scientific reports

OPEN

radioGWAS links radiome to genome to discover driver genes with somatic mutations for heterogeneous tumor image phenotype in pancreatic cancer

Dandan Zheng[1]✉, Paul M. Grandgenett[2], Qi Zhang[3], Michael Baine[4], Yu Shi[5], Qian Du[5], Xiaoying Liang[6], Jeffrey Wong[4], Subhan Iqbal[5], Kiersten Preuss[7], Ahsan Kamal[4], Hongfeng Yu[8], Huijing Du[9], Michael A. Hollingsworth[2]✉ & Chi Zhang[5]✉

Addressing the significant level of variability exhibited by pancreatic cancer necessitates the adoption of a systems biology approach that integrates molecular data, biological properties of the tumors, medical images, and clinical features of the patients. In this study, a comprehensive multi-omics methodology was employed to examine a distinctive collection of patient dataset containing rapid autopsy tumor and normal tissue samples as well as longitudinal imaging with a focus on pancreatic cancer. By performing a whole exome sequencing analysis on tumor and normal tissues to identify somatic gene variants and a radiomic feature analysis to tumor CT images, the genome-wide association approach established a connection between pancreatic cancer driver genes and relevant radiomic features, enabling a thorough and quantitative assessment of the heterogeneity of pancreatic tumors. The significant association between sets of genes and radiomic features revealed the involvement of genes in shaping tumor morphological heterogeneity. Some results of the association established a connection between the molecular level mechanism and their outcomes at the level of tumor structural heterogeneity. Because tumor structure and tumor structural heterogeneity are related to the patients' overall survival, patients who had pancreatic cancer driver gene mutations with an association to a certain radiomic feature have been observed to experience worse survival rates than cases without these somatic mutations. Furthermore, the association analysis has revealed potential gene mutations and radiomic feature candidates that warrant further investigation in future research endeavors.

Pancreatic ductal adenocarcinoma (PDAC) is a critical global health problem, with the mortality rate remaining the highest among major solid cancers. Despite decades of clinical and research efforts, the 1-year survival rate is 20% , and the 5-year survival rate remained single digit for many years and only recently rose to 10%[1,2]. Clearly, new and synergistic approaches are needed to battle this ferocious disease.

Pancreatic cancer is often detected at late stages and has a weak response to current chemotherapy and a poor overall prognosis. Some hereditary risks were discovered, suggesting that up to 15% of pancreatic cancer is attributable to genetic causes[3]. Genomic analysis plays an important role in understanding the complex biology of pancreatic cancer development and progression, and in identifying novel treatments targeting specific molecular pathways. However, the hallmark of pancreatic cancer is a high degree of heterogeneity in the biology

[1]Department of Radiation Oncology, University of Rochester Medical Center, Rochester, NY, USA. [2]Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, USA. [3]Department of Mathematics and Statistics, University of New Hampshire, Durham, NH, USA. [4]Department of Radiation Oncology, University of Nebraska Medical Center, Omaha, NE, USA. [5]School of Biological Sciences, University of Nebraska, Lincoln, NE, USA. [6]Department of Radiation Oncology, Mayo Clinic, Jacksonville, FL, USA. [7]Department of Nutrition and Health Sciences, University of Nebraska, Lincoln, NE, USA. [8]School of Computing, University of Nebraska, Lincoln, NE, USA. [9]Department of Mathematics, University of Nebraska, Lincoln, NE, USA. ✉email: dandan_zheng@urmc.rochester.edu; mahollin@unmc.edu; czhang5@unl.edu

of pancreatic tumor progression. Clonal variations were observed in premalignant and malignant tumors that result in different and multiple biological properties of tumors that progress to kill the patient[4,5]. These differences manifest as tumors that progress with different biological properties, which affects the nature of the cells that grow and metastasize and the capacity of these cells to influence and organize their tumor microenvironment. This heterogeneity can be seen both at the molecular level, with non-consensus mutations and gene expression patterns, at the histological level, with different cell types and structures within the tumor, or at the tumor imaging level, with various appearances on CT images[6,7]. For the molecular level of intertumoral heterogeneity, whole exome sequencing analyses revealed a complex mutational landscape for PDAC[8]. Although mutations of some genes, such as KRAS and TP53, occur at rates of up to >50%, the frequency of most other recurrently mutated genes is less than 10%, and there is a long tail of infrequently mutated genes among the PDAC patient population[9,10]. This degree of heterogeneity has previously been underestimated or understated in the literature and in studies that undertake the discovery of biomarkers related to disease progression. It is also reflected by the results of population-based DNA sequencing and RNA expression studies to date, in that no consistent pattern of mutations or RNA expression profiles have yet been defined that accurately predict biological aspects of disease progression[11,12].

Tackling this high degree of heterogeneity in pancreatic cancer demands a system science approach that integrates molecular data, biological properties of the tumors, and clinical features of the patients. The quantitative approaches for medical imaging analysis, such as extracting radiomic features, are perfect for globally assessing the heterogeneity of PDAC at the tumor imaging level[13]. In this work, we present a radiome-wide and genome-wide association approach to identify the driver genes for heterogenicity at the tumor phenotype level. This method was conducted in a unique patient population for which a large amount of tumor and normal tissue samples were collected in a rapid autopsy immediately following the patient's demise. Our work demonstrates the feasibility of this novel systematic approach to providing new insight into the molecular mechanisms of pancreatic cancer progression.

Currently, whole exome sequencing (WES) is widely used to identify cancer-driver genes by searching for genes with a high rate of somatic mutation recurrence in multiple patient samples. What roles do these WES-identified cancer driver genes play in intertumor heterogeneity? This is the scientific question that we wish to answer. We collected a cohort of PDAC patients who had both tumor and healthy tissues from rapid autopsy, and pancreatic contrast-enhanced CT images. WES was conducted on both tumor tissues and healthy tissues. After conducting a comprehensive WES analysis and a tumor image radiomic analysis on these patients, we performed an association study to identify the cancer-driver genes that are significantly associated with image features. Our results shed some new light on tumor genomic and morphological heterogeneity in PDAC.

Results

Patient and dataset information
The patient and tumor clinical characteristics of our studied cohort are listed in Table 1. The patients in the cohort had a median of 6 (range 3–30) serial pancreas contrast CTs available. The CT used for analysis, i.e., the last CT in the series for each patient, was acquired a median of 34 (range 1–579) days before the date of death.

Somatic single-nucleotide variants (SNVs)
Driver somatic SNVs are genetic changes in a cell that drive the development and progression of cancer. In pancreatic cancer, several driver mutations have been identified that contribute to the development of the disease[14,15]. In this study, we only retained translationally consequential SNVs, i.e., the missense variant, stop codon gain, start codon loss, sequence feature, splice donor variant, and intron variant. Detailed information on these somatic SNVs obtained from WES for this population is shown in Supplementary Table S1. The number of SNV recurrences in the above categories was counted for the patient population. Based on the single SNV recurrence, the mutations Chr12:25245350 C->T|A (G12D and G12V mutations) on KRAS had the highest recurrence rate, in 14 out of 26 patients (54%). These mutations, G12D and G12V, have been reported as the most common ones in pancreatic cancer, recurring at about 45% and 35%, respectively[16,17]. The second largest recurrence is 9 (35%), for the SNV Chr12:25245351 C->G|A (G12R and G12C mutations) also on KRAS. These two SNVs caused the same missense variant on the amino acid sequence. Subsequently, the SNVs, Chr2:130074357 on *POTEF* and Chr7:152358679 on *KMT2C* had 7 (27%) and 6 (23%) recurrences out of 26 patients, respectively. For each gene, we also calculated the number of individuals carrying variants in any given gene. The top-ranked genes are *KRAS* (22), *TP53* (17), *KMT2C* (17), *LRP1B* (14), *FGFR2* (13), *RGPD3* (11), *EWSR1* (10), and *RGPD4* (10) (Supplementary Table S2).

The *KRAS* gene had SNVs in 22 out of all 26 patients (84.6%). This agrees with the discovery that KRAS mutations are found in more than 90% of PDACs[18–20]. For gene *TP53*, 17 out of 26 patients (65.4%) had mutations, and previous studies showed that *TP53* mutations were found in approximately 50% of PDACs and are associated with a poor prognosis. 17 out of 26 patients (65.4%) also had *KMT2C* mutations. Histone Lysine Methyltransferase (KMT2) family genes are frequently mutated in multiple cancer types[21]. Histone Lysine Methyltransferase 2C (KMT2C), also known as myeloid/lymphoid or mixed-lineage leukemia protein 3 (MLL3), is among the most frequently mutated cancer genes in major cancer types[22,23]. These genes have somatic mutations in most PDACs.

Gene *CDKN2A* had SNVs in 9 out of all 26 patients (34.6%). Previous studies showed that CDKN2A mutations are found in approximately 29% of PDACs and are associated with a poorer prognosis[24]. CDKN2A is a tumor suppressor gene that encodes the p16INK4A protein (hereafter mentioned as CDKN2A). As in its name, CDKN2A is a negative regulator of cell cycle progression (the G1-to-S phase transition) by disturbing the complex formation between CDK4/6 and cyclin D[25,26].

Characteristic	Number of patients (percentage) Median (range)
Gender	
Male	16 (60%)
Female	10 (40%)
Median age (range)	71 (40–89)
Tumor site in pancreas	
Head	20 (77%)
Neck	1 (4%)
Tail	1 (4%)
Body	4 (15%)
Stage at diagnosis	
II	9 (35%)
III	4 (15%)
I V	13 (50%)
Use of radiotherapy	
Yes	8 (31%)
No	18 (69%)
Use of chemotherapy	
Yes	25 (96%)
No	1 (4%)
Survival days	331 (27–2282)

Table 1. Patient and tumor clinical characteristics.

Radiomic features

Radiomic features can be used to capture the heterogeneity of the tumor phenotype in the context of pancreatic cancer. Texture analysis can be used to quantify the spatial arrangement of pixel intensities within the tumor. With close-to-a-thousand radiomic features, a total of 944, extracted from each tumor volume-of-interest (VOI), a heatmap was generated to show the tumor radiomic feature pattern of the studied population (Fig. 1). The detailed data on radiomic features for patients in this cohort are listed in Supplementary Table S3. The radiomic feature pattern did not show direct correlations with the patient clinical data listed in Table 1. After feature selection using a recursive correlation pruning step for clustering with a correlation coefficient cutoff = 0.8, 170 representative radiomic features were kept for final clusters. Out of the 170 radiomic features, 59 features had a coefficient of variation (CV), the ratio of the standard deviation to the mean, greater than 1, indicating higher heterogeneity among the studied population. These 170 radiomic features and their CVs are shown in Supplementary Table S4. Radiomic feature wavelet.LHH.firstorder.Skewness had the largest CV, at 35.9. Figure 2 shows 3 tumor images with different values of wavelet.LHH.firstorder.Skewness. This feature is related to the asymmetry of the distribution of pixel values after applying a wavelet filter. The large CV indicates the high heterogeneity of the tumor pixel gray level distribution. The smallest CV including glcm.Idmn (5.98×10^{-3}), glcm.InverseVariance (4.04×10^{-2}), and original.shape.Sphericity (9.65×10^{-2}), indicating high homogeneity of these features among all patients. Feature IDMN (inverse difference moment normalized) is used to assess the local homogeneity of
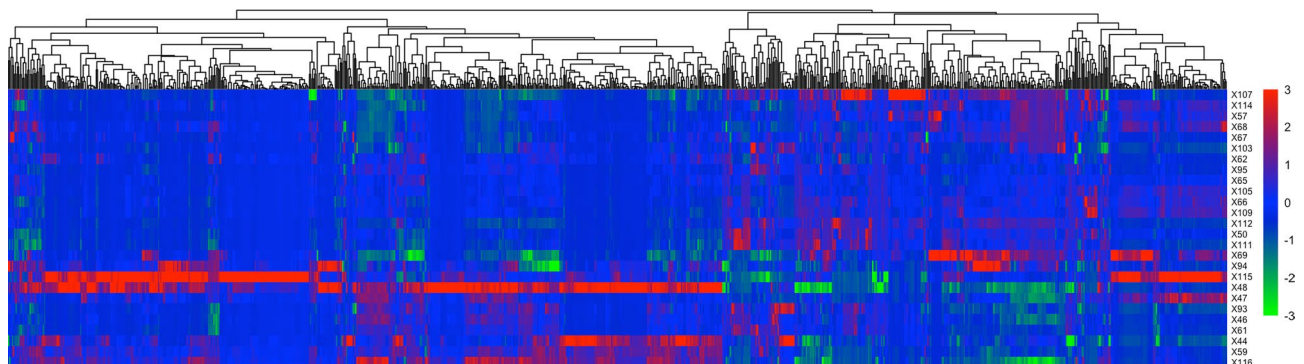

Figure 1. Radiomic features. Each tumor from a patient (row) has 944 radomics features (column). The color indicates the Z-score of a patient for a given radiomic feature. The heatmap was generated using the function *heatmap.2()* from the R package of gplot (3.1.3.1) (https://search.r-project.org/CRAN/refmans/gplots/html/heatmap.2.html).
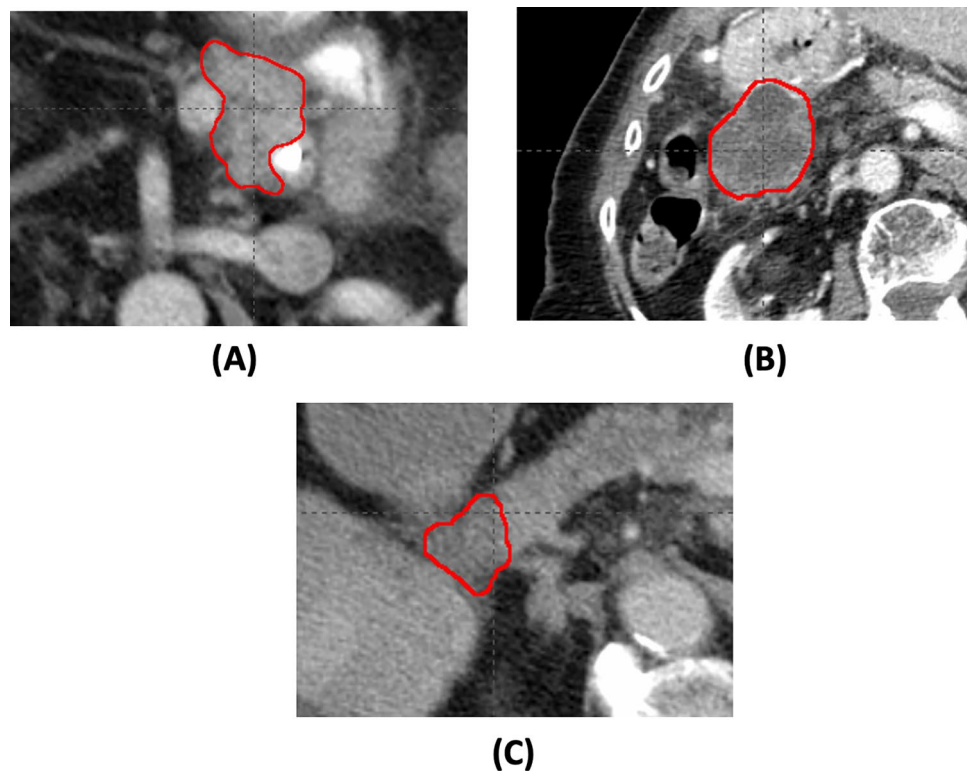
Figure 2. Example tumor images of three patients. The corresponding tumor images have the value of the radiomic feature, wavelet.LHH.firstorder.Skewness, -4.14 (**A**), -2.31×10^{-3} (**B**), and 1.31 (**C**), respectively.

VOI. Because all tumor images have very low local homogeneity, the values of IDMN for all tumor images were high, > 0.98. Sphericity is a measurement of the roundness of the tumor region's morphology relative to a sphere. It is a measure without dimensions, independent of scale and orientation. The majority of tumors in this data set are in their later stages, and hence, the tumors are not in a round shape.

Radiome-wide and genome-wide association

This association analysis found several significant associations (P-value $< 10^{-4}$) between radiomic features, from all 944 radiomic features, and genes with many somatic variants in the tumors. Figure 3 shows the distribution of all P-values of the associations between radiomic features and genes. Table 2 lists the radiomic features and genes with somatic variants with significant associations. Interestingly, there was no significant association found between radiomics features and driver genes with a high recurrence frequency of somatic variants, such as *KRAS*
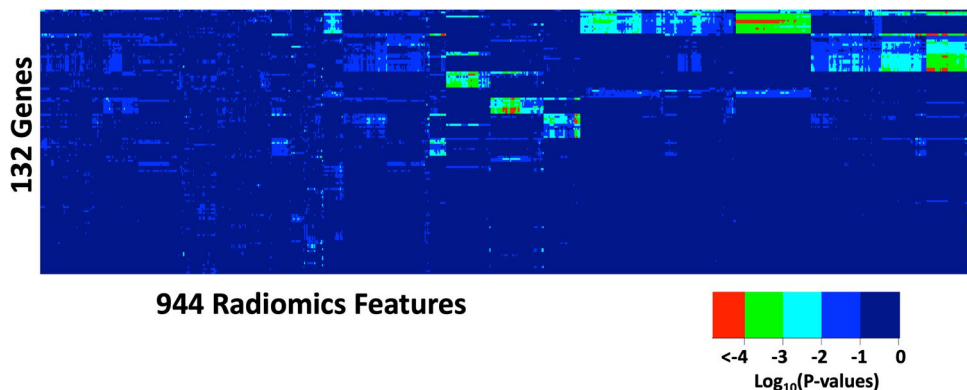

Figure 3. P-values of the associations between radiomic features and genes. Association studies were conducted between all 944 radomics features (column) and 132 genes with many somatic variants in the tumors (row). The color indicates the value of $\log_{10}$(P-values). The heatmap was generated using the function *heatmap.2()* from the R package of gplot (3.1.3.1) (https://search.r-project.org/CRAN/refmans/gplots/html/heatmap.2.html).

Gene	Recurrence frequency	Radiomic features (descriptions)	P-values
EGF	6	original.glrlm.RunVariance (a measure of the variance in runs for the run lengths of the original image's GLRLM, which quantifies gray level runs, defined as the length in number of pixels, of consecutive pixels that have the same gray level value)	2.36×10^{-5}
EPHB3	3	original.glrlm.RunVariance	2.97×10^{-5}
POTEF	7	original.glrlm.RunVariance	3.42×10^{-5}
UBC	4	wavelet.HHH.firstorder.Skewness (a measure of the asymmetry of the distribution of values about the mean value on the wavelet transformed image with an HHH filter)	2.69×10^{-5}
POTEJ	3	wavelet.HHH.firstorder.Skewness	4.91×10^{-5}
EGFR	3	wavelet.HHH.firstorder.Skewness	8.55×10^{-5}
CHD2	3	wavelet.HHL.glszm.LargeAreaHighGrayLevelEmphasis (a measure of the proportion in the image of the joint distribution of larger size zones with higher gray-level values on the GLSZM of the wavelet transformed image with an HHL filter)	3.50×10^{-5}
EPOR	3	wavelet.HHL.glszm.LargeAreaHighGrayLevelEmphasis	4.79×10^{-5}
SDHAP1	6	wavelet.HHL.glszm.LargeAreaHighGrayLevelEmphasis	9.40×10^{-5}
RANBP2	5	original.glszm.LargeAreaEmphasis (a measure of the distribution of large area size zones on the original image's GLSZM)	7.06×10^{-5}
KDR	5	wavelet.HLH.ngtdm.Busyness (A measure of the change from a pixel to its neighbor on the NGTDM of the wavelet transformed image with an HLH filter)	4.52×10^{-5}
JUN	3	wavelet.HLH.ngtdm.Busyness	8.31×10^{-5}
NOTCH1	4	wavelet.HLH.ngtdm.Busyness	8.31×10^{-5}
ARID1A	3	wavelet.LHH.glcm.ClusterProminence (a measure of the skewness and asymmetry of the GLCM of the wavelet transformed image with an LHH filter)	8.63×10^{-5}
RGPD6	8	wavelet.LHH.glcm.ClusterShade (a measure of groupings of voxels with similar gray-level values of the GLCM of the wavelet transformed image with an LHH filter)	3.91×10^{-5}
KIT	3	wavelet.LHH.glcm.ClusterShade	4.01×10^{-5}
GFRA2	3	wavelet.LHH.glcm.ClusterShade	5.70×10^{-5}
FUBP1	3	wavelet.LHH.glcm.ClusterShade	8.63×10^{-5}
FANCB	4	wavelet.LLH.glcm.ClusterShade (same as above, with an LLH filter)	1.56×10^{-5}
SPEN	4	wavelet.LHL.ngtdm.Contrast (a measure of the spatial intensity change of the NGTDM of the wavelet transformed image with an LHL filter)	1.46×10^{-5}
PRKG1	3	wavelet.LHL.ngtdm.Contrast	5.81×10^{-5}
CDKN2A	9	wavelet.LHL.ngtdm.Contrast	7.74×10^{-5}
BCORL1	4	wavelet.LHL.ngtdm.Contrast	9.18×10^{-5}
KMT2B	6	wavelet.LHL.ngtdm.Contrast	9.31×10^{-5}

Table 2. Associations between genes and radiomic features.

and *TP53*. Genes that had a significant association with radiomic features have a recurrent rate of 3 (11.5%) to 7 (26.9%) among the patients.

These genes that are significantly associated with radiomic features are related to tumor formation and progress. For example, the *EGF* and *EGFR* genes are important for cancer cell proliferation and spread in the body[27,28]. The *EGFR* gene with somatic SNVs is associated with wavelet.HHH.firstorder.Skewness which indicates the tumor pixel value distribution asymmetry on the CT images. Mutations in the *EGFR* gene, which encodes epidermal growth factor receptors, enable cancer cells to grow and proliferate. The expression and function of the mutant *EGFR* gene may contribute to the varying patterns of tumor growth, leading to the distinct skewness among patients. Some transcription factor genes, like RGPD6, which was the most commonly mutated in other types of cancers[29], also had a significant association with the radiomic feature, wavelet.LHH.glcm.ClusterShade, which is a descriptor of the tumor texture pattern.

Several genes are associated with the same radiomic feature. This indicates these genes may be involved in the same biological pathway or are involved in synergistic interactions. For example, *NOTCH1*, *JUN*, and *KDR* genes were all associated with the radiomic feature, wavelet.HLH.ngtdm.Busyness. Literature suggests that both *Notch* and *JUN* genes are related to cell apoptosis[30], and Notch-1 promotes JNK/c-Jun activation[31]. Loss of function of either JUN or NOTCH-1 can result in similar observed biological effects. In our cohort, only one patient carried somatic mutations in both the *JUN* and *NOTCH1* genes, suggesting that the two mutations may be an alternative. Gene *KDR* encodes the Kinase insert domain receptor, also known as vascular endothelial growth factor receptor 2 (VEGFR-2). The corresponding radiomic feature, ngtdm, is a Neighboring Gray Tone Difference Matrix that quantifies the difference between a gray value and the average gray value of its neighbor pixels. Busyness is a measure of the change from a pixel to its neighbor. A high value for busyness indicates rapid changes in intensity between pixels and their neighborhoods. This indicates that somatic SNVs in *NOTCH1*, *JUN*, and *KDR* genes may cause cancer cells to develop at various speeds and result in varying localized colonization of different cell types in the tumor.

Three genes, *EGF*, *EPHB3*, and *POTEF*, were all associated with original.glrlm.RunVariance. Three patients have SNVs in at least two of the three genes. A gray level run length is defined as the number of consecutive pixels that have the same gray level value. The gray level run length matrix (GLRM) consists of all run lengths in a VOI. Run Variance is the measure of the variance in runs for the run lengths, and hence, is related to the intratumor heterogeneity. As an example, Fig. 4 shows the images of two tumors with or without an *EGF* gene
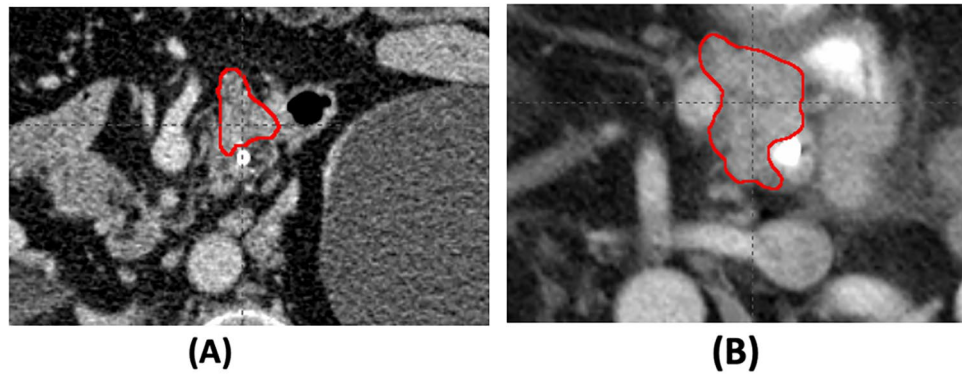
Figure 4. Tumor image of two different patients with (**A**) or without (**B**) an *EGF* gene mutation. The CT scans were obtained from two patients, one of whom had SNVs on the *EGF* gene, whereas the other did not. The corresponding tumor images have the value of the radiomic feature, original.glrlm.RunVariance, 0.249 (**A**) and 0.992 (**B**), respectively.

mutation. The association between these three genes, *EGF*, *EPHB3*, and *POTEF*, and Run Variance indicates their roles in the heterogeneous colonization of cancer cells. For example, the gene, *EPHB3*, has been found to be involved in the signaling conduction of colonizing cells[32].

Genes, including *SPEN*, *PRKG1*, *CDKN2A*, *BCORL1*, and *KMT2B*, had a significant association with wavelet. LHL.ngtdm.Contrast, which is a texture feature. The Neighboring Gray Tone Difference (NGTD) is defined as the difference between a gray value and the average gray value of its neighbors within a distance. The value of Contrast is a measure of the local intensity variation and the spatial intensity change. Figure 5 shows an example of the tumor CT images of two patients with or without the SNV in the *SPEN* gene. Its high value indicates that the tumor region has large changes between voxels and their neighborhood. This association suggests that these five genes are involved in tumor growth and tumor shape regulation. Especially, the gene *CDKN2A* has somatic SNVs in 9 patients (34.6%) out of 26 patients. The gene *CDKN2A*, whose gene product is the cyclin-dependent kinase inhibitor 2A, plays an important role in cell cycle regulation and demonstrates tumor suppressor activity. Inactivation of CDKN2A leads to uncontrolled cell growth[33].

Discussion

Previous whole-genome sequencing and variation analysis discovered that mutations on genes, *KRAS*, *TP53*, *CDKN2A*, *ARID1A*, *ROBO2*, *KDM6A*, and *PREX2*, are important in pancreatic cancer[34]. Activating mutations of KRAS are nearly ubiquitous, being found in more than 90% of PDACs[18–20]. Among all cancers, KRAS mutations are present in ~ 25% of tumors[35] and frequently in lung, colorectal, and pancreatic cancers[18,36–38]. Actually, the RAS family, including KRAS, NRAS, and HRAS, is the most frequently mutated gene family in all different types of cancers[16]. The inactivation of TP53 reoccurs at rates of > 50% in pancreatic cancer. The gene, *TP53,* is the most frequently mutated gene in different cancers at rates ranging from 38 to 50%, such as ovarian, esophageal, colorectal, head and neck, larynx, lung, and pancreatic cancers[39]. These genes, KRAS and TP53, are important
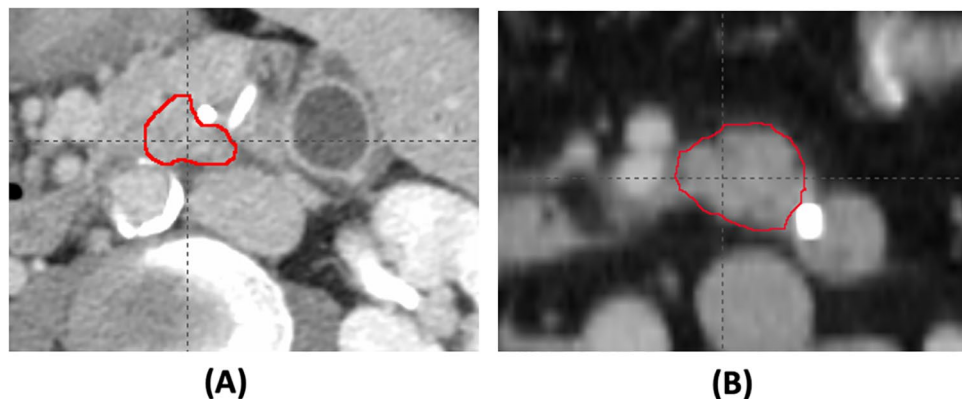

Figure 5. Tumor image of two different patients with (**A**) or without (**B**) a *SPEN* gene mutation. The CT scans were obtained from two patients, one of whom had SNVs on the *SPEN* gene, whereas the other did not. The corresponding tumor images have the value of the radiomic feature, wavelet.LHL.ngtdm.Contrast, 0.631 (**A**) and 3.02×10^{-2} (**B**), respectively.

in the formation of cancer and hence have been extensively studied before, but they are not often directly related to intertumor or intratumor heterogeneity. In pancreatic cancer, the prevalence of recurrently mutated genes then drops to ~ 10% for a handful of genes involved in chromatin modification, DNA damage repair, and other mechanisms resulting in significant intertumoral heterogeneity[34]. These driver mutations can be used to inform diagnostic and treatment strategies for pancreatic cancer and provide a better understanding of the underlying biology of the disease. To understand the role of these genes with low recurrence mutations, the association between gene-mutation profiles and radiomic features was examined in this work. The significant association between sets of genes and radiomics features identified genes that contribute to morphological heterogeneity. These genes include *NOTCH1*, *JUN*, *KDR*, *EGF*, *EPHB3*, *POTEF*, *SPEN*, *PRKG1*, *CDKN2A*, *BCORL1*, and *KMT2B*.

This work discovered that several genes with mutations have significant associations with the same radiomic features. Some genes associated with the same radiomic feature are in the same regulatory pathway and work together for a regulatory cascade. We conducted enrichment analysis on 24 genes shown in Table 2 against both the Gene Ontology (GO) database[40] and the KEGG database[41]. Figure 6 shows several genes are involved either in the same biological process described by the GO term or in the same KEGG pathway. For example, the KEGG database annotated both *NOTCH1* and *JUN* in the same pathway, hsa01522, "Endocrine resistance". *NOTCH1* and *JUN* are associated with wavelet.HLH.ngtdm.Busyness and the gene product NOTCH1 promotes JNK/c-Jun activation[30,31]. Loss of function of either *JUN* or *NOTCH1* can result in cells evading cellular death pathways. Gene *JUN* had somatic mutations in three patients, and *NOTCH1* gene had somatic mutations in four patients. Only one patient had mutations in both *JUN* and *NOTCH1* genes, and this patient had 220 somatic mutations, which is much more than the average number (51) of somatic mutations per patient in our dataset. In contrast to these alternative genes, the other genes that were associated with the same radiomic feature appear to work in different pathways, and several somatic mutations in multiple genes need to work together. For example, *EGF*, *EPHB3*, and *POTEF* are all associated with original.glrlm.RunVariance, but *EGF* works for cell proliferation and *EPHB3* works for colonizing cells. In this case, SNVs in multiple genes have a high chance to occur in the same patient. The knowledge of various recurrence patterns of cancer driver genes associated with a specific pheno-typic feature has the potential to guide combination therapy for cancer using multi-target medicines, which has recently garnered a great deal of attention as one of the most promising cancer-fighting tools[42,43]. In our novel approach, radiomic features expand the big data space that we could integrate and leverage for novel discoveries.

Tumor heterogeneity can refer to intratumor heterogeneity, including heterogeneity of structures within a single tumor, or intertumor heterogeneity if tumors are compared among patients. The tumor structural hetero-geneity can be quantified by medical images[44], especially via radiomic features[45,46]. Radiomic features, such as tumor shape and texture features, can be used to quantify tumor heterogeneity, which gives them the potential to serve as imaging-based heterogeneity biomarkers[47]. Texture analysis can be used to quantify the spatial arrange-ment of pixel intensities within the tumor. Usually, tumor image textural features could be extracted in several
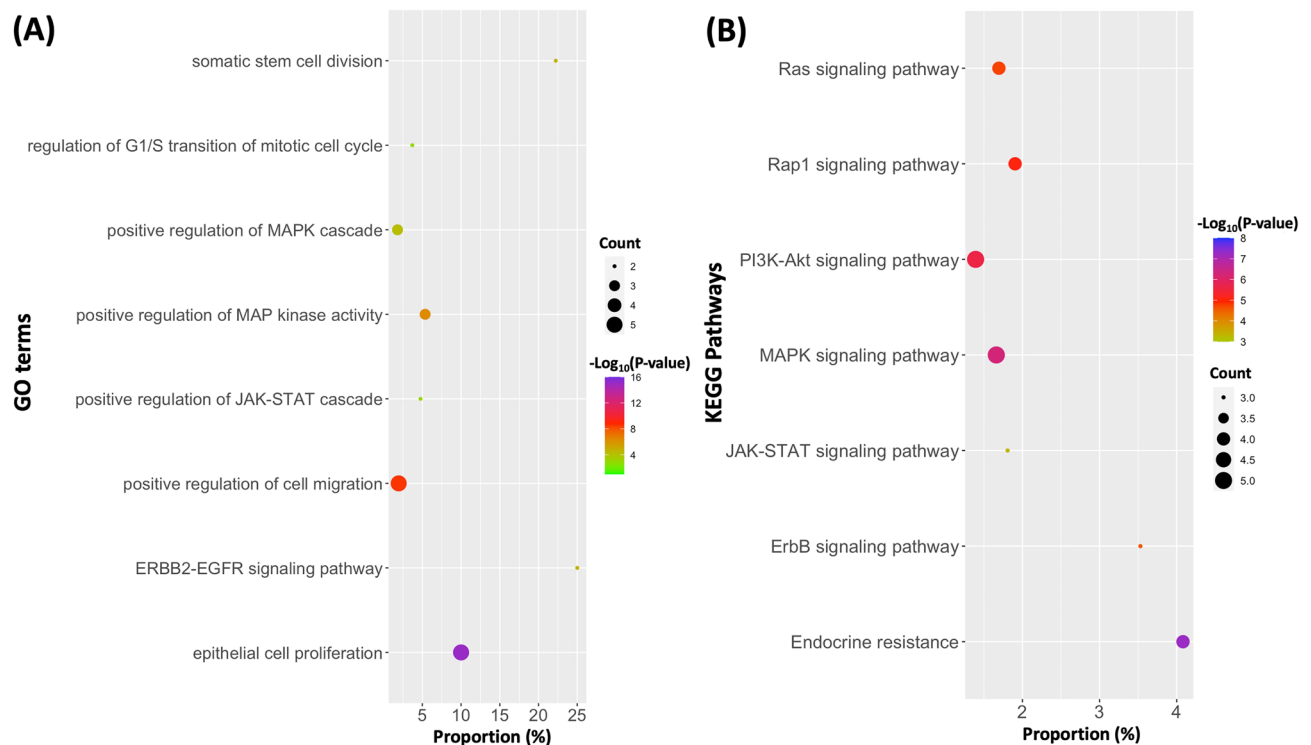

Figure 6. Enrichment of significant genes in GO terms and KEGG pathways. In all significant genes, several (2–5) genes are involved in the same biological processes in the GO database (**A**) or the same pathways in the KEGG database (**B**). Y-axis shows the names of GO terms or KEGG pathways, and X-axis is the percentage of our target genes in all genes annotated by a given GO term or KEGG pathway.

different ways, such as using GLRM-based approaches. In this study, we found a significant association between pancreatic cancer driver genes, *EGF*, *EPHB3*, and *POTEF*, and original.glrlm.RunVariance. This association links the molecular level mechanism and their outcomes at the level of tumor structural heterogeneity. Because tumor structure and tumor structural heterogeneity are related to the patients' overall survival, patients who had pancreatic cancer driver gene mutations with an association to a certain radiomic feature showed worse survival rates than cases without those somatic mutations. For example, we used a public cohort of pancreatic cancer in the TCGA database[48] and collected the survival information of patients with somatic mutations on genes, *CDKN2A*, *PRKG1*, and *BCORL1*, which had a significant association with wavelet.LHL.ngtdm.Contrast from our study. Figure 7 shows the survival curves comparing the patients with somatic mutations on genes, *CDKN2A*, *PRKG1*, and *BCORL1*, and the other patients without. The survival time of patients with somatic mutations in these driver genes is shorter than that of the other patients (FDR-adjusted P-value = 0.0348).

In this study, we explored genome-wide and radiome-wide association investigation. Screening hundreds of thousands of genetic variants across the entire genome, genome-wide association studies (GWAS) have been widely used to identify disease-specific genetic variants and use them in broad clinical and biological applications. For pancreatic cancer, previous GWAS have identified new and useful risk loci[49]. For this extremely lethal disease, these types of findings can be especially important owing to the complexity as well as the dynamic heterogeneity over the rapid progression of the disease. However, GWAS are expensive to conduct and require tissue samples that are not readily available. For pancreatic cancer, this is especially the case because tissue samples are usually only procured during surgery owing to the high risks associated with biopsy, and only around one-fifth of all pancreatic cancer patients are operable due to late stages at detection[50]. The association study approach described in this work can help link the radiome with the genome, i.e., link phenotypic radiomic information from medical images with genotypic information, therefore enabling a much broader and longitudinal genome-wide search for this highly dynamic disease.

The Pancreatic Cancer Rapid Autopsy provides a unique dataset that allows a comprehensive investigation of pancreatic cancer with a systems approach. In this proof-of-concept study, a genome-wide and radiome-wide association analysis was conducted on whole-exome sequencing data from both primary tumor and normal pancreatic tissue from the rapid autopsy. Our approach normalizes the SNVs identified on tumor tissue by those on normal tissue, i.e., the approach teases out all somatic SNVs and selects only tumor SNVs. This way, each patient acts as their own control, thereby suppressing the immense background noise and focusing only on the tumor-specific genomic signals. Potentially, similar "normalization" approaches could be applied to zoom in on the genomic changes between the primary tumor and each of the metastatic tumors, and the temporal changes of all tumors over the course of disease progression and treatments. For the former, the large tissue collection from the rapid autopsy is uniquely valuable by providing primary and all metastatic lesions as well as normal organ tissues. This radiome-genome association approach we established in this work could facilitate these investigations with relevant radiomic features. For the latter, direct genomic assessment is not possible as tissue samples cannot be collected repeatedly along the time course. On the other hand, because periodical medical imaging is already part of the cancer care routine, the relevant radiomic features can be used as surrogates to assess the longitudinal genetic changes accompanying the rapid and heterogeneous progression of this vicious disease. Together, this approach and the future investigations it enables may help decipher the mechanisms and pathways of how pancreatic cancer cells progress and respond to treatments and shed light on better treatment options.

Radiogenomics is an existing branch of radiomics[51]. Combining genomic data and imaging features has been shown to yield imaging biomarkers and provide valuable information for diseases, especially cancer. However,
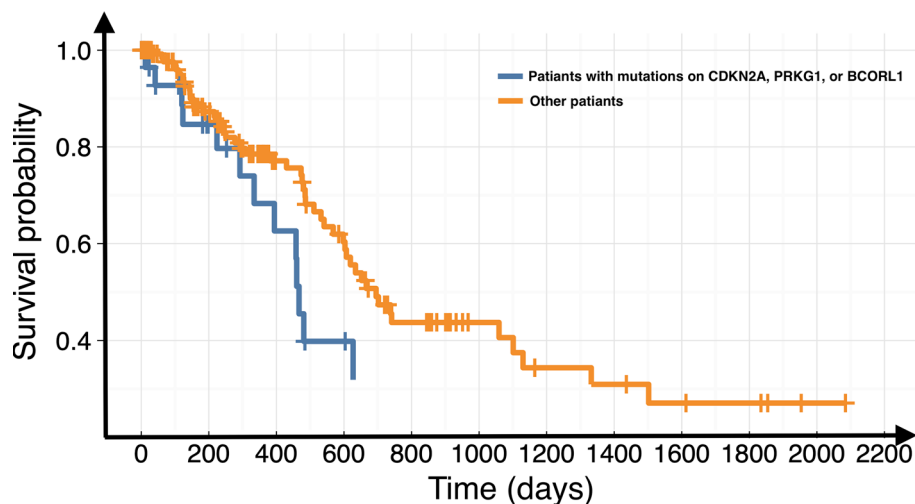

Figure 7. Survival curves for patients with different SNVs. Patients (blue) with somatic SNVs on genes, *CDKN2A, PRKG1,* and *BCORL1,* have shorter overall survival time than other pancreatic cancer patients (orange).

among different cancers, there is a relative paucity of radiogenomics literature for pancreatic cancer, largely due to the limited known molecular markers for this highly heterogeneous disease and the difficulty of obtaining simultaneous imaging data and genomic data at well-synchronized time points for this rapidly progressing disease. Furthermore, all existing radiogenomics studies, including those on pancreatic cancer[52], focused only on known molecular markers. In contrast, our genome-wide and radiome-wide association study applies a system-wide search through large-scale radiomic and genomics data to explore novel imaging and genomic biomarkers and mechanisms. In this proof-of-concept preliminary study, radiomics and primary tumor whole genome information were correlated for pancreatic cancer patients. Using the last pancreatic CT scan of the patient, the imaging date was reasonably close to the date of death when the tissues used for genomic investigations were collected through rapid autopsy, comparable in quality to that obtained by surgical resection.

While the concept of a system-wide large-scale radiome- and genome-wide association study is innovative and the results encouraging, the study is not without limitations. First, the cohort size is rather small, with only 26 patients. A validation dataset was also not available to strengthen the robustness of our findings. We hope that our novel work and proof-of-concept findings will catalyze future development and examination of such datasets, which are crucial for advancing multiomics integration studies. Our dataset size is limited by data availability and more so by the substantial cost associated with the whole-exome sequencing of each tumor and tissue sample. This cost highlights the potential benefits of this multiomcis association approach to developing imaging surrogates for these costly large-scale screenings. At the same time, as our study applied the somatic SNVs of each patient as their own control, tumor SNVs can be identified with much higher specificity in even a small cohort. Another limitation is the slight heterogeneity of the imaging data in terms of the CT scanner model and acquisition protocol, as well as the time point of the scan relative to tissue sample collection. These variances are inevitable as the unique dataset came from the Pancreatic Rapid Autopsy Program, which is retrospective and curated carefully over more than a decade. On the other hand, coming from a single institution, the CT scanners used for this study were all from a single vendor and on a single line of models, and the imaging protocols were largely similar. The imaging dates for the cohort ranged from 1 to 224 days before patient death but were relatively synchronized with the tissue collection with imaging dates within 1 or 2 months for most patients. This timing misalignment between the genetic samples and imaging samples could potentially act as a confounding variable, especially for patients with longer time elapses between the two samples and those with more rapid progression and mutation. Our novel exploration and proof-of-concept findings will hopefully help motivate future curation of better synchronized biological and imaging data, for example, by adding clinical imaging of the deceased patient right before the autopsy. For large-scale omics research, false positives are always a potential challenge, especially when the sample size is small. In our study, we employed some strategies to minimize false positives, such as applying the gene-based burden-testing approach and selecting a stricter significance level at P-value $= 10^{-4}$, which was used or suggested by other works in GWAS[53,54].

Future applications of this novel methodology could include genome and radiome association studies on metastatic lesions to investigate their similarities and differences with the primary pancreatic tumor in terms of these molecular and imaging metrics. This could further validate the new approach and help us understand the complex tumor mutations that occur during the progression of pancreatic cancer. The application of the discovered imaging biomarkers to additional patients as well as the longitudinal images of these patients would be an additional investigation of value. The latter may shed light on the temporal changes of relevant genomic markers as the disease progresses and the patient responds to treatment. Currently, marker genes' expression levels are used to identify molecular subtypes of PDAC that are widely accepted[18,55]. For example, the hypermethylated *EGFR* gene, which is associated with the radiomic feature wavelet.HHH.firstorder.Skewness, indicates the subtype of pancreatic progenitor. Therefore, it is important to explore the radiomic features concerning the molecular subtypes of PDAC in the future.

Materials and methods
Study population
Patients were included in this study from a unique database of the University of Nebraska Medical Center Pancreatic Cancer Rapid Autopsy Program. For over a decade, the program has been collecting large quantities of tumor and tissue samples from autopsies performed within hours of patient demise. In the rapid autopsy, all primary and metastatic tumors and a large number of tissue samples, such as liver, lung, spleen, and kidney, are collected under rapid conditions that produce tissue that is comparable in quality to that obtained by surgical resection. The resected autopsy samples are reviewed and annotated by at least two pathologists, in concert with lab members who conduct the autopsies. Twenty-six patients from the pancreatic cancer rapid autopsy program, with comprehensive, unique tissue sample collection and longitudinal contrast-enhanced CT images, were included in this study. All data collection was approved by the Institutional Review Boards (IRB) of the University of Nebraska Medical Center (Protocols: 728-16-EP and 127-18-EP), and all methods were performed in accordance with the relevant guidelines and regulations.

DNA isolation and whole exome sequencing (WES)
DNAs were extracted from tumor tissues and healthy tissues in the liver, kidney, etc. Illumina TruSeq DNA Exome kit was used for exon capture. Sequencing was carried out using Illumina 2×100 bp paired-end sequencing on a HiSeq 2500 instrument according to the manufacturer's recommendation.

Genome-wide identification of somatic single-nucleotide variants (SNVs)
When applied to many samples of the same cancer type, the identification of the cancer driver gene can be conducted to search for multiple recurrences of somatic mutations in the same gene. With the WES data for

the tumor and tumor-free organ tissues from 26 patients, tumor-specific somatic SNVs were identified with VarScan2[56] after the standard read preprocessing and read-mapping by BWA[57]. Based on the FDR-adjusted P-values calculated by VarScan2, we retained the significant somatic mutations with a cut-off of the adjusted P-value $< 10^{-5}$ for subsequent analyses.

Imaging studies

For the 26 patients included in the study, varying numbers (3–30) of contrast-enhanced abdominal CT scans were acquired per standard pancreatic cancer care from diagnosis to longitudinal monitoring, using Lightspeed VCT, Lightspeed Pro 16, or Lightspeed RT16 (GE Healthcare, Boston, Massachusetts, USA). For the image acquisition, patients received ISOVUE injection with bolus triggering arterial phase imaging about 30 s and venous phase about 60 s after injection. These scans used a slice thickness of 1.5–5 mm with an in-plane resolution of 0.6–0.8 mm. For the purpose of this study, the last available CT scan prior to the patient's death was used for radiomic analysis for the patient. This way, we could get the closest match between radiomic information from the imaging and the genomic information from the rapid autopsy.

Radiomic feature extraction

Pancreatic tumor volume-of-interest (VOI) was manually segmented by two experienced clinical investigators using a consistent window/level setting and reconciled disagreements to mitigate intra- and inter-observer uncertainty. From each tumor, VOI, 944 radiomic features were extracted using the radiomic module on 3D Slicer (version 4.10)[58] and visualized using an interactive visualization platform. A resampled $2 \times 2 \times 2$ mm^3 voxel size and a bin width of 25 were used for feature extraction. The features are defined in compliance with feature definitions as described by the Imaging Biomarker Standardization Initiative (IBSI)[59] and can be divided into original features (107 features), Laplacian of Gaussian features (LoG, 93 features), and wavelet features (744 features). The original features can be subdivided into 6 classes, including 14 Shape features, 18 First Order statistical features, 38 Gray Level Dependence Matrix (GLDM) features, 16 Gray Level Run Length Matrix (GLRLM) features, 16 Gray Level Size Zone Matrix (GLSZM) features, and 5 Neighboring Gray Tone Difference Matrix (NGTDM) features. The wavelet features included all except Shape features calculated on the filtered images with all 8 combinations of applying either a High or a Low pass filter in each of the three dimensions. All features are pre-selected to eliminate features unstable to respiratory motion and inter-observer contouring uncertainty[60].

Genome-wide association analysis between radiomic features and somatic mutations

Based on our discovered genes with a high reoccurrence rate of somatic mutations and radiomic features from the corresponding tumor, we conducted an association study between these genes and radiomic features of CT scans from the same population. Here, we employed the gene-based burden-testing approach. For this approach, we used the number of individuals carrying variants in each gene to associate with traits in cohorts[61]. We applied the sequence kernel association test (SKAT)[62] to test the association between each radiomic feature and somatic mutations within each gene. SKAT was originally designed to test the association between a trait and the rare variants in a genomic region and was based on a variance-component score test in a mixed-model framework. It was shown to have much higher power than many other burden tests for gene-based GWAS. We focused on the 132 genes with somatic mutations in at least three patients (the specific mutations can be different among these patients) and tested their association with 944 radiomic features. Here, we used principal component analysis (PCA) to estimate the population structure in our dataset. The population structure can be addressed by including principal components (PCs) as covariates[63,64]. The top two PCs of the somatic mutation matrix are used as covariates in the null model. In our dataset, the reoccurrences of individual somatic mutations are generally low due to the small sample size. Aggregating their potentially heterogeneous effects using SKAT is expected to improve the detection power. The output p-values were adjusted for multiple tests based on Benjamini–Hochberg procedure.

TCGA data and survival analysis

The R package, RTCGA, (https://rtcga.github.io/RTCGA/) and the full set of somatic mutations discovered by TCGA data for the cohort of Pancreatic Cancer from Firehose (https://gdac.broadinstitute.org/)[48] were used to get mutation and survival information of 185 Pancreatic Cancer patients for survival analysis. The function, *kmTCGA()*, in RTCGA was used to plot Kaplan–Meier estimates of survival curves for survival data from patients with or without given mutations. The function, *pairwise_survdiff()*, in the R package of survminer was used to have the comparisons of multiple survival curves.

Informed consent

Informed consent was obtained from all subjects and/or their legal guardian(s). All methods were performed in accordance with the relevant guidelines and regulations.

Data availability

The raw sequences of WES that were used in this study are available in the NCBI BioProject database with links to BioProject accession number PRJNA1041040 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1041040). All somatic SNV and radiomic data that were generated and analysed in this study are included in this article and the Supplementary Information files.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30. https://doi.org/10.3322/caac.21442 (2018).
2. Bengtsson, A., Andersson, R. & Ansari, D. The actual 5-year survivors of pancreatic ductal adenocarcinoma based on real-world data. *Sci. Rep.* **10**, 16425. https://doi.org/10.1038/s41598-020-73525-y (2020).
3. Klein, A. P. Genetic susceptibility to pancreatic cancer. *Mol. Carcinog.* **51**, 14–24. https://doi.org/10.1002/mc.20855 (2012).
4. Hanahan, D. Hallmarks of cancer: New dimensions. *Cancer Discov.* **12**, 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059 (2022).
5. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: Implications for targeted therapeutics. *Br. J. Cancer* **108**, 479–485. https://doi.org/10.1038/bjc.2012.581 (2013).
6. Al-Hawary, M. M. *et al.* Pancreatic ductal adenocarcinoma radiology reporting template: Consensus statement of the Society of Abdominal Radiology and the American Pancreatic Association. *Radiology* **270**, 248–260. https://doi.org/10.1148/radiol.13131184 (2014).
7. Chakraborty, J. *et al.* Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients. *PLoS ONE* **12**, e0188022. https://doi.org/10.1371/journal.pone.0188022 (2017).
8. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405. https://doi.org/10.1038/nature11547 (2012).
9. Yang, S. *et al.* Detection of mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. *Cancer Biol. Ther.* **18**, 158–165. https://doi.org/10.1080/15384047.2017.1281499 (2017).
10. Qian, Y. *et al.* Molecular alterations and targeted therapy in pancreatic ductal adenocarcinoma. *J. Hematol. Oncol.* **13**, 130. https://doi.org/10.1186/s13045-020-00958-3 (2020).
11. Ahmad, J. *et al.* Disease progression detection via deep sequence learning of successive radiographic scans. *Int. J. Environ. Res. Public Health* **19**, 480. https://doi.org/10.3390/ijerph19010480 (2022).
12. Shen, Z. *et al.* Platelet transcriptome identifies progressive markers and potential therapeutic targets in chronic myeloproliferative neoplasms. *Cell Rep. Med.* **2**, 100425. https://doi.org/10.1016/j.xcrm.2021.100425 (2021).
13. Preuss, K. *et al.* Using quantitative imaging for personalized medicine in pancreatic cancer: A review of radiomics and deep learning applications. *Cancers* **14**, 1654. https://doi.org/10.3390/cancers14071654 (2022).
14. Grant, T. J., Hua, K. & Singh, A. Molecular pathogenesis of pancreatic cancer. *Prog. Mol. Biol. Transl. Sci.* **144**, 241–275. https://doi.org/10.1016/bs.pmbts.2016.09.008 (2016).
15. Makohon-Moore, A. & Iacobuzio-Donahue, C. A. Pancreatic cancer biology and genetics from an evolutionary perspective. *Nat. Rev. Cancer* **16**, 553–565. https://doi.org/10.1038/nrc.2016.66 (2016).
16. Moore, A. R., Rosenberg, S. C., McCormick, F. & Malek, S. RAS-targeted therapies: Is the undruggable drugged? *Nat. Rev. Drug Discov.* **19**, 533–552. https://doi.org/10.1038/s41573-020-0068-6 (2020).
17. Bannoura, S. F., Khan, H. Y. & Azmi, A. S. KRAS G12D targeted therapies for pancreatic cancer: Has the fortress been conquered? *Front. Oncol.* **12**, 1013902. https://doi.org/10.3389/fonc.2022.1013902 (2022).
18. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52. https://doi.org/10.1038/nature16965 (2016).
19. Amintas, S. *et al.* KRAS gene mutation quantification in the resection or venous margins of pancreatic ductal adenocarcinoma is not predictive of disease recurrence. *Sci. Rep.* **12**, 2976. https://doi.org/10.1038/s41598-022-07004-x (2022).
20. Bryant, K. L., Mancias, J. D., Kimmelman, A. C. & Der, C. J. KRAS: Feeding pancreatic cancer proliferation. *Trends Biochem. Sci.* **39**, 91–100. https://doi.org/10.1016/j.tibs.2013.12.004 (2014).
21. Tubbs, A. & Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**, 644–656. https://doi.org/10.1016/j.cell.2017.01.002 (2017).
22. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339. https://doi.org/10.1038/nature12634 (2013).
23. Husmann, D. & Gozani, O. Histone lysine methyltransferases in biology and disease. *Nat. Struct. Mol. Biol.* **26**, 880–889. https://doi.org/10.1038/s41594-019-0298-7 (2019).
24. Zhang, X. *et al.* Characterization of the genomic landscape in large-scale Chinese patients with pancreatic cancer. *EBioMedicine* **77**, 103897. https://doi.org/10.1016/j.ebiom.2022.103897 (2022).
25. Zhao, R., Choi, B. Y., Lee, M. H., Bode, A. M. & Dong, Z. Implications of genetic and epigenetic alterations of CDKN2A (p16(INK4a)) in cancer. *EBioMedicine* **8**, 30–39. https://doi.org/10.1016/j.ebiom.2016.04.017 (2016).
26. Serra, S. & Chetty, R. p16. *J. Clin. Pathol.* **71**, 853–858. https://doi.org/10.1136/jclinpath-2018-205216 (2018).
27. Tokunaga, A. *et al.* Clinical significance of epidermal growth factor (EGF), EGF receptor, and c-erbB-2 in human gastric cancer. *Cancer* **75**, 1418–1425 (1995).
28. da Cunha Santos, G., Shepherd, F. A. & Tsao, M. S. EGFR mutations and lung cancer. *Annu. Rev. Pathol.* **6**, 49–69. https://doi.org/10.1146/annurev-pathol-011110-130206 (2011).
29. Yan, J., Li, P., Gao, R., Li, Y. & Chen, L. Identifying critical states of complex diseases by single-sample Jensen–Shannon divergence. *Front. Oncol.* **11**, 684781. https://doi.org/10.3389/fonc.2021.684781 (2021).
30. Hoeck, J. D. *et al.* Fbw7 controls neural stem cell differentiation and progenitor apoptosis via Notch and c-Jun. *Nat. Neurosci.* **13**, 1365–1372. https://doi.org/10.1038/nn.2644 (2010).
31. Cheng, Y. L. *et al.* Evidence that neuronal Notch-1 promotes JNK/c-Jun activation and cell death following ischemic stress. *Brain Res.* **1586**, 193–202. https://doi.org/10.1016/j.brainres.2014.08.054 (2014).
32. Alfaro, D. *et al.* EphB2 and EphB3 play an important role in the lymphoid seeding of murine adult thymus. *J. Leukoc. Biol.* **98**, 883–896. https://doi.org/10.1189/jlb.1HI1114-568R (2015).
33. Niederacher, D., Yan, H. Y., An, H. X., Bender, H. G. & Beckmann, M. W. CDKN2A gene inactivation in epithelial sporadic ovarian cancer. *Br. J. Cancer* **80**, 1920–1926. https://doi.org/10.1038/sj.bjc.6690621 (1999).
34. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501. https://doi.org/10.1038/nature14169 (2015).
35. Beganovic, S. Clinical significance of the KRAS mutation. *Bosn J. Basic Med. Sci.* **9**(Suppl 1), S17–S20. https://doi.org/10.17305/bjbms.2009.2749 (2009).
36. Zhu, C. *et al.* Targeting KRAS mutant cancers: From druggable therapy to drug resistance. *Mol. Cancer* **21**, 159. https://doi.org/10.1186/s12943-022-01629-2 (2022).
37. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337. https://doi.org/10.1038/nature11252 (2012).
38. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616. https://doi.org/10.1038/ng.3564 (2016).
39. Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: Origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008. https://doi.org/10.1101/cshperspect.a001008 (2010).
40. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. https://doi.org/10.1038/75556 (2000).

41. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
42. Makhoba, X. H., Viegas, C. Jr., Mosa, R. A., Viegas, F. P. D. & Pooe, O. J. Potential impact of the multi-target drug approach in the treatment of some complex diseases. *Drug Des. Dev. Ther.* **14**, 3235–3249. https://doi.org/10.2147/DDDT.S257494 (2020).
43. Petrelli, A. & Giordano, S. From single- to multi-target drugs in cancer therapy: When aspecificity becomes an advantage. *Curr. Med. Chem.* **15**, 422–432. https://doi.org/10.2174/092986708783503212 (2008).
44. Just, N. Improving tumour heterogeneity MRI assessment with histograms. *Br. J. Cancer* **111**, 2205–2213. https://doi.org/10.1038/bjc.2014.512 (2014).
45. Eloyan, A., Yue, M. S. & Khachatryan, D. Tumor heterogeneity estimation for radiomics in cancer. *Stat. Med.* **39**, 4704–4723. https://doi.org/10.1002/sim.8749 (2020).
46. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006. https://doi.org/10.1038/ncomms5006 (2014).
47. Alic, L., Niessen, W. J. & Veenland, J. F. Quantification of heterogeneity as a biomarker in tumor imaging: A systematic review. *PLoS ONE* **9**, e110300. https://doi.org/10.1371/journal.pone.0110300 (2014).
48. Deng, M., Bragelmann, J., Kryukov, I., Saraiva-Agostinho, N. & Perner, S. FirebrowseR: An R client to the Broad Institute's Firehose Pipeline. *Database* **2017**, 160. https://doi.org/10.1093/database/baw160 (2017).
49. Klein, A. P. *et al.* Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat. Commun.* **9**, 556. https://doi.org/10.1038/s41467-018-02942-5 (2018).
50. McGuigan, A. *et al.* Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. *World J. Gastroenterol.* **24**, 4846–4861. https://doi.org/10.3748/wjg.v24.i43.4846 (2018).
51. Lo Gullo, R., Daimiel, I., Morris, E. A. & Pinker, K. Combining molecular and imaging metrics in cancer: Radiogenomics. *Insights Imaging* **11**, 1. https://doi.org/10.1186/s13244-019-0795-6 (2020).
52. Hinzpeter, R. *et al.* CT radiomics and whole genome sequencing in patients with pancreatic ductal adenocarcinoma: Predictive radiogenomics modeling. *Cancers* **14**, 224. https://doi.org/10.3390/cancers14246224 (2022).
53. Hammond, R. K. *et al.* Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci. *Elife* **10**, 206. https://doi.org/10.7554/eLife.62206 (2021).
54. Feng, S. *et al.* Methods for association analysis and meta-analysis of rare variants in families. *Genet. Epidemiol.* **39**, 227–238. https://doi.org/10.1002/gepi.21892 (2015).
55. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178. https://doi.org/10.1038/ng.3398 (2015).
56. Koboldt, D. C. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576. https://doi.org/10.1101/gr.129684.111 (2012).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 (2009).
58. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339 (2017).
59. Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338. https://doi.org/10.1148/radiol.2020191145 (2020).
60. Parr, E. *et al.* Radiomics-based outcome prediction for pancreatic cancer following stereotactic body radiotherapy. *Cancers* **12**, 051. https://doi.org/10.3390/cancers12041051 (2020).
61. Guo, M. H., Plummer, L., Chan, Y. M., Hirschhorn, J. N. & Lippincott, M. F. Burden testing of rare variants identified through exome sequencing via publicly available control data. *Am. J. Hum. Genet.* **103**, 522–534. https://doi.org/10.1016/j.ajhg.2018.08.016 (2018).
62. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93. https://doi.org/10.1016/j.ajhg.2011.05.029 (2011).
63. Peloso, G. M. & Lunetta, K. L. Choice of population structure informative principal components for adjustment in a case-control study. *BMC Genet.* **12**, 64. https://doi.org/10.1186/1471-2156-12-64 (2011).
64. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909. https://doi.org/10.1038/ng1847 (2006).

Author contributions
D.Z., C.Z., and M.H. conceived the project. M.B. and A.K. performed expert contouring. P.G., D.Z., Y.S., Q.D., J.W., S.B., and K.P. curated the data. C.Z., D.Z., Q.Z., X.L., H.Y., and H.D. performed data analysis. C.Z., D.Z., and M.H. drafted the manuscript. M.H. provided expert knowledge and funding support. All authors reviewed and approved the manuscript.

Funding

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-62741-5.

Correspondence and requests for materials should be addressed to D.Z., M.A.H. or C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.