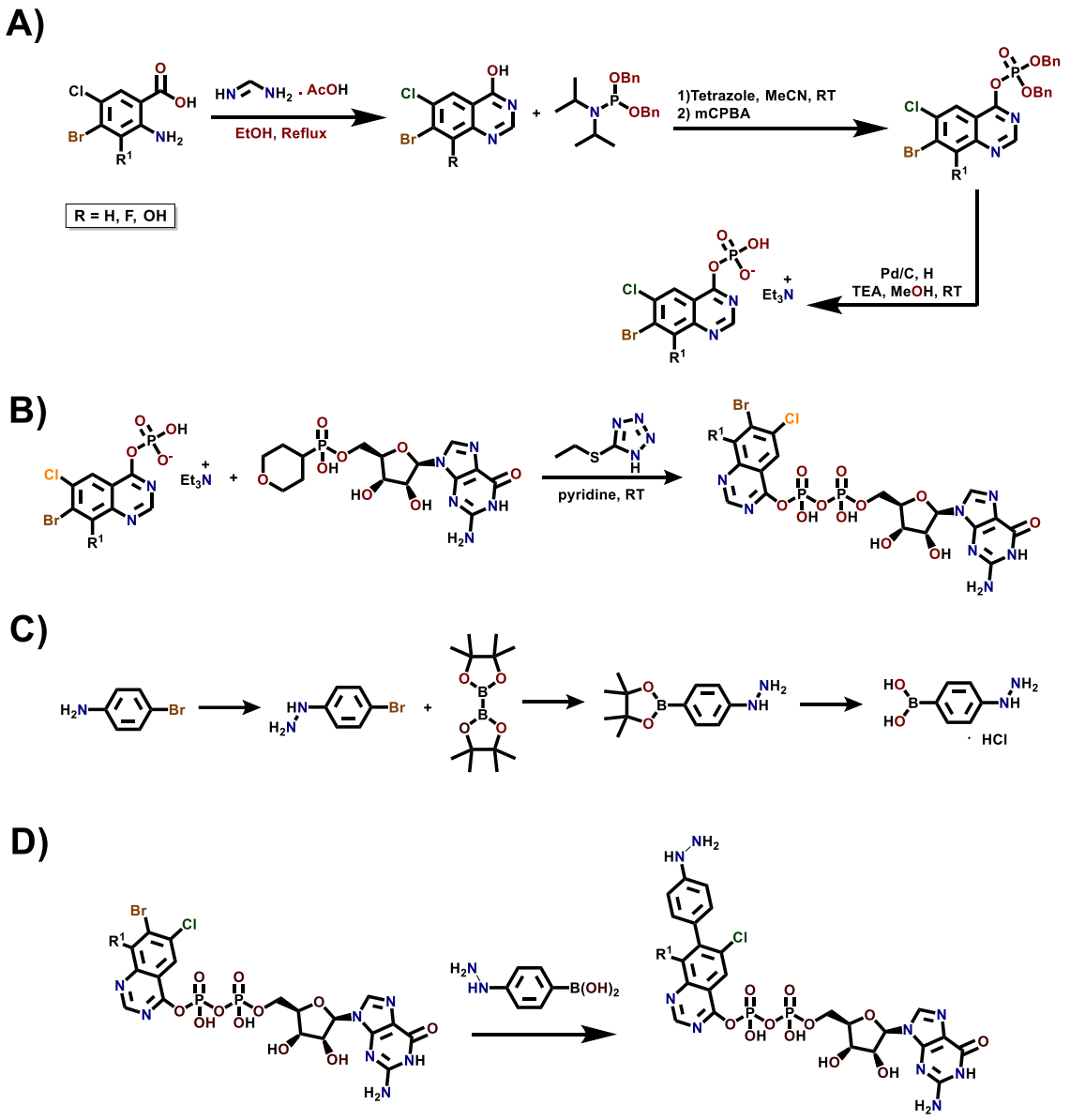


Figure 2. Chemical reactions of NNFC probe development: synthesis of GMP nucleotide.

The NIR fluorescent agent was synthesized using the heptamethine cyanine MHI-148 (Figure 3A) which would be conjugated with the hydrazine linker by a nucleophilic substitution reaction (Figure 3B) [40].

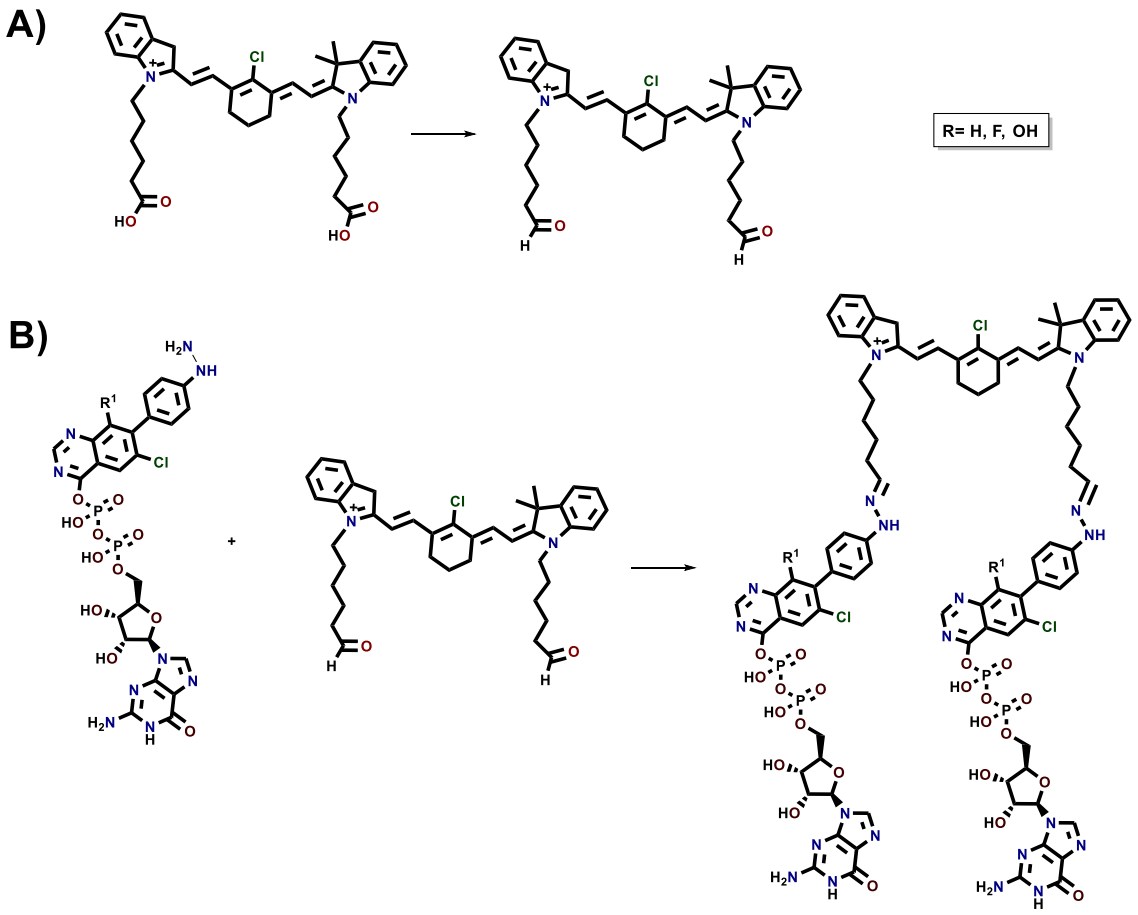


Figure 3. Chemical reactions of NNFC probe development: development of the nucleotide-NIR fluorescent conjugate probes. **A.** Reduction of heptamethine cyanine; **B.** conjugation of the NIR fluorescent agent with GMP nucleotide by a nucleophilic substitution reaction.

Results

Frequency distribution of *KRAS* in human cancer

We selected the TCGA PANCAN dataset to analyze the frequency distribution of *KRAS* in human cancer at a patient level [42]. The TCGA PANCAN study encompasses 12,841 cancer samples obtained in clinical settings, of which 12,833 provide information regarding the *KRAS* gene (Supplementary Material S5).

As expected, the PANCAN analysis revealed that *KRAS* mutation hotspots are predominantly located at codons 12, 13, and 61, with frequencies of 71.25%, 11.09%, and 4.89%, respectively (Figure 4A). Notably, 12.28% of the mutated samples harbor a p.G12C mutation (n=83; Figure 4A). As depicted in Figure 4B, the most prevalent cancer type with KRAS p.G12C mutations in the PANCAN dataset is Lung Adenocarcinoma (LUAD), accounting for 69.88% of samples, followed by Colon Adenocarcinoma (COAD) at 8.43%, and Uterine Corpus Endometrial Carcinoma (UCEC) at 4.82%. Pancreatic Adenocarcinoma (PAAD) represents only 1.20% of the KRAS p.G12C total samples.

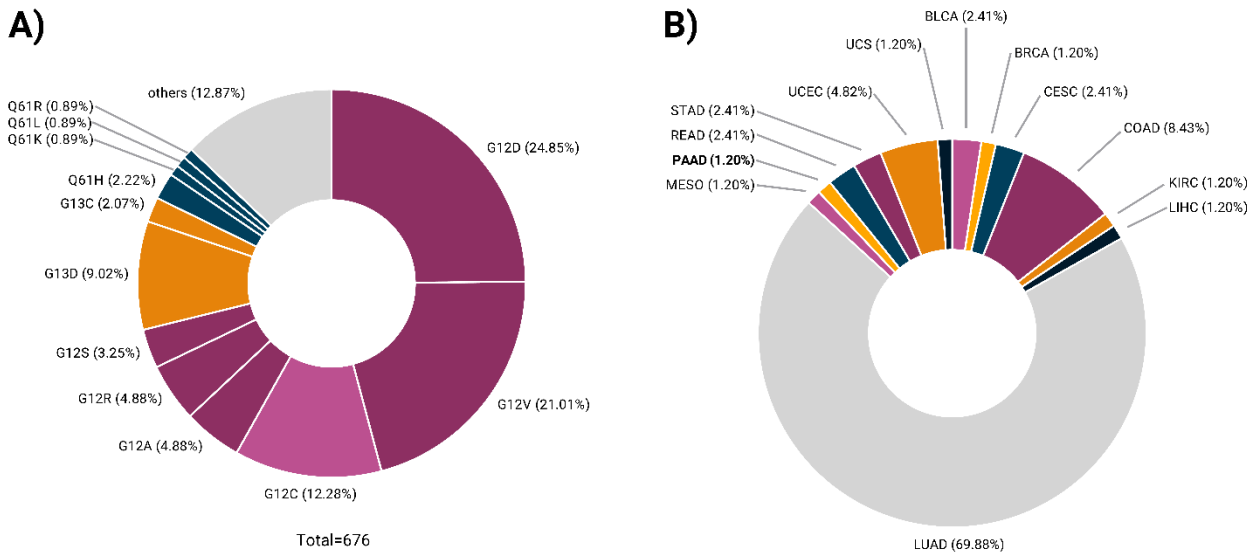


Figure 4. Frequency distribution of **A** mutation hotspots in *KRAS*; and **B.** KRAS p.G12C in The Cancer Genome Atlas (TCGA): Pan-Cancer (PANCAN) dataset. *Legend: BLCA=bladder urothelial carcinoma; BRCA=breast invasive carcinoma; CESC=cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD=colon adenocarcinoma; KIRC=kidney renal clear cell carcinoma; LIHC=liver hepatocellular carcinoma; LUAD=lung adenocarcinoma; MESO=mesothelioma; PAAD=pancreatic adenocarcinoma; READ=rectum adenocarcinoma; STAD=stomach adenocarcinoma; UCS=uterine carcinosarcoma; UCEC=uterine corpus endometrial carcinoma.*

The dataset of Pancreatic Cancer with *KRAS* information comprises only 196 samples (Supplementary Material S7), among which 59% exhibit mutations, contrasting with 31% classified as wild-type samples and 10% unknown (missing information about *KRAS* mutation status). Notably, more than half (55%) of this dataset presents mutations in codon 12 (Figure 5).

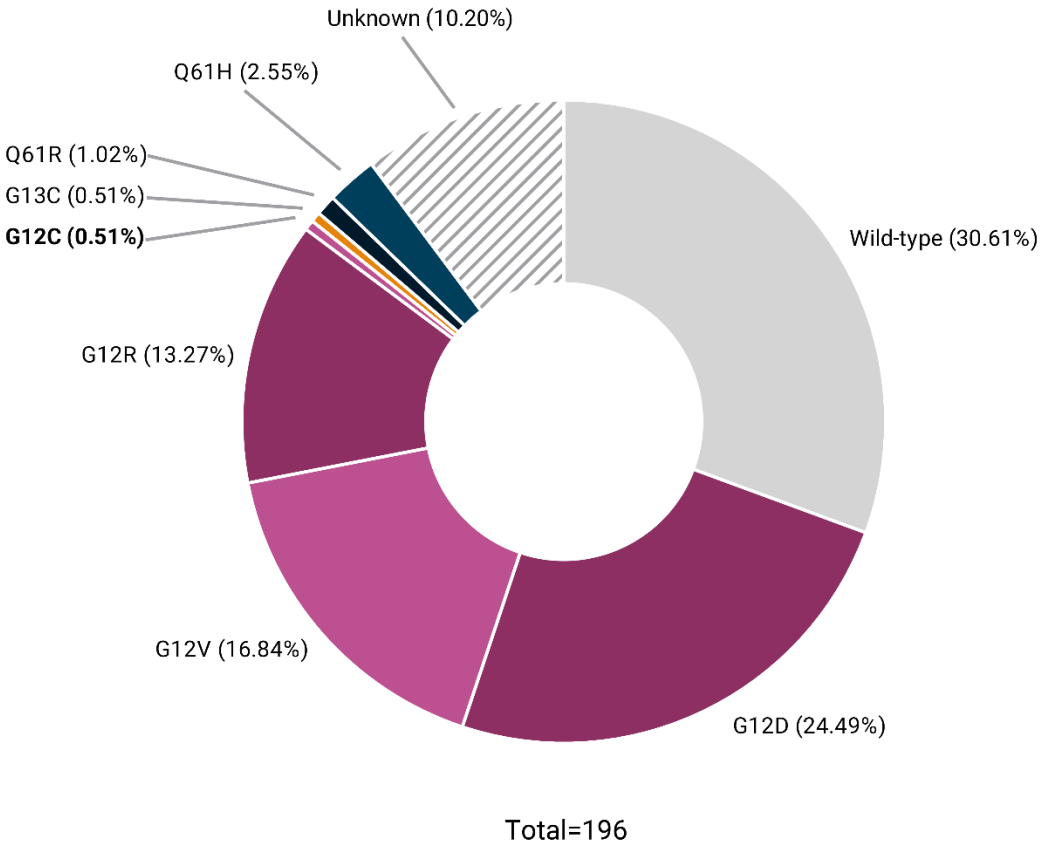


Figure 5. The Cancer Genome Atlas (TCGA): distribution of KRAS mutations in the Pancreatic Ductal Adenocarcinoma (PAAD) dataset.

KRAS protein structure comparison of pockets on mutants

For this comparison, two binding pockets on the KRAS protein are of significance: firstly, the binding pocket accommodating GDP/GTP, referred to herein as pocket 1; and secondly, the site where sotorasib, a known inhibitor, binds to KRAS p.G12C, herein designated as pocket 2.

In the KRAS-WT protein, pocket 1 exhibits an open configuration at the end where the phosphates of the ligands are situated (Figure 6A). Mutations at codon 12 induce alterations on the surface of this specific end of pocket 1, particularly where it interfaces with pocket 2. In mutated proteins, the presence of a cysteine (Cys) residue at codon 12 generates a substantial tunnel connecting both pockets (G12C; Figure 6B), while an aspartic acid (Asp) residue at the same

position entirely obstructs the communication between pockets (G12D; Figure 6C). Similarly, a tunnel linking both pockets emerges when glycine (Gly) at codon 12 is substituted by arginine (Arg) (G12R; Figure 6D) or valine (Val) (G12V; Figure 6E). The tunnel formed by the G12R mutation is narrow (G12R; Figure 6D), whereas G12V exhibits a narrow tunnel with a bifurcated opening on the pocket 2 side, significantly impeding communication between pockets (G12V; Figure 6E). Although a mutated protein with alanine (Ala) at codon 12 of KRAS does not obstruct communication between pockets, pocket 2 is nearly completely occluded, thereby impeding the binding of most ligands at this site (G12A; Figure 6F).

Mutations at codon 13 induce modifications to the surface of pocket 1 close to its interface with pocket 2. In KRAS-WT, the presence of a glycine (Gly) residue at this position maintains an open configuration at the end of pocket 1 (Figure 6A). However, substitution of glycine with aspartic acid (Asp) partially obstructs the bottom of pocket 1 where it interfaces with pocket 2 (G13D; Figure 6G).

Mutations in KRAS at amino acid position 61 modify the surface of pocket 2 near its interaction site with pocket 1. When the amino acid glutamine (Gln) is substituted with histidine (His), it partially obstructs both pockets at the interface between them (Q61H; Figure 6H). Conversely, an alanine (Ala) at codon 61 partially obstructs the opposite end of pocket 2 (Q61A; Figure 6I).

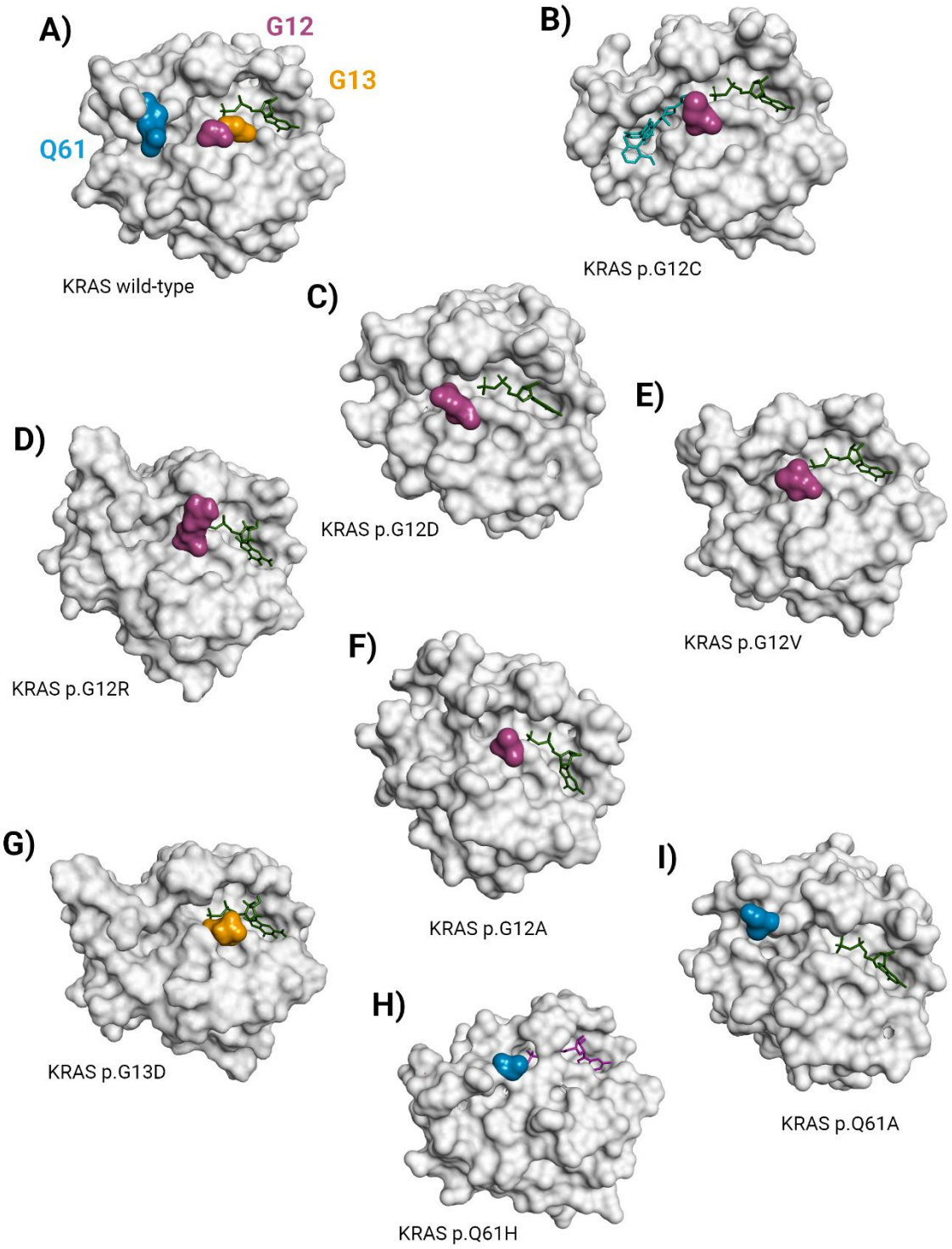


Figure 6. Pocket comparison of the active site in crystallized structure of KRAS variants. **A.** KRAS Wild-Type (WT) – PDB ID: 5W22; **B.** KRAS p.G12C – PDB ID: 6OIM; **C.** KRAS p.G12D – PDB ID: 5US4; **D.** KRAS p.G12R – PDB ID: 4QL3; **E.** KRAS p.G12V – PDB ID: 7C40; **F.** KRAS p.G12A – PDB ID: 5VP7; **G.** KRAS p.G13D – PDB ID: 4TQA; **H.** KRAS p.Q61H – PDB ID: 3GFT; **I.** KRAS p.Q61A – PDB ID: 5VQ1. *Legend: purple surface=residue 12; orange surface=residue 13; blue surface=residue 61; green ligand=GDP (guanine diphosphate); aqua ligand=AMG 510 (sotorasib); and pink ligand=GNP (nucleoside diphosphate; guanine triphosphate analogue).*

Molecular docking

AMG 510 (established inhibitor; Figure 1A) and GDP (endogenous activator; Figure 1B) were used as references to establish the docking box coordinates in KRAS p.G12C (PDB ID: 6OIM). The possible pharmacological interactions were inferred by PyRx using empirical information about the conformation of receptor-ligand complexes and the experimental affinity measurements following AutoDock Vina protocol [43]. Thus, the three compounds' binding energy predicted by docking were analyzed and classified according to their binding affinity; the interaction energy values were expressed in Kcal/mol as shown in Table 3.

Table 3. Interaction energies of ligand-receptor complexes with KRAS p.G12C (crystallized).

Compound	Binding affinity (Kcal/mol)
AMG 510	-7.2
GDP	-10.5
Fluor NNFC	-6.6
Hydrogen NNFC	-5.1
Hydroxyl NNFC	-7.0

Legend: AMG 510=Sotorasib; GDP=guanosine diphosphate; NNFC=nucleotide-near-infrared fluorescent conjugate.

Twenty-seven poses of the three NNFC probes (Figures 1C-E) were generated through docking. According to AutoDock Vina score function, the conformation with less energy involvement, indicating increased binding affinity, represent the most stable three-dimensional conformations. These conformations closely resemble the ligand-receptor complexes in their biologically active states [47]. The most stable conformations for each NNFC probe in KRAS p.G12C are shown in Figures 7B-D.

The analysis of Table 3 and Figure 8 highlight the significant relevance of amino acid interactions in the binding of NNFC probes to the active site of KRAS p.G12C. Common residues involved in interactions with pocket 1 across all NNFC probes and AMG 510 include Cys12, Gly13, Gly15, Lys16, Phe28, Tyr32, Pro34, Ala59, and Asn85. Additionally, hydrogen and hydroxyl interactions are observed with Ala18, Asp30, Leu120, and Lys147 residues, which are shared with AMG 510 (Figures 8A, 8C-E).

All NNFC compounds exhibit common interactions with GDP involving residues Cys12, Gly13, Gly15, Lys16, Phe28, Asp30, Tyr32, Pro34, and Lys117 (Figures 8B-E). Fluor NNFC interacts with residues Ala11 and Val14 (Figure 8C), Hydrogen NNFC interacts with residues Ser17, Ala18, and Leu120 (Figure 9D), while Hydroxyl NNFC interacts with Ala11, Val14, Ala18, Asn116, Asp119, Leu120, Ser145, Ala146, and Lys147 (Figure 8E) that are also observed in the KRAS-GDP complex (Figure 8B).

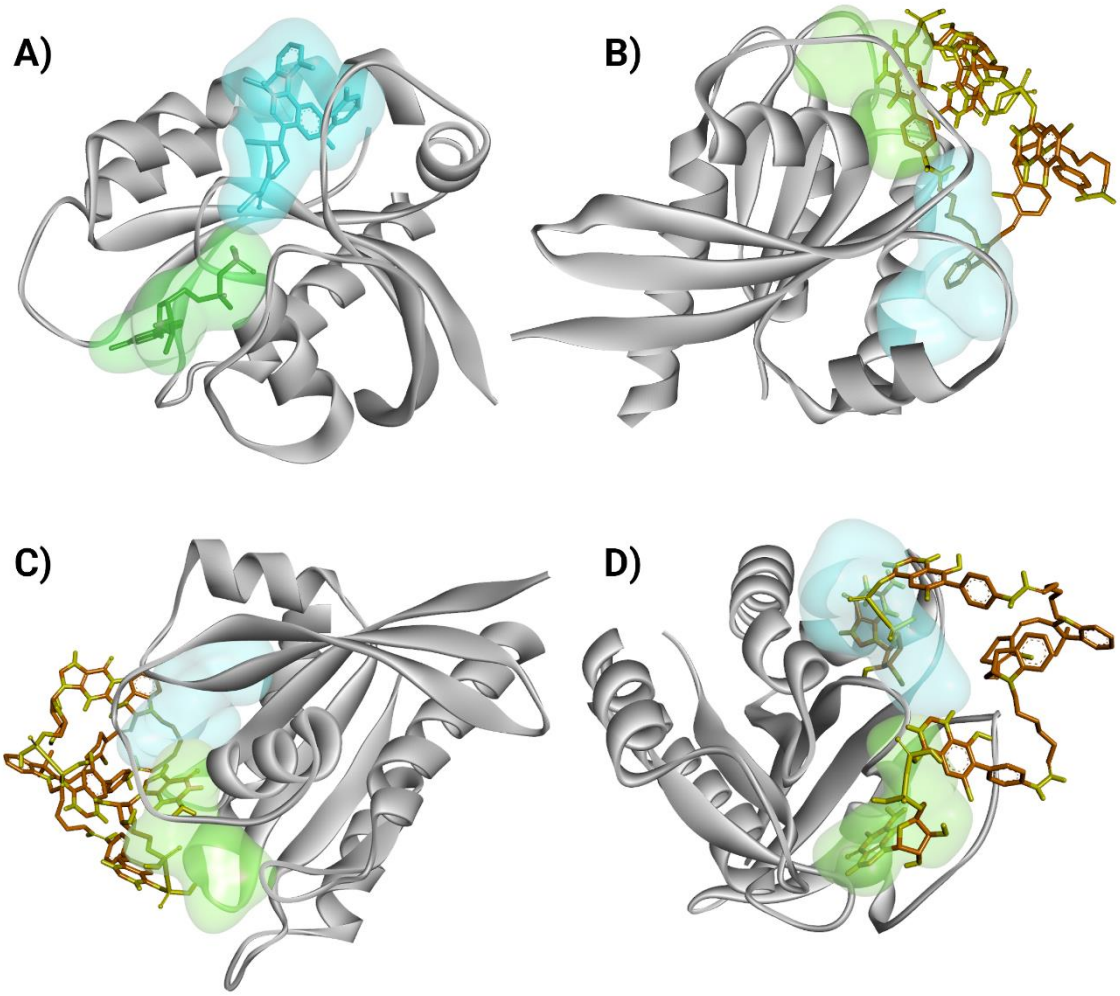


Figure 7. Best predict docking conformations for each tested probe. **A.** KRAS P.G12C with AMG 510 (sotorasib, known inhibitor) interaction pocket highlighted in blue, and GDP (guanosine diphosphate) pocket highlighted in green; **B.** Fluor NNFC; **C.** Hydrogen NNFC; **D.** Hydroxyl NNFC. *Legend: NNFC=nucleotide-near-infrared fluorescent conjugate; gray protein= KRAS P.G12C obtained from the PDB 6OIM; green stick format=GDP; blue stick format=sotorasib; yellow/orange stick format= tested compounds (carbon atoms in orange, non-carbon atoms in yellow).*

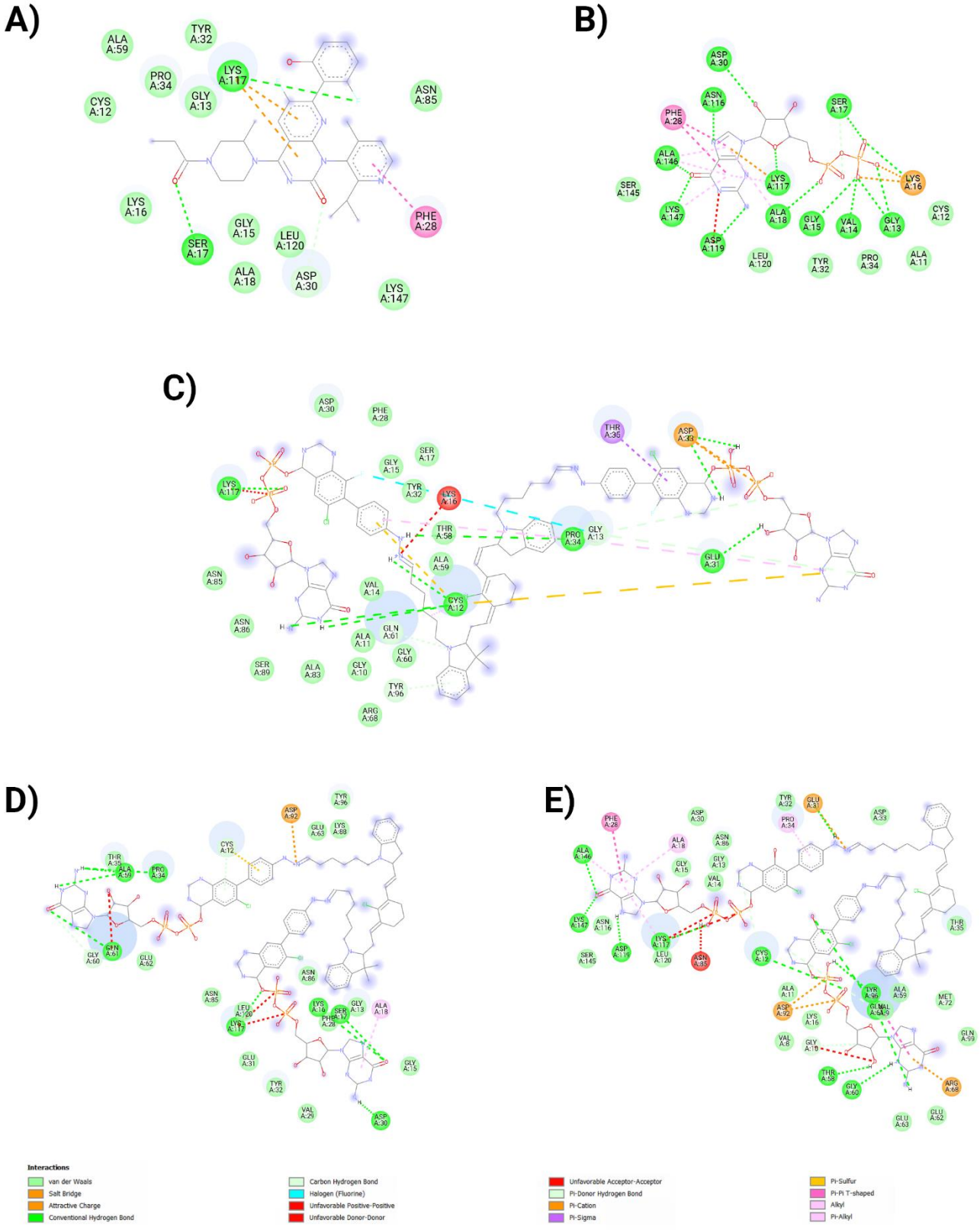


Figure 8. Amino acids interaction with **A.** AMG 510 (Sotorasib); **B.** GDP (guanosine diphosphate); **C.** Fluor NNFC**; D.** Hydrogen NNFC; and **E.** Hydroxyl NNFC. *Legend: NNFC=nucleotide-near-infrared fluorescent conjugate.*

Discussion and conclusion

There are a lot of available data about the *KRAS* gene and its homonymous protein as it is one of the most studied and well-characterized genes related to cancer [3,5]. However, this data is not well distributed amongst different mutations and databases (Table 4). The variants of KRAS protein with available crystals are all included in Table 4, while TCGA PANCAN has other codon mutations identified in its patients' samples.

Table 4. Cross-reference of available KRAS hotspots (codons 12, 13, and 61) mutations in RCSB PDB Database (protein) and TCGA (RNA-Seq-based transcriptome of patient samples).

KRAS MUTATION	RCSB PDB	TCGA PANCAN
Wild type	yes	yes
p.G12A	no	yes
p.G12C	yes	yes
p.G12D	yes	yes
p.G12R	yes	yes
p.G12S	no	yes
p.G12V	yes	yes
p.G12A	yes	yes
p.G13C	no	yes
p.G13D	yes	yes
p.G13G	no	yes
p.Q61A	yes	no
p.Q61H	yes	yes
p.Q61K	no	yes
p.Q61L	no	yes
p.Q61R	no	yes

This asymmetric distribution is expected as RCSB PDB is an archive of structural data of biological macromolecules. Complex experiments are needed to obtain the structural conformation of a macromolecule, so the biological importance of that specific protein is relevant for the protein to be selected, thus only the most common mutations have their structure available at RCSB PDB [48]. In contrast, TCGA PANCAN has data of tumor samples of patients available, thus all mutations detected in the patient cohort are included in the dataset. Even if a

mutation is not frequent it could be detected in the TCGA samples, albeit in less frequency than the most common mutations [42].

Although the TCGA dataset can offer insights into the distribution of cancer in the human population, it is important to note that the PANCAN dataset does not exhibit an even distribution of samples among different types of cancer. Cancers with higher mortality rates, which may not be among the most incident cancers worldwide, such as pancreatic cancer (ranked as the 6th most lethal cancer and 12th most incident cancer according to GLOBOCAN 2022), tend to have fewer samples deposited in the TCGA database [49]. Consequently, the limited number of samples available can pose challenges for performing genomic analysis, as we rely on the availability of data. In the case of *KRAS* mutations in pancreatic cancer, our analysis was restricted to frequency analysis due to the small number of samples for different mutations, which rendered them statistically insignificant.

Until recently, KRAS was considered a non-druggable target because the only known drug-binding pocket available was the GDP/GTP binding site, which has a high affinity for both GDP and GTP, therefore targeting only this binding site would be virtually impossible. KRAS as a druggable target became a possibility with the discovery of a dynamic pocket in the mutant KRAS p.G12C [50].

The analysis of the available structural data about the variants of KRAS corroborates the literature, as the new pocket in KRAS P.G12C can be seen in Figure 6B. In addition to our comparison of protein structures (Figure 6), it's evident that not only does KRAS p.G12C present a potential target, but hotspot mutations at codons 12, 13, and 61 in *KRAS* also alter the surrounding surface

of the GDP/GTP binding pocket. This alteration could potentially create new binding sites. However, drugs must be designed for each mutation, taking into consideration their specific surface characteristics.

The structural analysis also explains why a mutation in KRAS hotspots constitutively activated the protein. The intrinsic catalytic rate of KRAS is very low, thus the transformation of GTP (active state) in GDP (inactive state) requires the facilitation of a GAP protein [51]. The GTP-bound KRAS protein is the subtract required for GAP-assisted hydrolysis of GTP [52], thus the surface alterations in the proximity of the nucleotide-binding pocket interfere with this process, resulting in a constitutively activated protein. Mutations in codons 12 and 13 alter the surface of the end of the nucleotide pocket (Figures 6B-G), preventing the coupling of GAP to the GTP-KRAS protein, while mutations in residue 61 alter the surface in the proximity of the GDP-GTP pocket (Figures 6H-I), obstructing the coordination of a water molecule required for GTP hydrolysis [53].

Sotorasib (Lumakras®) is the first approved KRAS-target treatment by the FDA [29]. Although it was approved for the treatment of KRAS p.G12C lung cancer, there is evidence (ClinicalTrials Identifier: NCT00287391) that Sotorasib is also effective against other KRAS p.G12C solid tumors, including pancreatic cancer [29,54].

The NNFC probes we developed interact *in silico* with amino acids important for both AMG 510 (Sotorasib) and GDP interaction with the KRAS P.G12C protein, thus their binding pattern is similar to a known inhibitor and the endogenous activator. The tested compounds show an excellent affinity (Fluor NNFC: -6.6; Hydrogen NNFC: -5.1; and Hydroxyl NNFC: -7.0 Kcal/mol; Table 3)

for this protein based on the binding energy results obtained by docking using the crystallized KRAS p.G12C.

Targeting cysteine residues is a common strategy in the development of targeted covalent inhibitors as it could result in a strong interaction with the ligand [55], thus the mutant cysteine in G12C is the target of our new covalent inhibitors. Sotorasib has a weaker interaction with cys12 than the Fluor and Hydroxyl NNFCs, which have more than 1 interaction with the residue, including a conventional hydrogen bond (Figures 8B-C, 8E).

Although interacting with less affinity than GDP to the molecule, the increased binding affinity of Hydroxyl NNFC can be explained by the greater number of interactions, especially conventional hydrogen bonds (Figure 8E), which also explains the known high affinity of GDP and KRAS (Figure 8B). This suggests a possible competition through the GDP binding site, increasing the inhibition of this protein. Moreover, this compound shows a considerable number of interactions with the amino acids of the active site, which would stabilize the compound in an appropriate position to interact with the KRAS p.G12C active site.

In this context, our findings, along with the literature presented herein, support the potential of our NNFC probes as both diagnostic biomarkers and therapeutic targets in pancreatic cancer. Their selectivity renders them excellent candidates for further investigation in drug development studies.

Supplementary material

SUPPLEMENTARY MATERIAL S1. Raw phenotypic data of tumor samples. The Cancer Genome Atlas (TCGA) Pan-Cancer dataset obtained from the webserver UCSC Xena (https://xenabrowser.net/; last access in 23-MAY-22).

Available at: https://drive.google.com/file/d/1bd9c3ifsk-TRmdvlTjZqHvMDDm5t4Pzi/view?usp=sharing

SUPPLEMENTARY MATERIAL S2. Somatic mutation raw data of KRAS gene. The Cancer Genome Atlas (TCGA) Pan-Cancer dataset obtained from the webserver UCSC Xena (https://xenabrowser.net/; last access in 23-MAY-22).

Available at: https://drive.google.com/file/d/1WltQUwI3pQTnO-3iX_1aVufSVwmJxH_y/view?usp=drive_link

SUPPLEMENTARY MATERIAL S3. KRAS gene expression raw data. The Cancer Genome Atlas (TCGA) Pan-Cancer dataset obtained from the webserver UCSC Xena (https://xenabrowser.net/; last access in 23-MAY-22).

Available at: https://drive.google.com/file/d/1n9jc6vCk3ugBKLI3GiJ77cjwVOFOhKuq/view?usp=drive_link

SUPPLEMENTARY MATERIAL S4. Merged table of cancer type, KRAS gene expression, and somatic mutation information. The Cancer Genome Atlas (TCGA) Pan-Cancer dataset obtained from the webserver UCSC Xena (https://xenabrowser.net/; last access in 23-MAY-22).

Available at: https://drive.google.com/file/d/1VudOJ8PQmZHdr023mR5_KSAU4AJcvW9P/view?usp=drive_link

SUPPLEMENTARY MATERIAL S5. Supplementary material S5. Pan-cancer samples matched tumor type, mutated codon, and amino acid change (n=12,833).

Available at: https://docs.google.com/spreadsheets/d/1lmmCFfp1LurobIL5xbvEBzwEpt3AmYMD/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

SUPPLEMENTARY MATERIAL S6. Number of pan-cancer samples per mutated codon and amino acid change.

Available at: https://docs.google.com/spreadsheets/d/1lmmCFfp1LurobIL5xbvEBzwEpt3AmYMD/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

SUPPLEMENTARY MATERIAL S7. Pancreatic cancer samples matched tumor type, mutated codon, and amino acid change (n=196).

Available at: https://docs.google.com/spreadsheets/d/1lmmCFfp1LurobIL5xbvEBzwEpt3AmYMD/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

SUPPLEMENTARY MATERIAL S8. Number of pancreatic cancer samples per mutated codon and amino acid change.

Available at: https://docs.google.com/spreadsheets/d/1lmmCFfp1LurobIL5xbvEBzwEpt3AmYMD/edit?usp=drive_link&ouid=113652411378247866839&rtpof=true&sd=true

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.

6. Concluding remarks

The results presented in this study shed light on critical aspects of pancreatic cancer, particularly the association between *KRAS* mutations and immune profiles within the tumor microenvironment.

In Chapter 1, the identification of distinct methylation profiles and their correlation with *KRAS* mutations and patient survival outcomes underscores the intricate interplay between epigenetic alterations, genetic mutations, and disease prognosis. Furthermore, the discovery of three patterns of the tumor immune microenvironment emphasizes the heterogeneity of cell composition in the tumoral tissue and the complexity of interactions between tumor and non-tumor cells in pancreatic cancer. These findings underscore the importance of understanding these patterns for developing targeted therapeutic strategies.

Building upon these findings, Chapter 2 delves deeper into the structural implications of *KRAS* mutations and their potential therapeutic implications. The high prevalence of *KRAS* mutations, particularly in codon 12, highlights the significance of targeting specific mutation sites for effective treatment. The structural analysis underscores the need for allele-specific drug design to effectively target the unique surface characteristics created by these driver mutations.

Moreover, the promising binding affinity of the designed compounds to the KRAS p.G12C mutation presents exciting prospects for diagnostic and therapeutic advancements in pancreatic cancer treatment. These findings not only provide insights into the molecular mechanisms underlying pancreatic cancer but also offer potential avenues for the development of personalized and targeted therapeutic interventions.

In conclusion, the integration of molecular, genetic, and structural analyses presented in this study contributes to advancing the knowledge of pancreatic cancer biology and highlights potential avenues for improving patient outcomes through targeted therapeutic approaches. Further research in this direction holds great promise for advancing the field of pancreatic cancer treatment and ultimately improving patient management and survival.

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.

8. Appendix

In addition to the primary research presented in this thesis, I have also contributed to other bibliographical productions during my doctoral studies. This appendix section includes those works, which are as follows:

Papers published in academic journals (featured on the publication covers):

- Oliveira RJ, da Silveira IOMF, das Neves SC, Mitsuyasu B, Martins AC, Berno C, Mohammad J, Raj H, de Araujo FHS, Hortelan CR, Machado L, da Silva Júnior EN, Vilela MLB, Nascimento VA, Beatriz A, da Silva Gomes R. *ZIM, a Norbornene Derived from 4-Aminoantipyrine, Induces DNA Damage and Cell Death but in Association Reduces the Effect of Commercial Chemotherapeutics*. *Chem Res Toxicol*. 2023 Jan 16;36(1):66-82. DOI: 10.1021/acs.chemrestox.2c00275.

- Correa WA, das Neves SC, Oliveira RJ, Kassuya CA, Navarro SD, Faustino Martins AC, Saroja B, Mitsuyasu B, Ostaciana Maia Freitas da Silveira I, Vitor N, Coelho HRS, Vilela MLB, do Nascimento VA, de Lima DP, Beatriz A, da Silva Gomes R. *Chemotherapeutic Mechanism of Action of the Synthetic Resorcinolic Methyl 3,5-dimethoxy-2-octanoylbenzoate*. Chem Res Toxicol. 2024 Feb 19;37(2):259-273. DOI: 10.1021/acs.chemrestox.3c00269.

Abstracts published in conference proceedings

- Rainho CA, Assumpcao JHM, France FAS, Callegari DP, Barbosa BM, Takeda AAS, Sforcin JM. Interactive e-Poster P20.28.A: *Impact of natural compounds in the DNA methylation of RASSF1A in triple-negative breast cancer cell lines and epi-drug discovery*. In: Abstracts from the 53rd European Society of Human Genetics (ESHG) Conference. Eur J Hum Genet. 2020 Dec 1;28(Suppl 1):141–797. DOI: 10.1038/s41431-020-00739-z.

- Barbosa BM, Martins ACF, Gomes RS, Rainho CA. Abstract B040: *In silico drug design using KRASG12C as a theranostic target for pancreatic cancer*. In: AACR Special Conference: Pancreatic Cancer (Boston, 2022). Cancer Res. 2022 Nov 15;82(22_Supplement): B051. DOI: 10.1158/1538-7445.PANCA22-B040.

- Martins ACF, Barbosa BM, Badenoch B, Gomes RS. Abstract B051: *Design, synthesis, and in vitro evaluation of enedione derivatives as promising selective COPZ1 inhibitor*. In: AACR Special Conference: Pancreatic Cancer (Boston, 2022). Cancer Res. 2022 Nov 15;82(22_Supplement): B051. DOI: 10.1158/1538-7445.PANCA22-B051.

- BARBOSA BM, GOMES RS, RAINHO CA. *Unveiling the heterogeneity of pancreatic cancer: insights from cell-type deconvolution of bulk DNA methylation data*. In: 68th Brazilian Congress of Genetics, 2023, Ouro Preto. 68th Brazilian Congress of Genetics e-book. 2023:90. ISBN: 978-65-89156-08-6.

- BARBOSA BM, FRANCE FAS, ASSUMPCAO JHM, SCHELLINI SA, VIVEIROS MMH, RAINHO CA. *Humulene increases the expression of CXCL8 but not IL-8 secretion in the triple-negative breast cancer cell line BT-20*. In: XVI Congress of MutaGen-Brasil, 2023, Londrina. Procedings XVI Congress of MutaGen-Brasil. 2023:96. ISBN: 978-65-85111-11-9.

- Martins ACF, Barbosa BM, Silveira IOMF, Gomes RS. Abstract 4500: *Enedione derivates as a potential cancer treatment through the inhibition of COPZ1*. In: AACR Annual Meeting (Orlando, 2023). Cancer Res. 2023 Apr 1;83(7_Supplement): 4500. DOI: 10.1158/1538-7445.AM2023-4500.