

1 **A central research portal for mining pancreatic clinical and molecular datasets and**
2 **accessing biobanked samples.**

3

4 J. Oscanoa^{1,a}, H Ross-Adams^{1,a}, Abu Z M Dayem Ullah^{1,2}, TS Kolvekar^{1,2}, L Sivapalan¹,
5 E Gadaleta¹, GJ Thorn¹, M Abdollahyan¹, A Imrali², A Saad^{2,4}, R Roberts², C Hughes²,
6 PCRFTB, HM Kocher^{2,3,4}, C Chelala^{1,2,*}

7

8 ¹Centre for Cancer Biomarkers and Biotherapeutics, Barts Cancer Institute, Queen Mary
9 University London, UK, EC1M 6BQ

10 ²Pancreatic Cancer Research Fund Tissue Bank, Centre for Tumour Biology, Barts Cancer
11 Institute, Queen Mary University London, UK, EC1M 6BQ

12 ³Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University London, UK,
13 EC1M 6BQ

14 ⁴Barts and the London HPB Centre, The Royal London Hospital, Barts Health NHS Trust,
15 Whitechapel, London E1 1BB

16

17 *Corresponding Author: Prof C Chelala, c.chelala@qmul.ac.uk

18 ^aThese authors contributed equally.

19

20 **PCRFTB:** Mo Abu Hilal (Southampton), Bilal Al Sarireh (Swansea), Somaiah Aroori
21 (Plymouth), Ali Arshad (Southampton), Satyajit Bhattacharya (The London Clinic, London),
22 Brian Davidson (Royal Free, London), John Isherwood (Leicester), Deep Malde (Leicester),
23 Stuart Robinson (Newcastle), Zahir Soonawalla (Oxford).

24

25

26

27

28

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

29

30 **Abstract**

31 The Pancreatic Expression Database (PED) is a powerful resource dedicated to the mining
32 and analysis of pancreatic -omics datasets. Here, we demonstrate the biological
33 interpretations that are possible because of vital updates that have transformed PED into a
34 dynamic analytics hub accommodating an extensive range of publicly available datasets.
35 PED now hosts clinical and molecular datasets from four primary sources (Cancer Genome
36 Atlas, International Cancer Genome Consortium, Cancer Cell Line Encyclopaedia and
37 Genomics Evidence Neoplasia Information Exchange) that together form the foundation of
38 omics profiling of pancreatic malignancies and related lesions (n=7,760 specimens). Several
39 user-friendly analytical tools to explore and integrate the molecular data derived from these
40 primary specimens and cell lines are now available. Crucially, PED is integrated as the data
41 access point for Pancreatic Cancer Research Fund Tissue Bank – the only national
42 pancreatic cancer biobank in the UK. This will pioneer a new era of biobanking to promote
43 collaborative studies and effective sharing of multi-modal molecular, histopathology and
44 imaging data from biobank samples (>60,000 specimens from >3,400 cases and controls;
45 2,037 H&E images from 349 donors) and accelerate validation of *in silico* findings in patient-
46 derived material. These updates place PED at the analytical forefront of pancreatic
47 biomarker-based research, providing the user community with a distinct resource to facilitate
48 hypothesis-testing on public data, validate novel research findings, and access curated,
49 high-quality patient tissues for translational research. To demonstrate the practical utility of
50 PED, we investigate somatic variants associated with established transcriptomic subtypes
51 and disease prognosis: several patient-specific variants are clinically actionable and may be
52 leveraged for precision medicine.

53

54 **Introduction**

55 Pancreatic ductal adenocarcinoma (PDAC) is predicted to become the second leading cause
56 of cancer-related mortality worldwide before 2040¹. It has dismal 5-year survival rates of 3-
57 15%^{1,2}, largely due to late disease detection and few effective treatment options. Alarminglly,
58 the incidence of early-onset pancreatic cancer is also increasing, in contrast to most other
59 solid tumours³. Better tools for patient stratification and treatment response are thus

60 essential to improve survival outcomes. However, the pancreatic cancer research
61 community is relatively small – investigators tend to develop bespoke collections of samples
62 that may be unusable beyond the breadth and scope of their ethical approval and storage
63 conditions. Sample collection protocols also vary widely, further limiting the translation of
64 derived results to clinical benefits.

65
66 Most existing biomarkers for monitoring treatment or assessing prognosis are not based on
67 the molecular attributes of PDAC tumours and have shown limited sensitivity and/or
68 specificity in prospective settings⁴. Numerous studies into the genomic and transcriptomic
69 determinants of tumour development and progression are available, but their findings are
70 dispersed across multiple resources, and can be difficult to access and translate into
71 meaningful survival or treatment benefits for patients by those without computational
72 expertise. This highlights the pressing need for simplified, integrated data mining and
73 analysis tools to improve accessibility to clinical and molecular information from disparate
74 sources and enable laboratory and clinical researchers to easily and effectively cross-query
75 large multi-omics datasets, to fuel new discoveries in pancreatic diseases.

76
77 Here, we present the latest release of the Pancreatic Expression Database ([PED](#)), an
78 intuitive, online portal that links numerous multi-modal datasets to an active biobank where
79 users can both validate their findings, and/or apply for samples to confirm *in silico* findings.
80 To remain abreast of the evolving nature of integrative multi-omics workflows and in
81 response to user feedback, we have made vital new updates to PED's Analytics Hub,
82 broadened the range of datasets available and placed this essential resource at the centre
83 of the established framework for PED (Table 1). We have also integrated PED as the major
84 bioinformatics platform of the UK's national Pancreatic Cancer Research Fund Tissue Bank
85 (PCRFTB), to facilitate investigative biomarker-based research and data sharing between
86 clinicians and scientists. This is powered by a customised version of [SNPNexus](#), a versatile
87 platform for the functional annotation of known and novel sequence variation⁵, designed to
88 reduce the analytical burden associated with large-scale genomic datasets and facilitate the
89 straightforward identification of biologically and clinically relevant genetic variants in patients.

90

91 The PCRFTB is the world's first national pancreas tissue bank and has been collecting
92 blood, urine, saliva and solid tissue samples from patients recruited at nine participating
93 centres across the UK NHS since 2015, making it a valuable resource for translational
94 research. Tissues are available from patients with pancreatic and hepatobiliary diseases,
95 including resectable and unresectable cancer. Blood, urine and saliva samples from
96 patients, their first-degree relatives and other healthy volunteers are also available, as well
97 as cancer organoids and cancer-associated fibroblasts. (Figure 1). These are accompanied
98 by extensive, verified clinical, histological and imaging data that is continually updated –
99 median 300 data points per visit, with some donors providing longitudinal samples at multiple
100 visits throughout their treatment journey. Best practice and ongoing technical research
101 ensures available biological materials are of high quality, to support reliable and reproducible
102 results^{6,7} In addition to samples, digitised radiological images are available for 171 patients
103 with malignant, pre-malignant and benign pancreatic diagnoses, and 2,037 H&E images
104 from 349 donors are also currently available.

105

106 The PCRFTB is further supported by a Data Return policy to maximise the use of available
107 samples by linking each patient/donor with an enriched 'digital fingerprint' encompassing
108 molecular, transcriptomic, proteomic, imaging and longitudinal clinical data. All donors
109 provide written, informed consent, and all samples are collected, processed and stored at
110 each of the participating centres (Barts, Leicester, Swansea, Oxford, Royal Free (London),
111 Southampton, Newcastle, Plymouth, The London Clinic) under one Research Ethics
112 Committee reference (13/SC/0593, renewed 18/SC/0629, renewed 23/SC/0282) and using
113 standardised protocols, quality assurance and quality control policies ensuring consistency
114 across the collection.

115

116 Via PED, researchers can directly query and apply to PCRFTB for samples, specimens
117 and/or imaging data that match user-determined criteria. A link to the PCRFTB Tissue
118 Request System allows users to submit Expressions of Interest directly to the Tissue Bank
119 using an online application form.

120

121 By providing a dynamic hub for the analysis of publicly available pancreatic datasets and
122 ongoing research data generated from biobanked samples, PED allows researchers to
123 access a broad range of pancreas-specific molecular information freely and quickly. The
124 flexibility of this hub allows molecular alterations with biological and clinical relevance to be
125 identified and prioritised for downstream validation.

126

127 **The Analytics Hub**

128 We have updated the web-based Analytics Hub to include a broad range of pancreas-
129 related, publicly available -omics datasets, together with expanded analytical features and
130 visualisation options (Table 1), based on feedback from our diverse international user
131 community.

132

133 ***Publicly available data sources***

134 Building on the 2018 release⁸, the newly formed analytics hub now hosts publicly available
135 clinical and molecular datasets from four core sources: The Cancer Genome Atlas (TCGA)⁹
136 whole exome sequencing (WES) data, updated to filter specifically for adenocarcinoma
137 samples vs others; the Cancer Cell Line Encyclopaedia (CCLE)¹⁰, updated to include
138 somatic mutation and mRNA expression data for the complete set of 60 pancreas cell lines
139 from primary and metastatic tumours; the Genomics Evidence Neoplasia Information
140 Exchange (GENIE)¹¹ v13.0, updated to include simple somatic mutations and clinical data
141 from the *complete* set of 6,633 patients with pancreatic cancer of any type; and the now
142 archived complete International Cancer Genome Consortium (ICGC)^{12,13} dataset, including
143 whole genome and RNA sequencing data from both adenocarcinoma (PACA-AU; PACA-CA)
144 and neuro/endocrine tissues (PAEN-IT; PAEN-AU). These sources host data generated by
145 both national and international consortia efforts to sequence and analyse cancer genomes
146 and biology, including pancreatic malignancies. Analysed and quality-controlled data files
147 were downloaded from the respective sources and used without further processing. PED
148 2024 uses the most recent data releases, including linked clinical data when available.

149

150 **Advanced filtering of public datasets**

151 Public datasets may be queried according to the clinical characteristics of each study cohort
152 (Figure 2A). Filtering options have been selected based on relevance to disease development
153 and pathogenesis, and the depth of annotation provided in available clinical data for each
154 cohort. Implemented filters are accompanied by dedicated visualizations of clinical summaries
155 for each respective study cohort (Figure 2B). These include filters based on patient-related
156 factors (cancer type, sex, age, diabetes, family history, ethnicity, survival; Figure 2C), and
157 tumour characteristics (stage, grade, *KRAS* somatic mutation status) (Figure 2D), which allow
158 for trends in data to be clearly observed: e.g. survival beyond 3 years is very low for PDAC
159 compared to neuroendocrine tumours (Figure 2B). Crucially, it is possible to filter each dataset
160 by diagnosis, allowing researchers to focus on different pancreatic lesions individually (e.g.
161 IPMN, ductal adenocarcinoma, neuroendocrine, adenosquamous, mucinous) that have
162 different molecular alterations and clinical prospects, since using unstratified sample sets has
163 been shown to yield unreliable results.^{14,15}

164

165 **Characterising the genomic characteristics of established PDAC molecular subtypes**

166 PED facilitates the stratification and analysis of TCGA and ICGC cohorts according to their
167 molecular subtype classifications, as determined by hallmark transcriptomic^{16–18} and genomic
168 (ICGC only)¹⁹ studies, and recent histopathology-based artificial intelligence (AI) predictions
169 in matched TCGA samples²⁰(Figure 2E). We demonstrate the implementation of this clinically
170 useful feature below.

171

172 Collisson et al. (2011) originally identified 3 subclasses of PDAC tumours with different
173 clinical outcomes and treatment responses, termed *quasi-mesenchymal* (QM) (worst
174 prognosis), *classical* (best prognosis) and *exocrine-like*, using hybridisation array-based
175 mRNA expression data from primary untreated resected PDAC¹⁶. Next, Moffitt et al (2015)
176 analysed bulk tumour tissues from treatment-naïve primary resected PDAC tumours using
177 virtual microdissection to exclude transcripts native to the normal pancreas and the tumour
178 microenvironment, and reported 2 distinct tumour subtypes (*basal* and *classical*) as well as 2

179 classifications based on peritumoural stromal tissues (*activated* and *normal*)¹⁷. *Basal*
180 subtype tumours were associated with a poorer overall patient survival compared to *classical*
181 tumours, which overlapped significantly with the Collisson *classical* subtype. Subsequently,
182 Bailey et al. performed RNA-sequencing of bulk primary untreated resected tumour tissues
183 from 328 PDAC tumours and resolved four stable tumour classes (*squamous*, *pancreatic*
184 *progenitor*, *immunogenic* and *aberrantly differentiated endocrine exocrine* (ADEX)), each
185 governed through the differential expression of transcription factors and their targets
186 involved in lineage specification during pancreatic development¹⁸. *Squamous* subtype
187 tumours overlapped with previous *basal* (Moffitt) and *QM* (Collisson) classifications and were
188 associated with the poorest overall prognosis in patients. Recent investigations of these
189 proposed classifications have corroborated the presence of 2 overarching transcriptomic
190 subtypes of PDAC tumours comprising *basal-like/squamous* and *classical/progenitor* that
191 have shown relevance for defining survival outcomes in patients, with remaining subtypes
192 (*exocrine-like*, *ADEX*) shown to have confounding associations with poor tumour
193 cellularity^{21,22}. Most recently, Saillard et al (2023) used an artificial intelligence model
194 trained and validated on 5 independent surgical and biopsy cohorts with RNAseq and
195 histology data (n=598), including n=126 TCGA samples to further refine these tumour
196 subtypes²⁰. This approach recapitulated the known *basal/classical*¹⁷ tumour subtypes at the
197 whole H&E slide level but detected variable proportions of basal cells in samples previously
198 categorised as classical subtype when slides were analysed at 112 μm tile size level. This
199 changed survival outcomes in 39% of cases classed as *classical* subtype by bulk RNAseq
200 analysis, with the impact apparently proportional to the percentage basal cell content.²⁰

201

202 At the DNA level, whole genome sequencing (WGS) and copy number variant (CNV)
203 analysis performed on 100 treatment-naïve, macro-dissected PDAC tissues (ICGC Australia)
204 identified four disease subtypes with distinct patterns of structural variation (*scattered*, *locally*
205 *rearranged*, *stable*, *unstable*) and clinical utility, with *unstable* subtype characterised by a
206 very high degree of genomic instability throughout the genome and encompassing defects in
207 DNA damage repair (DDR) pathways that confer susceptibility to PARP inhibition (PARPi)
208 and/or platinum chemotherapies¹⁹.

209

210 **Application of subtype-specific filtering criteria in study cohorts**

211 Subtype-specific characteristics can be explored using the TCGA (n=185) and newly added
212 ICGC Australia (PACA-AU, n=461) and Canada (PACA-CA, n=317) pancreatic cancer
213 cohorts. Molecular subtype classifications according to Collisson¹⁶, Moffitt¹⁷, Bailey¹⁸ and/or
214 Saillard²⁰ are available for n=134 of the 156 confirmed PDAC patients included in the TCGA
215 cohort²¹. Alternatively, ICGC-AU cohorts can be analysed according to the subtype
216 classifications proposed by Bailey et al (2016)¹⁸ (n=95 patients total; 81 PDAC) and/or
217 Waddell et al (2015)(n=86 patients total; 85 PDAC).¹⁹ Given the prognostic relevance of
218 these transcriptomic subtypes, here we explore the associated genomic features of TCGA
219 PDAC tumours classed unanimously as either classical/progenitor (n=27) or QM/basal-
220 like/squamous (n=16) by all three subtyping systems (Collisson/Moffit/Bailey), as an
221 example of PED Analytics Hub.

222

223 **Subtype-specific somatic variations**

224 Comparisons between the genomic characteristics of each TCGA subtype showed different
225 gene sets mutated in *classical* and *basal* subtypes (Figure 3A, B). This was also true for
226 PACA-AU prognostic subtypes (progenitor vs squamous) (Supplementary Figure 1A, B), but
227 with little consensus between the two cohorts (Supplementary Figure 1C). To more reliably
228 identify subtype-specific genetic variations robust to inevitable inter-study variability (e.g.
229 tissue heterogeneity; WES vs WGS²³), we considered the *union* of the top 25 most
230 frequently mutated genes between similar prognostic groups for the two largest datasets
231 (TCGA+ICGC). This revealed a handful of common genes detected at >10% prevalence
232 (*KRAS*, *TP53*, *CDKN2A*, *MUC16*, *LRP1B*, *AFF2*, *FAT4*), but with most genes/variants being
233 subtype-specific (Figure 3C), consistent with (1) the early acquisition of these common
234 alterations during PDAC tumour development and (2) the molecular heterogeneity of PDAC
235 tumours²¹.

236

237

238 **Identifying treatment biomarkers**

239 *In silico* functional analysis of *all* patient-specific somatic variants identified, using the most
240 recent, freely available embedded Cancer Genome Interpreter (CGI)²⁴ analytic tool,
241 identified numerous biomarkers of response/resistance to existing clinical treatments and/or
242 pharmacological inhibitors (Supplementary Table S2) in various cancer contexts, including
243 multiple variants in *KRAS*, *TP53*, *CDKN2A* and *LRP1B*. Although *KRAS* is extensively
244 mutated across both prognostic groups, *KRAS* p.G12C driver variants were unique to the
245 best prognosis subtype tumours (Supplementary Table S2). This variant, rare in PDAC (<1%
246 patients)²⁵, has been shown to preferentially drive the RAF/RAL pathway, while the more
247 common *KRAS*^{G12D} mutation (~30% PDAC patients) favours the PI3K/AKT pathway^{25,26}.
248 Targeted *KRAS*^{G12C} inhibitor sotorasib has recently received FDA approval for treatment of
249 mNSCLC²⁷, and has been shown to be safe and effective in treatment of advanced mPDAC,
250 in a Phase I/II trial in n=38 patients²⁸. Additionally, certain variants in *CDKN2A* (L104V,
251 E120*, R58*, R80*) have been linked with treatment resistance to PD1 inhibitors, and
252 treatment response to CDK4/6 inhibitors in cutaneous melanoma (CM; Supplementary Table
253 S2), highlighting the usefulness of PED in identifying potentially clinically relevant patient-
254 and cancer-specific therapeutic targets.

255

256 Of the somatic variants identified, CGI oncogenic classifications (bioactivity) (Supplementary
257 Table S1), revealed several to be TIER 1 predicted driver variants, i.e. the gene activity is
258 confirmed relevant to cancer, with mutations identified effecting oncogenic transformation.
259 Only one was identified in the poorest outcome patient group – a splice donor variant in
260 central cell-cycle regulator gene *ATM*, and linked with response to cisplatin chemotherapy,
261 PARPi by olaparib, and PD1/PD-L1 inhibition in other solid tumours (Supplementary
262 Table S2). However, this variant was uncommon in the sample set (1/16). Conversely,
263 several TIER 1 oncogenic driver variants were identified in the best prognosis patient group,
264 with some associated with response/resistance to specific drugs in other solid tumours,
265 suggesting possible utility in pancreatic cancer: *ARID1A* (p.E1542*, p.Q1277*; responsive to
266 EZH2, PD1, ATR & PARP inhibitors), *RNF43* (W159*, responsive to porcupine inhibitor) and

267 *TP53* (S166*, Y205S, D259V, M246R, S241F, L194H, N131I; resistance to CDK4/6 inhibitor
268 abemaciclib, cisplatin, MDM2 inhibitor; responsive to ATR inhibitor AZD6738, doxorubicin,
269 decitabine, gemcitabine, mitomycin C).

270

271 Results of the above detailed output are also summarised in alluvial plots, showing clinically-
272 actionable targets present in selectable proportions of the filtered dataset (5%-25%) and
273 their responsiveness/resistance to available drugs (Supplementary Figure 2A, B), as well as
274 a visual summary of the number of druggable gene categories represented in the selected
275 dataset (Supplementary Figure 2C, D), that shows different clinically actionable genome
276 targets between the two prognostic groups. While the most commonly mutated genes are
277 common between prognostic subtypes ($\geq 25\%$ of patients from both groups contain similar
278 *KRAS* and *TP53* variants), distinct biomarker/drug combinations are apparent when less
279 common variations are considered: only one other gene in the poor prognosis group was
280 identified as harbouring variants linked to (among others) responsiveness to small molecule
281 AURKA-VEGF inhibitor ilorasertib (CDKN2A R58*), whereas the better prognosis subgroup
282 is associated with 6 additional gene/biomarker candidates (*RNF43*, *PIK3CA*, *ERCC4*,
283 *CTNNB1*, *CDKN2A*, *ARID1A*), that have shown promise in treating other solid tumours.
284 Additionally, by exploring the available CCLE database of 60 pancreatic cell lines, *in vitro*
285 models with/without *KRAS* and/or *TP53* variants may be identified to support downstream
286 functional studies.

287

288 These results demonstrate the value of PED in contextualising individual patient genetic
289 profiles in suggesting possible treatment options, or refining research areas to pursue for
290 more effective, stratified approaches in pancreatic cancer.

291

292 ***Characterising patterns of gene expression in best- and worst-outcome tumours***

293 Inspecting the top 250 differentially expressed genes in both TCGA and ICGC filtered patient
294 subgroups confirms scant overlap between genes differentially expressed in best and worst-
295 outcome PDAC tumours. However, *classical/progenitor* tumours appear to have higher
296 *TP53* expression compared to *QM/basal/squamous* tumours, consistent with its role as

297 tumour suppressor (Figure 4A) and mirrored in ICGC *progenitor* and *squamous* patient
298 subgroups¹⁸ (Supplementary Figure 3). Subtype-specific differences in expression were also
299 observed for *MUC16* (encoding CA125 membrane glycoprotein) which was over-expressed
300 in *QM/basal/squamous* subtype tumours, compared to *classical/progenitor* cases (Figure
301 4B). Considering all n=402 confirmed PDAC in PACA-AU (the largest single transcriptomic
302 dataset), PED shows that *MUC16* over-expression is associated with significantly reduced
303 patient survival (logrank p=0.011; HR=2.23) (Figure 4C), where no association was found for
304 neuroendocrine tumours (Figure 4D). Elevated CA125 has also recently been shown to be
305 an independent prognostic marker of significantly shorter survival in n=207 resectable PDAC
306 patients, both before and after treatment²⁹. Functionally, CA125 over-expression has been
307 shown to promote tumourigenesis *in vitro* and *in vivo*^{30,31}, and monoclonal antibody mAb
308 AR9.6 has recently shown potential as a specific and effective inhibitor of CA125 and its
309 oncogenic effects in pancreatic and ovarian cancers^{31,32}.

310

311 ***Identifying clinically actionable genomic alterations in KRAS wild-type PDAC tumours***

312 Several studies have explored the genetic landscape of *KRAS* wild-type tumours, delineating
313 several alterations that occur frequently in the absence of mutant *KRAS*^{21,33–35}. In addition,
314 a significant enrichment for somatic aberrations that target the RAS-MAPK pathway, either
315 upstream or downstream of *KRAS*, has been observed in up to one-third of *KRAS* wild-type
316 tumours^{21,35}, where *BRAF* alterations were prevalent and mutually exclusive with *KRAS*
317 mutations³⁵. However, alterations within genes that are not typically associated with *RAS*
318 signalling have also been widely identified across *KRAS* wild-type PDAC tumours, and
319 require further investigation to determine functional relevance^{18,21}.

320

321 ***Alternative oncogenic drivers amongst KRAS wild-type PDAC tumours***

322 We used GENIE¹¹ as the largest available resource to identify n=756 *KRAS* wild-type PDAC
323 samples. To identify other likely molecular drivers in these tumours, predictions from CGI
324 were analysed to evaluate the distribution of altered genes and their associated pathways.
325 Mutations were detected across several genes previously reported to be altered in *KRAS*

326 wild-type PDAC cases, including *TP53* (mutated in >40% of the samples), *GNAS* and
327 *BRAF*³⁶ (Figure 5A). *In silico* biomarker predictions also showed that *ARID1A*, *BRAF*,
328 *CDKN2A*, *GNAS*, *PIK3CA* and *TP53* variants demonstrated therapeutic potential in
329 response to PARP, tyrosine kinase and VEGF inhibitors, immunotherapies and several
330 chemotherapies (Figure 5B).

331

332 The analysis of altered signalling pathways amongst mutated genes across *KRAS* wild-type
333 samples revealed frequent alterations in pathways associated with MAPK signalling, P53
334 signalling, neurotrophin, cell cycle, wnt and apoptosis signalling, consistent with previous
335 characterisations of core biological pathways involved in PDAC development and
336 progression^{18,19,21,30,36}. (Figure 5C). These findings highlight the utility of PED for the
337 prioritisation of functionally and biologically relevant variants amongst subgroups of PDAC
338 tumours, with important implications for the characterisation of distinct molecular pathologies
339 and the identification of novel therapeutic opportunities.

340

341 Using the PED Analytics Hub, we demonstrate how our integrated high-performance
342 visualisation and analysis tool can be used to investigate the link between genomic and
343 transcriptomic features and phenotypes of pancreatic cancer, providing an important step in
344 defining potential subtype-specific therapeutic vulnerabilities.

345

346 **The PCRFTB Data Return Module**

347 The vision of precision medicine has driven unprecedented interest into biomarker-based
348 studies (genomics, transcriptomics, proteomics) for pancreatic cancer, which are being
349 adopted across research and development from early discovery through to clinical research
350 and trials³⁷. Fundamental to biomarker research is access to quality biospecimens and
351 samples that have been well annotated with clinical and molecular data⁶. Whilst many
352 biobanks have invested heavily in the IT infrastructure of sample management, most
353 platforms are facing challenges in the effective sharing of returned data to drive investigative
354 research across the pancreatic research community. A major challenge is the rise of so-
355 called 'big data' from e.g. NGS and images that need to be integrated with large quantities of

356 primary/secondary care information and other real-world healthcare data. Traditional
357 biobanks are not usually set up to leverage these innovations, which have the potential to
358 improve patient outcomes and accelerate the development and delivery of new therapies.
359 PED is now the primary bioinformatics platform of the PCRFTB, providing a unique
360 integrated resource of biological materials and associated clinical, molecular and
361 radiological/imaging data.

362

363 In addition to providing a direct link to sample requests from the PCRFTB, the newly
364 incorporated data return module in PED hosts clinical and molecular data returned to the
365 bank from studies undertaken using PCRFTB specimens, where published findings have
366 been made available for researchers to review and analyse prior to submitting a tissue
367 request. Studies are categorised according to the type of -omics data generated for each
368 project (i.e. genomic, transcriptomic or proteomic), with alternative data types (e.g.
369 summaries of staining or imaging results generated from experimental investigations)
370 classified “other”, to simplify use. These classifications are presented in a summary table,
371 which also provides a description of each project and details the different sample types (i.e.
372 blood, tissues, cell lines) and cohort sizes used for each project. Users also have the option
373 view this information as clinical summary plots for each individual cohort, prior to exploring
374 available molecular data from each study.

375

376 Like other tissue repositories^{38,39}, PCRFTB has implemented a data return policy, where
377 anonymised data derived from banked samples is returned to the tissue bank on completion
378 of the study and made freely available to the research community, regardless of whether the
379 study is ultimately published⁶. As associated sample datasets develop, more in-depth
380 integrative analyses will be possible. Under *PCRFTB Data*, PED lists the details of any
381 returned data and associated sample characteristics available for analysis, while *PCRFTB*
382 *Research Projects* links directly to the relevant published report. So far, this has further
383 enriched the data available for banked tissue sample and/or patients, and currently includes
384 data on stromal⁴⁰, urinary miRNA⁴¹, metallomic⁴², and volatile organic compound⁴³
385 biomarkers, proteomics (ELISA)⁴⁴, circulating tumour cells (CTCs) and xenotransplantation

386 models^{45,46}, germline and somatic mutations⁴⁷, a phase I clinical trial⁴⁸, and risk⁴⁹ and
387 recurrence⁵⁰ predictions incorporating electronic health record data. To date, PCRFTB has
388 processed 38 EoI, supported 33 research projects and 19 peer-reviewed publications.
389 Ultimately, each study contributes to the development of a 'digital fingerprint' for each
390 patient, linking multi-modal data with longitudinal clinical information.

391

392 PED Analytics Hub is the web-based portal through which this enriched dataset can be
393 accessed and compared with large scale pancreatic -omics data, with the unique benefit of
394 also providing access to additional patient samples (*via* PCRFTB) for subsequent validation
395 of molecular alterations with clinical potential. Here, we demonstrate the added value of
396 tissue banking to precision cancer medicine, to translate research findings into prognostic
397 and therapeutic tools using well-annotated curated tissues and associated clinical data.

398

399 **Discussion**

400 A rapid expansion in high-throughput genomic and transcriptomic profiling of pancreatic
401 diseases necessitates the development of sophisticated yet user-friendly analytics hubs to
402 host the growing compendium of molecular and clinical datasets and enable integrated
403 mining and analysis of available results. The updated PED now supports a range of data
404 modalities to enable users across the diverse international pancreatic research community to
405 identify and investigate trends in molecular data across disparate cohorts of patients,
406 samples and cell lines easily and effectively. PED now provides an unprecedented
407 opportunity to characterise distinct variations in both tumour mutation landscapes and gene
408 expression profiles that are associated with prognostic molecular subtypes. Candidate
409 tumour drivers and biomarkers predictive of response to existing and novel clinical
410 treatments can be identified and visualized, allowing suitable targets for downstream
411 validation and pharmacological testing to be prioritised without the need for laborious data
412 retrieval or processing tasks. Furthermore, PED is the gateway to a national tissue bank
413 repository of >60,000 samples from >3,400 patients and a growing repository of digitised
414 radiological and H&E images, that researchers can access independently and apply for
415 donor samples that match with their research question. The selection of samples is

416 significantly improved by the availability of high-quality clinical data, curated and maintained
417 by PCRFTB.
418
419 These major updates to the PED infrastructure are further underpinned by its recent
420 adoption as the bioinformatics platform of the PCRFTB, pioneering a new generation in
421 biobanking to support effective data sharing and promote collaborative studies,
422 democratizing access to complex cancer genomics. By harmonising PCRFTB samples with
423 clinical and molecular information from datasets returned to the biobank, PED provides an
424 essential platform to support translational pancreatic research and fuel discoveries that can
425 manifest clinically meaningful benefits for patients. PCRFTB recently launched
426 internationally (<https://www.pcrf.org.uk/news/tissue-bank-launches-internationally/>),
427 improving opportunities for high-quality research into earlier diagnosis and treatment of
428 pancreatic cancer. As -omics driven research continues to drive efforts to advance the
429 characterisation of pancreatic diseases, PED's design supports the ongoing data analysis,
430 integration and visualisation needs of the growing research community.
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445

446 **Table 1. Summary of the 2024 updates to PED**

Features	2018 release ⁸	2024 release
The PED Analytics Hub		
Publicly available data sources (pancreas-specific)		
PubMed ^a	X	
TCGA ^b	X	X
ICGC		X
GENIE ^c	X	X
CCLE ^d	X	X
Analytical features		
Principal components analysis	X	X
Gene expression profiles	X	X
Correlation analyses	X	X
Gene networks	X	X
Survival analyses ^e	X	X
Variant identification	X	X
Somatic gene interactions	X	X
Reactome & oncogenic pathway analyses		X
Improved clinical annotations to visualize & query publicly available datasets		X
MAFtools genomic analyses and summary visualisations		X
Tumour mutational burden		X
Clinically actionable genes/proteins & associated drugs		X
Cohort comparison by clinical/molecular feature		X
Gene intersections between filtered datasets		X
The PCRFTB Data Module		
Data return module to host both -omics and experimental datasets		X
Integrated primary/secondary care clinical data		X
Apply for samples		X

447

448

449

450 ^a Less-used literature mining module
451 ^b Updated to include essential filters based on cancer subtype
452 ^c Expanded from 445 adenocarcinoma or neuroendocrine tumours to full set of 6,633
453 pancreatic cancers of all types, with additional clinical and molecular information.
454 ^d Somatic variant dataset; now expanded to include the full set of 60 primary and metastatic
455 tumour derived cell lines.
456 ^e Expanded to include analyses based on mutational status and mRNA level

457
458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495 References

- 496
- 497 1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7–30
498 (2020).
 - 499 2. Pereira, S. P. *et al.* Early detection of pancreatic cancer. *Lancet Gastroenterol Hepatol*
500 **5**, 698–710 (2020).
 - 501 3. Abboud, Y. *et al.* Increasing Pancreatic Cancer Incidence in Young Women in the.
502 *Gastroenterology* 1–12 (2023) doi:10.1053/j.gastro.2023.01.022.
 - 503 4. Brezgyte, G., Shah, V., Jach, D. & Crnogorac-jurcevic, T. Non-invasive biomarkers for
504 earlier detection of pancreatic cancer—a comprehensive review. *Cancers (Basel)* **13**,
505 1–25 (2021).
 - 506 5. Oscanoa, J. *et al.* SNPnexus: A web server for functional annotation of human
507 genome sequence variation (2020 update). *Nucleic Acids Res* **48**, W185–W192 (2020).
 - 508 6. Balarajah, V. *et al.* Pancreatic cancer tissue banks: where are we heading? *Future*
509 *Oncology* **12**, 2661–2663 (2016).
 - 510 7. Imrali, A. *et al.* Validation of a Novel, Flash-Freezing Method: Aluminum Platform.
511 *Current Protocols in Essential Laboratory Techniques* **21**, 1–17 (2020).
 - 512 8. Marzec, J. *et al.* The Pancreatic Expression Database: 2018 update. *Nucleic Acids Res*
513 **46**, D1107–D1110 (2018).
 - 514 9. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat*
515 *Genet* **45**, 1113–1120 (2013).
 - 516 10. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of
517 anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
 - 518 11. Sweeney, S. M. *et al.* AACR project genie: Powering precision medicine through an
519 international consortium. *Cancer Discov* **7**, 818–831 (2017).
 - 520 12. Zhang, J. *et al.* International Cancer Genome Consortium Data Portal—a one-stop shop
521 for cancer genomics data. *Database (Oxford)* **2011**, bar026 (2011).
 - 522 13. Zhang, J. *et al.* The International Cancer Genome Consortium Data Portal. *Nat*
523 *Biotechnol* **37**, 367–369 (2019).
 - 524 14. Peran, I., Madhavan, S., Byers, S. W. & McCoy, M. D. Curation of the pancreatic ductal
525 adenocarcinoma subset of the cancer genome atlas is essential for accurate
526 conclusions about survival-related molecular mechanisms. *Clinical Cancer Research*
527 **24**, 3813–3819 (2018).
 - 528 15. Nicolle, R. *et al.* Prognostic biomarkers in pancreatic cancer: Avoiding errata when
529 using the TCGA dataset. *Cancers (Basel)* **11**, 1–10 (2019).
 - 530 16. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their
531 differing responses to therapy. *Nat Med* **17**, 500–503 (2011).
 - 532 17. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-
533 specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **47**, 1168–1178
534 (2015).
 - 535 18. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer.
536 *Nature* **531**, 47–52 (2016).
 - 537 19. Waddell, N. N. N. *et al.* Whole genomes redefine the mutational landscape of
538 pancreatic cancer. *Nature* **518**, 495–501 (2015).
 - 539 20. Saillard, C. *et al.* Pacpaint: a histology-based deep learning model uncovers the
540 extensive intratumor molecular heterogeneity of pancreatic adenocarcinoma. *Nat*
541 *Commun* **14**, (2023).
 - 542 21. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal
543 Adenocarcinoma. *Cancer Cell* **32**, 185–203.e13 (2017).

- 544 22. Sinkala, M., Mulder, N. & Martin, D. Machine Learning and Network Analyses Reveal
545 Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. *Sci Rep* **10**,
546 1212 (2020).
- 547 23. Ellrott, K. *et al.* Scalable Open Science Approach for Mutation Calling of Tumor
548 Exomes Using Multiple Genomic Pipelines. *Cell Syst* **6**, 271–281.e7 (2018).
- 549 24. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical
550 relevance of tumor alterations. *Genome Med* **10**, 25 (2018).
- 551 25. Kwan, A. K., Piazza, G. A., Keeton, A. B. & Leite, C. A. The path to the clinic: a
552 comprehensive review on direct KRASG12C inhibitors. *Journal of Experimental and*
553 *Clinical Cancer Research* **41**, (2022).
- 554 26. Ihle, N. T. *et al.* Effect of KRAS oncogene substitutions on protein behavior:
555 Implications for signaling and clinical outcome. *J Natl Cancer Inst* **104**, 228–239
556 (2012).
- 557 27. Nakajima, E. C. *et al.* FDA Approval Summary: Sotorasib for KRAS G12C-Mutated
558 Metastatic NSCLC. *Clinical Cancer Research* **28**, 1482–1486 (2022).
- 559 28. Strickler, J. H. *et al.* Sotorasib in KRAS p.G12C–Mutated Advanced Pancreatic Cancer .
560 *New England Journal of Medicine* **388**, 33–43 (2023).
- 561 29. Napoli, N. *et al.* Ca 125 is an independent prognostic marker in resected pancreatic
562 cancer of the head of the pancreas. *Updates Surg* **75**, 1481–1496 (2023).
- 563 30. Qi, Z. H. *et al.* RIPK4/PEBP1 axis promotes pancreatic cancer cell migration and
564 invasion by activating RAF1/MEK/ERK signaling. *Int J Oncol* **52**, 1105–1116 (2018).
- 565 31. Thomas, D. *et al.* Isoforms of MUC16 activate oncogenic signaling through EGF
566 receptors to enhance the progression of pancreatic cancer. *Molecular Therapy* **29**,
567 1557–1571 (2021).
- 568 32. Sharma, S. K. *et al.* ImmunoPET of Ovarian and Pancreatic Cancer with AR9.6, a Novel
569 MUC16-Targeted Therapeutic Antibody. *Clinical Cancer Research* **28**, 948–959 (2022).
- 570 33. Heining, C. *et al.* NRG1 Fusions in KRAS Wild-Type Pancreatic Cancer. *Cancer Discov* **8**,
571 1087–1095 (2018).
- 572 34. Luchini, C. *et al.* KRAS wild-type pancreatic ductal adenocarcinoma: molecular
573 pathology and therapeutic opportunities. *Journal of Experimental & Clinical Cancer*
574 *Research* **39**, 227 (2020).
- 575 35. Singhi, A. D. *et al.* Real-Time Targeted Genome Profile Analysis of Pancreatic Ductal
576 Adenocarcinomas Identifies Genetic Alterations That Might Be Targeted With Existing
577 Drugs or Used as Biomarkers. *Gastroenterology* **156**, 2242–2253.e4 (2019).
- 578 36. Philip, P. A. *et al.* Molecular Characterization of KRAS Wild-type Tumors in Patients
579 with Pancreatic Adenocarcinoma. *Clinical Cancer Research* **28**, 2704–2714 (2022).
- 580 37. Herbst, B. & Zheng, L. Precision medicine in pancreatic cancer: treating every patient
581 as an exception. *Lancet Gastroenterol Hepatol* **4**, 805–810 (2019).
- 582 38. Gadaleta, E., Pirrò, S., Dayem Ullah, A. Z., Marzec, J. & Chelala, C. BCNTB
583 bioinformatics: The next evolutionary step in the bioinformatics of breast cancer
584 tissue banking. *Nucleic Acids Res* **46**, D1055–D1061 (2018).
- 585 39. Speirs, V. Quality Considerations When Using Tissue Samples for Biomarker Studies in
586 Cancer Research. *Biomark Insights* **16**, (2021).
- 587 40. Goulart, M. R. *et al.* Pentraxin 3 is a stromally-derived biomarker for detection of
588 pancreatic ductal adenocarcinoma. *NPJ Precis Oncol* **5**, (2021).
- 589 41. Debernardi, S. *et al.* Noninvasive urinary miRNA biomarkers for early detection of
590 pancreatic adenocarcinoma. *Am J Cancer Res* **5**, 3455–3466 (2015).

- 591 42. Schilling, K. *et al.* Urine metallomics signature as an indicator of pancreatic cancer.
592 *Metallomics* **12**, 752–757 (2020).
- 593 43. Daulton, E. *et al.* Volatile organic compounds (VOCs) for the non-invasive detection of
594 pancreatic cancer from urine. *Talanta* **221**, (2021).
- 595 44. Debernardi, S. *et al.* A combination of urinary biomarker panel and PancRISK score for
596 earlier detection of pancreatic cancer: A case-control study. *PLoS Med* **17**, 1–23
597 (2020).
- 598 45. Raj, D. *et al.* Switchable CAR-T cells mediate remission in metastatic pancreatic ductal
599 adenocarcinoma. *Gut* **68**, 1052–1064 (2019).
- 600 46. Raj, D. *et al.* CEACAM7 is an effective target for CAR T-cell therapy of pancreatic
601 ductal adenocarcinoma. *Clinical Cancer Research* **27**, 1538–1552 (2021).
- 602 47. Sivapalan, L. *et al.* Longitudinal profiling of circulating tumour DNA for tracking
603 tumour dynamics in pancreatic cancer. *BMC Cancer* **22**, 1–17 (2022).
- 604 48. Kocher, H. M. *et al.* Phase I clinical trial repurposing all-trans retinoic acid as a stromal
605 targeting agent for pancreatic cancer. *Nat Commun* **11**, 4841 (2020).
- 606 49. Zardab, M. *et al.* Differentiating Ductal Adenocarcinoma of the Pancreas from Benign
607 Conditions Using Routine Health Records: A Prospective Case-Control Study. *Cancers*
608 (*Basel*) **15**, (2023).
- 609 50. Ang, A., Michaelides, A., Chelala, C., Ullah, D. & Kocher, H. M. Prognostication for
610 recurrence patterns after curative resection for pancreatic ductal adenocarcinoma.
611 *Ann Hepatobiliary Pancreat Surg* **28**, 248–261 (2024).

612

613

614

615

616

617

618

619

620

621

622

623

624

625 **Funding**

626 The PCRFTB is funded by PCRF. Organoid generation is supported by Barts Charity. HMK
627 and CC acknowledge the support of NIHR Barts BRC. This work was supported by Barts
628 Charity (grant code MGU0504) and Barts NIHR BRC (grant code BTXH1A1R), part of the
629 Precision Medicine programme.

630

631 **Competing interests**

632 H.M.K. received a research grant for conducting this trial (Celgene: institutional) and
633 educational grant support for attending or organizing conferences (Celgene, Baxalta, Mylan,
634 Medtronic, Oncosil: institutional) which are unrelated to this work. No other authors have
635 interests to declare.

636

637 **Ethics approvals**

638 All tissue/sample donors provide written, informed consent, and all samples were collected,
639 processed and stored at each of the participating centres (Barts, Leicester, Swansea,
640 Oxford, Royal Free (London), Southampton, Newcastle, Plymouth, The London Clinic) under
641 one Research Ethics Committee reference (13/SC/0593, renewed 18/SC/0629, renewed
642 23/SC/0282).

643

644 **Author contributions**

645 Conceptualization: CC; Methodology: JO, AZMDU, TSK, EG, GJT, MA; Software: JO;
646 Validation: LS, HRA; Investigation: LS, HRA; Resources: PCRFTB, AI, AS, RR, CH; Data
647 Curation: MA, TSK, AZMDU, EG, JO, RR, CH; Writing – original draft: LS; Writing – Review
648 & Editing: HRA, HMK, CC; Supervision: CC; Funding acquisition: HMK, CC. All authors read
649 and approved the final manuscript.

650

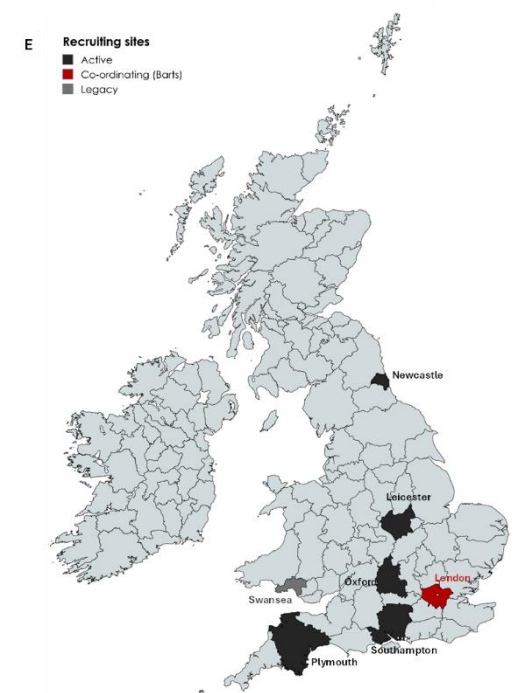
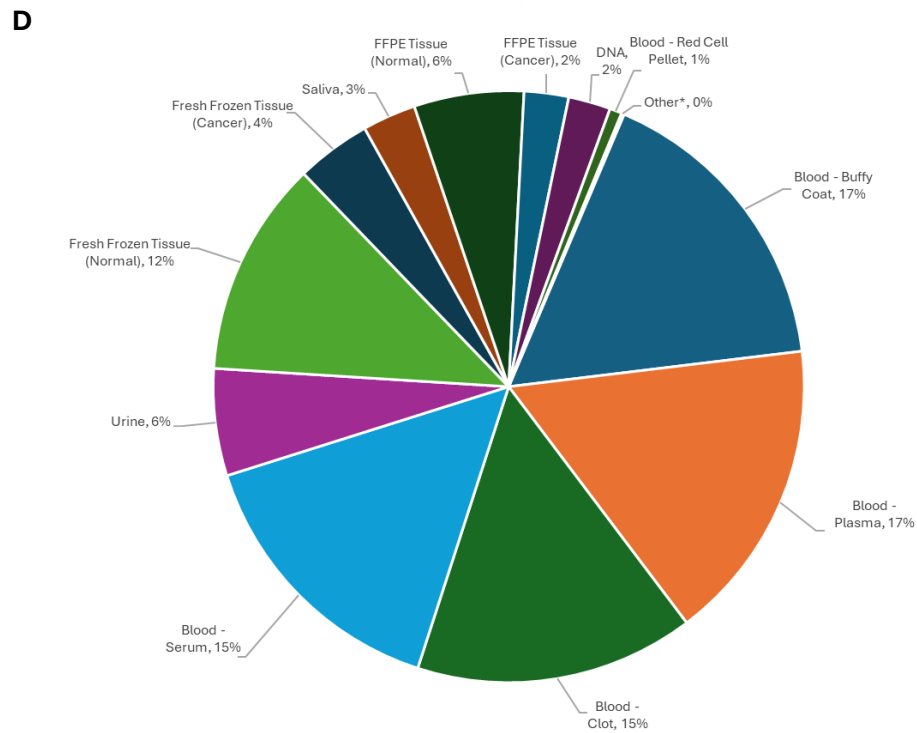
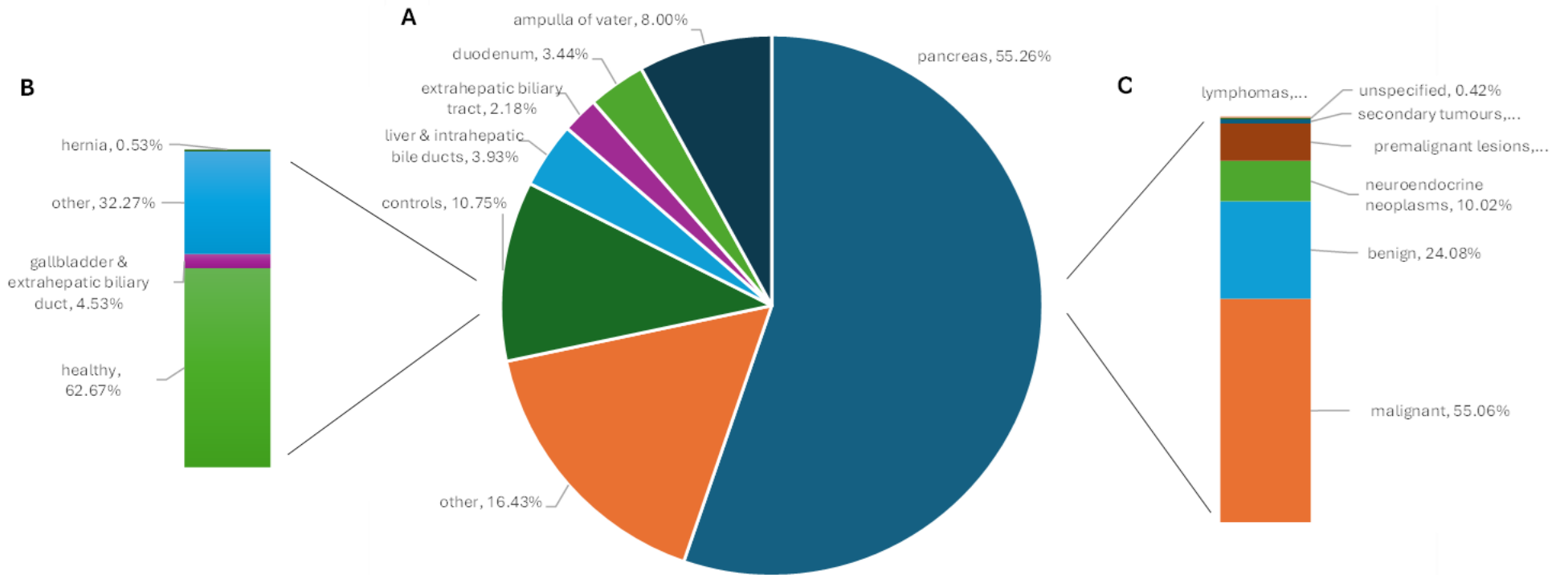


Figure 1. Summary of available tissue types.

(A) The proportion of >3 400 unique organ site tissues available for study in the UK national Pancreatic Cancer Research Fund Tissue Bank, with the breakdown of controls (B) and pancreas (C) highlighted. (D) Distribution of >60 000 [PCRFTB](#) specimens by type, across all patients. Details are updated weekly. Additionally, radiological imaging is available for 171 patients with malignant, pre-malignant and benign pancreatic diagnoses, and >2,000 H&E images from 349 donors. Samples can be applied for [here](#). (E) Geographical locations of PCRFTB patient recruitment sites.

Map created with [mapchart.net](#).

*=pancreatic juice, CTC, bile, organoids.

A PED

Pancreatic Expression Database
The bioinformatics portal of the PCRF TB
Powered by SNPnexus



Home Publicly Available Data PCRF TB Data Cohort Comparison

Apply for Samples

The Cancer Genome Atlas

Clinical filters

Age (30-90):

Sex: Female Male

Race: White Black or African American Asian

Diagnosis: Ductal adenocarcinoma

Tumor stage: I II III IV

Survival: Alive Deceased

Survival Period: <1 year 1-5 years >5 years

History of Diabetes: Yes No

Family History of Cancer: Yes No

Molecular filters

Missense mutation - KRAS: G12 G13 Q61 Absent

Missense mutation - TP53: Present Absent

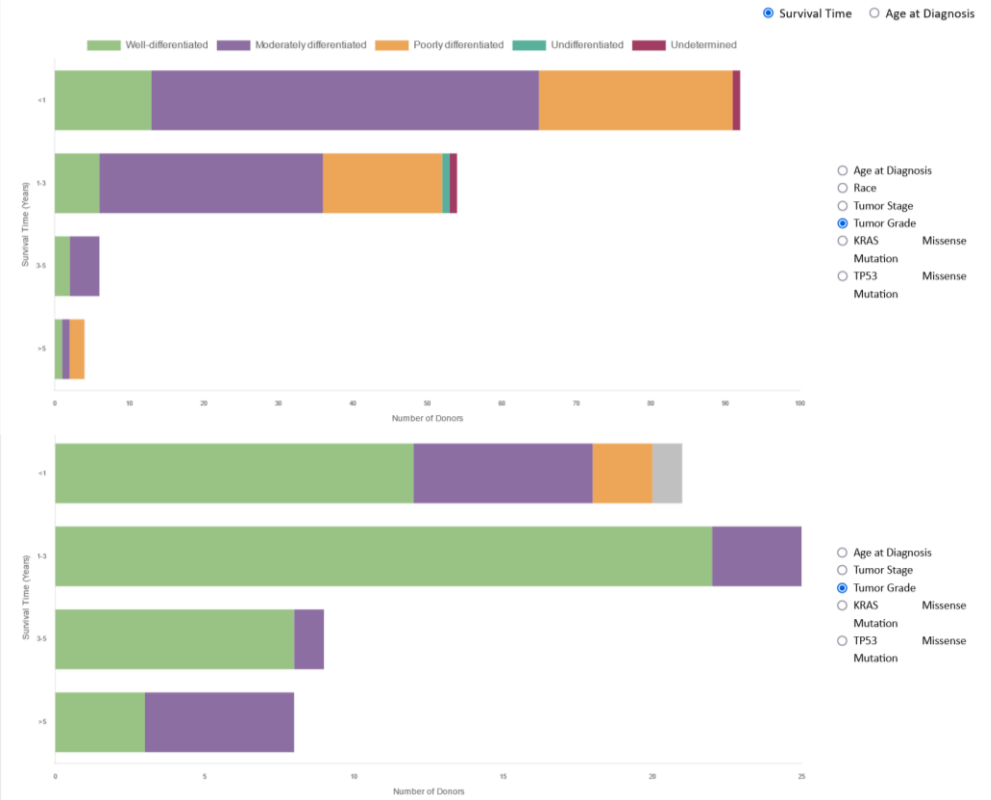
Subtype - Moffitt: Classical Basal

Subtype - Bailey: Progenitor Squamous Immunogenic ADEX

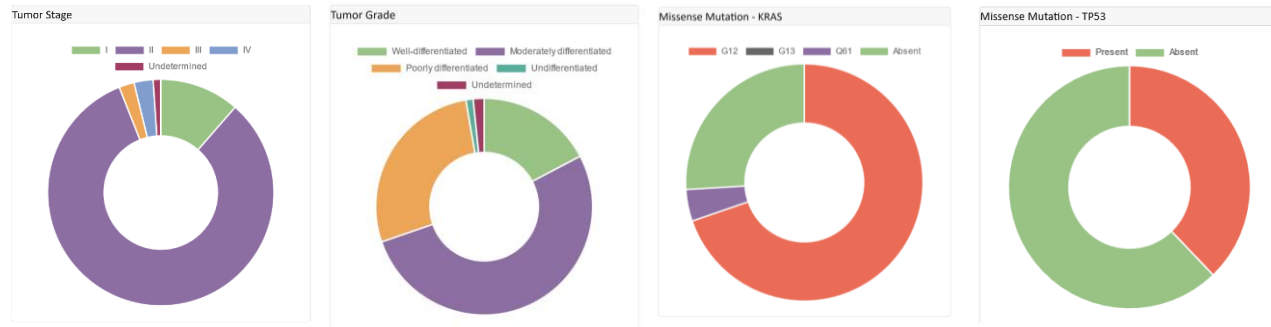
Subtype - Collisson: Classical Exocrine Quasi-mesenchymal

Subtype - PacPaint: Classical Basal

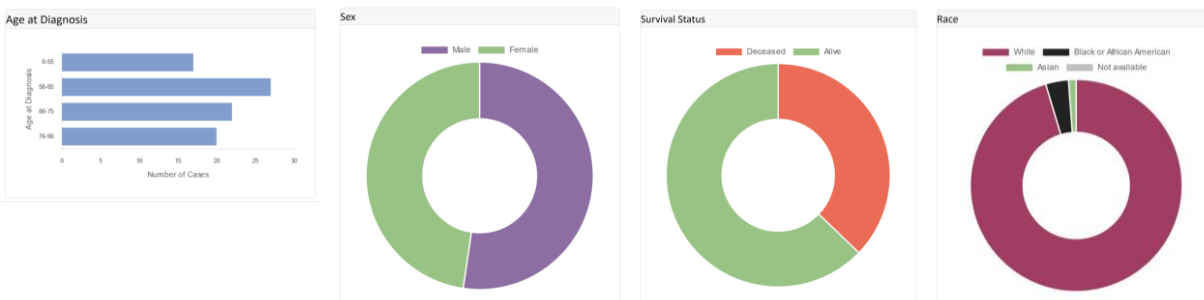
B



C



D



E

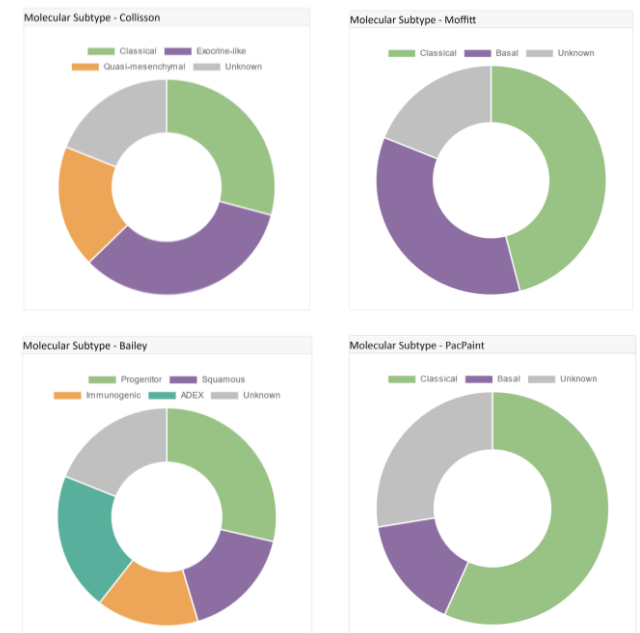


Figure 2. Advanced filtering options and clinical summaries for publicly available PDAC datasets. (A) Available data can be filtered according to various patient-related factors and tumour characteristics, including the stratification and analysis of cohorts according to KRAS and TP53 mutational status and established transcriptomic (TCGA, ICGC), genomic (ICGC) or histologically-derived AI (PacPaint) subtypes. **(B)** Dynamic bar charts allow multiple covariates to be viewed in relation to each other: e.g. survival trends in PDAC (TCGA) and neuroendocrine (ICGC PAEN-AU). **(C, D, E)** Each filtered attribute can be visualized as clinical summaries for each study cohort. Alternatively, data can be downloaded as .csv or .xls files, for offline analysis.

Figure 3. Transcriptomic stratification in PDAC reveals subtype-specific somatic variants

Oncoplots* of the top 25 most frequently mutated genes for consensus **(A)** n=27 *classical*-type and **(B)** n=16 *basal*-type PDAC cases (TCGA). **(C)** Overlap^ between the somatically mutated genes associated with best/worst prognosis subtypes across TCGA and ICGC PACA-AU cohorts combined. Genes highlighted in **bold** contain Tier 1 predicted oncogenic driver variants that have associated pharmacological inhibitors or chemotherapies (see Supplementary Table S1).

*Mutated genes are ranked in order of the *total* number of *mutations* in each given gene (where genes may have >1 mutation present; black 'multi-hit'), while the percentage to the right of each bar reflects the proportion of *samples* altered in the cohort. ^Created in [Venny 2.1](#).

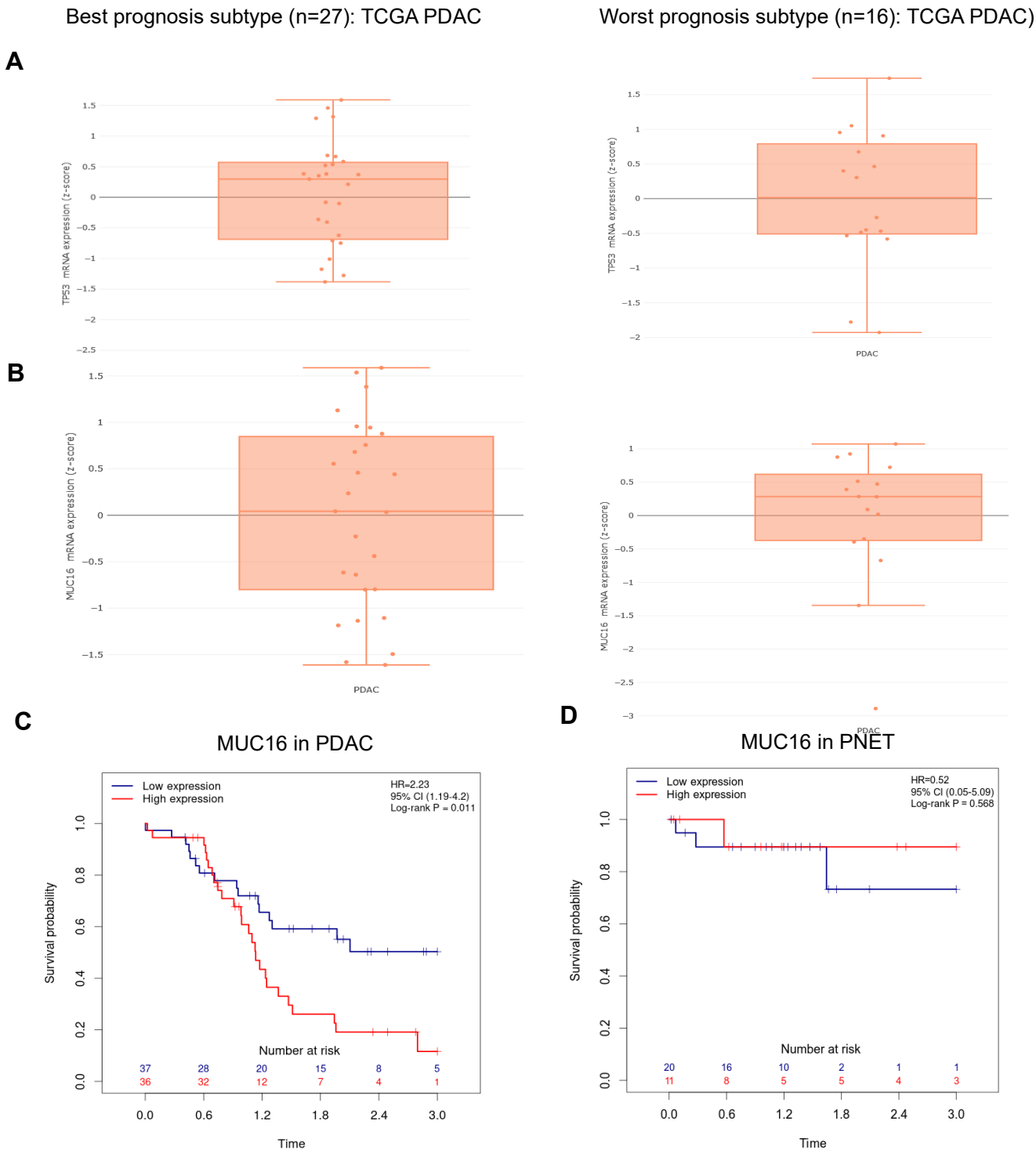
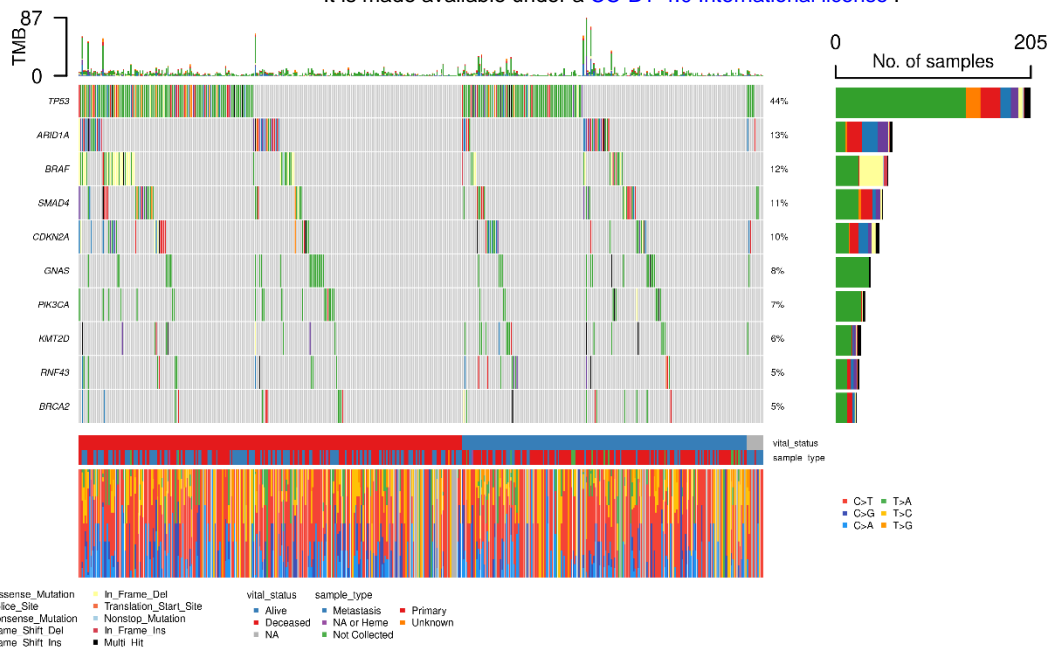


Figure 4. Differentially expressed genes between classical/progenitor and basal-like/QM/squamous TCGA PDAC tumours.

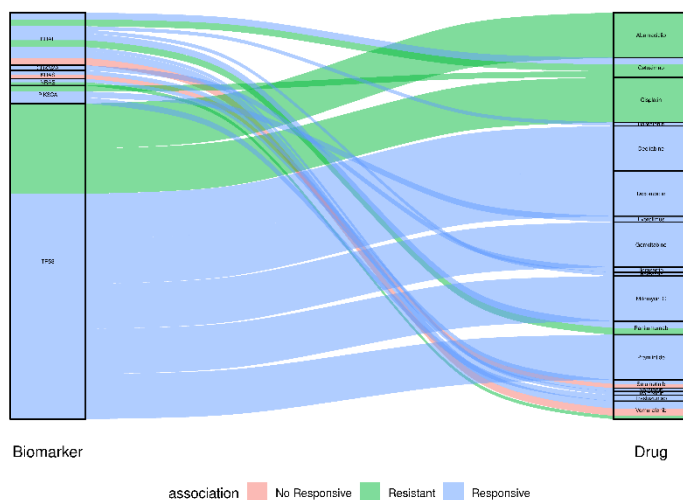
Box plots showing the trends of **(A)** TP53 and **(B)** MUC16 mRNA expression levels across all patients in each filtered TCGA PDAC group; best prognosis (n=27; classical/progenitor; left) and worst prognosis (n=16; basal/squamous/QM; right). **C.** Kaplan-Meier curve showing elevated MUC16 expression significantly associated with lower patient survival over 3 years, from n=402 PACA-AU PDAC patients with expression and outcome data (logrank p=0.011; hazard ratio (HR)=2.23). **D.** No association between MUC16 mRNA expression levels and outcome were observed in n=65 neuroendocrine carcinomas (PAEN-AU).

A



B

Clinically-actionable targets present in at least 5% of the patient cohort



C

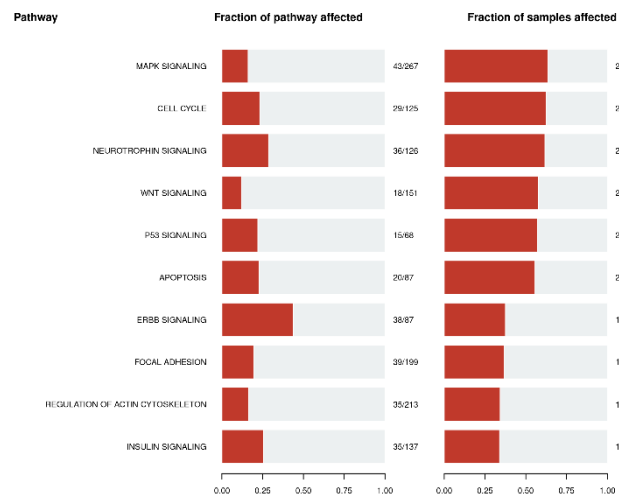


Figure 5. Frequently altered genes and biological pathways amongst n=756 *KRAS* wild-type PDAC tumours from GENIE

(A) Oncoplot showing the top 10 most frequently mutated genes in *KRAS* wild-type PDAC tumours (confirmed somatic missense mutations filtered out; insertions or duplications may still be present). **(B)** Alluvial plot showing gene targets harbouring any variants with therapeutic biomarker potential in $\geq 5\%$ of patients, as identified by the Cancer Genome Interpreter and based on data from [OncoKB](#), [CIVic](#) (Clinical Interpretation of Variants in Cancer) and the [Cancer Biomarkers database](#). **(C)** Altered biological pathways amongst *KRAS* wild-type PDAC tumours include MAPK and p53 signalling, as derived from the KEGG pathway database. The proportion of genes mutated in each pathway (left) and the proportion of all *KRAS* wild-type patients affected (x-axis) are given.

