GATA4 and GATA6 loss-of-expression is associated with extinction of the classical programme and poor outcome in pancreatic ductal adenocarcinoma

Mónica P. de Andrés, [1] Richard Jackson, [2] Irene Felipe, [1,3] Sladjana Zagorac, [1] Christian Pilarsky,[4] Anna Melissa Schlitter, [5,6] Gun-ho Jang, [7] Eithne Costello, [8] Steven Gallinger, [7,9-12] William Greenhalf, [8] Paula Ghaneh, [8] Thomas Knösel,[7] Jaime Martínez de Villarreal, [1,3] Daniel Palmer, [8] Petra Rümmele,[8] Wilko Weichert, [15] Markus Büchler, [16] Thilo Hackert, [16] John P. Neoptolemos, [16] Faiyaz Notta, [7,17,18] Núria Malats, [3,19] Paola Martinelli, [20*] Francisco X. Real [1,3,21]

[1] Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain

[2] Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom

[3] CIBERONC, Madrid, Spain

[4] Department of Surgery, Universitätsklinikum Erlangen, Erlangen, Germany

[5] Institute of Pathology, School of Medicine, Technische Universität München, Munich, Germany

[6] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[7] PanCuRx Translational Research Initiative, Ontario Institute for Cancer Research, Toronto, Ontario, Canada

[8] Department of Molecular and Clinical Cancer Medicine, University of Liverpool

[9] Department of Surgery, University of Toronto, Toronto, Ontario, Canada

[10] Wallace McCain Centre for Pancreatic Cancer, Princess Margaret Cancer Centre, University

[11] Health Network, Toronto, Ontario, Canada

[12] Hepatobiliary/Pancreatic Surgical Oncology Program, University Health Network, Toronto, Ontario, Canada

[13] Institute of Pathology, Ludwig-Maximilians-University Munich, Munich, Germany

[14] Pathologisches Institut, University of Erlangen, Erlangen, Germany

[15] Institute of Pathology of the Technical University Munich, Munich, Germany

[16] Department of General, Visceral, and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany

[17] Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada.

[18] Division of Research, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada
[19] Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre-CNIO, Madrid, Spain
[20] Institute of Cancer Research, Clinic for Internal Medicine I, Medical University of Vienna, Austria
[21] Departament de Medicina i Ciències de la Vida, Universitt Pompeu Fabra, Barcelona, Spain

* Current address: Boehringer Ingelheim RCV GmbH &Co KG, Vienna, Austria

Corresponding author: Francisco X. Real email preal@cnio.es
Word count: 3997

SUMMARY BOX

What is already known about this subject?

- Patients with classical-type PDAC have a better outcome
- Retrospective analyses suggest that classical-type PDAC is more sensitive to 5-FU-based chemotherapy
- GATA6 is a surrogate biomarker of classical tumours and its expression is associated with better survival
- GATA4 and GATA6 have overlapping and unique functions during pancreatic and gastrointestinal development

What are the new findings?

- Combined low expression of GATA4 and GATA6 has a higher transcriptomic impact than either of them alone
- GATA4 and GATA6 display a partially overlapping genome wide distribution
- In PDAC cells, knockdown of GATA4 has more subtle effects than knockdown of GATA6
- GATA4 plays a greater role in the maintenance of the classical phenotype than in repression of the basal programme
- In a large multicenter tissue microarray study, low expression of GATA4 and GATA6 is an independent predictor of survival in patients with resectable PDAC
- GATA4 levels are down-regulated in liver metastases and are negatively correlated with basal markers such as KRT5/6, KRT14, and TP63

How might it impact on clinical practice in the foreseeable future?

- The combined assessment of GATA4 and GATA6 expression may improve prognostic stratification of patients with PDAC and prediction of response to chemotherapy, compared with GATA6 alone
- These findings should be confirmed in prospective studies

Abstract

Objective: GATA6 is a key regulator of the classical phenotype in pancreatic ductal adenocarcinoma (PDAC). Low GATA6 expression associates with poor patient outcome. *GATA4* is the second most expressed GATA factor in the pancreas. We assessed whether, and how, GATA4 contributes to PDAC phenotype and analysed the association of expression with outcome and response to chemotherapy.

Design: We analysed PDAC transcriptomic data, stratifying cases according to *GATA4* and *GATA6* expression, and identified differentially expressed genes and pathways. The genome wide distribution of GATA4 was assessed as well as the effects of *GATA4* knockdown. A multicenter TMA study to assess GATA4 and GATA6 expression in samples (n=745) from patients with resectable was performed. GATA4 and GATA6 levels were dichotomized into high/low categorical variables; association with outcome was assessed using univariable and multivariable Cox regression models.

Results: *GATA4* mRNA is enriched in classical, compared to basal-like tumours. We classified samples in 4 groups as high/low for *GATA4* and *GATA6.* Reduced expression of *GATA4* had a minor transcriptional impact but low expression of *GATA4* enhanced the effects of *GATA6* low expression. GATA4 and GATA6 display a partially overlapping genome wide distribution, mainly at promoters. Reduced expression of both proteins in tumours was associated with the worst patient survival. *GATA4* and *GATA6* expression significantly decreased in metastases and negatively correlated with basal markers.

Conclusions: *GATA4* and *GATA6* cooperate to maintain the classical phenotype. Our findings provide compelling rationale to assess their expression as biomarkers of poor prognosis and therapeutic response.

Word count: 246

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) remains a deadly malignancy with a 5-year survival rate of only 10%[1]. Non-specific symptoms, late diagnosis, and aggressive biology, among other factors, account for the poor outcome[2]. Only 20% of patients have resectable tumours at presentation and the 5-year-survival rate after surgery alone with radical intent is only 8%, rising to 30-50% using adjuvant combination chemotherapy after resection[3,4]. The standard-of-care for patients with resectable disease is adjuvant therapy with modified 5-fluorouracil, leucovorin, irinotecan and oxaliplatin (FOLFIRINOX) or gemcitabine with capecitabine[4,5]. FOLFIRINOX or gemcitabine (with or without Abraxane) are used in locally advanced and metastatic disease[6,7]. Neoadjuvant chemotherapy is increasingly used in patients with borderline resectable PDAC and it may impact on patient outcome since PDAC is considered to be a systemic disease[8]. An improved understanding of the biological basis of PDAC aggressiveness is needed.

Tumour stratification through genomic analysis has unraveled a quite homogenous landscape, where most tumours share recurrent alterations in four genes (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*). In this background, each tumour has a private constellation of additional genomic alterations, rendering PDAC a highly heterogeneous disease. Transcriptomic profiling of PDAC samples has allowed a molecular classification consistently identifying two main tumour subtypes: classical and basal[9–12]. High expression of adhesion- and epithelial-associated genes[9], up-regulation of GATA6 and HNF1A, and better prognosis are hallmarks of the classical subtype[9,10] (also designated as “pancreatic progenitor”[11], and “classical A/B”[12]). Basal tumours ("quasimesenchymal"[9], "basal-like"[10], "squamous"[11], and "basal-like A/B"[12]) are characterized by down-regulation of endodermal transcription factors (e.g. GATA6, HNF1A, and PDX1) and epithelial markers (e.g. CDH1), poor survival, high expression of keratins of stratified epithelium (e.g. KRT14 and KRT5/6)[10], and of ΔNP63[11]. Increasing evidence suggests that PDAC display a continuous, rather than a dichotomous, phenotype resulting both from the co-occurrence of cells with classical and basal features in the same tumour as well as the existence of a continuum of phenotypes at the single cell level[3,12,13]. GATA6 has emerged as a master regulator of canonical differentiation programmes both in normal pancreas and in PDAC[3,14]. A *GATA6* superenhancer is involved in the classical PDAC phenotype[15] and, among all up-regulated pancreatic lineage transcription factors in this subtype, *GATA6* has been found to be recurrently amplified[12]. GATA6 expression is a surrogate biomarker for classical PDAC[16] and is associated with response to adjuvant therapy[3,16].

The GATA family of transcription factors comprises six zinc finger proteins that bind the consensus DNA sequence (T/A)GATA(A/G). They play complex roles in embryonic development and cancer[17]. GATA6 and GATA4 are key regulators of heart development: *GATA6* haploinsufficiency causes congenital heart defects in humans[18,19] and *GATA4* mutations cause cardiac septal defects and dilated cardiomyopathy[20,21]. Regarding the pancreas, *GATA6* mutations can lead to a broad phenotypic spectrum, ranging from adult-onset diabetes to pancreatic hypoplasia/agenesis[18,19]. *GATA4* inactivating mutations and deletions cause neonatal or childhood-onset diabetes, with variable degrees of exocrine insufficiency[22]. *Gata4* and *Gata6* share partially redundant functions during pancreas organogenesis, since concomitant inactivation results in pancreas agenesis while inactivation of only one of them does not impede pancreas formation[23,24]. *Gata4* inactivation in the Pdx1-expressing domain during mouse development results in pancreatic heterotopia in the stomach[25]. Accordingly, GATA4 and GATA6 control shared developmental programmes but also have gene-specific functions.

Here, we show that GATA4 is expressed at high levels in normal human adult pancreas and in a subset of PDAC. Through bioinformatics analysis of available transcriptomic datasets of PDAC[11,12], we find that GATA4 mRNA is down-regulated in the basal subtype and expression of both GATA transcripts is positively correlated. At the molecular level, low expression of GATA4 amplifies the transcriptomic effect of GATA6 loss. Orthogonal validation of GATA4 and GATA6 protein levels in a large series of clinically-annotated PDAC samples from patients undergoing surgical resection with curative intent in the context of clinical trials or standard of care treatment, revealed that concomitant low expression of GATA6 and GATA4 is consistently associated with worse patient outcome than GATA6 low expression alone. Finally, we show that low GATA4 and GATA6 expression is associated with liver metastasis. Our studies provide novel clinical insights supporting a cooperative interaction between GATA4 and GATA6 to sustain the classical PDAC phenotype and uncover GATA4 as a novel marker of tumour progression that may contribute to refine patient stratification.

METHODS AND MATERIALS

Details are provided as Supplementary Information

The following PDAC RNA-Seq datasets were used in the analyses: ICGC (n=96)[11] and PanCuRx (n=247).

Tumour classification

The distribution of GATA4 and GATA6 expression values was assessed and four categories were defined in both datasets: GATA4 high/GATA6 high ($G4^{Hi}/G6^{Hi}$), GATA4 low/GATA6 high ($G4^{Lo}/G6^{Hi}$), GATA4 high/GATA6 low ($G4^{Hi}/G6^{Lo}$), and GATA4 low/GATA6 low ($G4^{Lo}/G6^{Lo}$).

Differential gene expression analysis, Weighted Gene Correlation Network Analysis (WGCNA), and Gene Set Enrichment Analysis (GSEA)

Differential gene expression analysis was performed using Comparative Marker Selection (version 10.1) and Comparative Marker Selection Viewer (version 9) tools[26] through GenePattern using the default parameters. The R package WGCNA was applied to analyse the PanCuRx dataset. The topological overlap matrix (TOM) algorithm was used to identify modules of densely interconnected genes[27]. GSEA was performed using the tool “Investigate Gene Sets” provided by the Molecular Signatures Database v7.4[28].

Patient cohorts and clinical information

Tissue samples for tissue micro arrays (TMAs) were obtained from patients who underwent pancreatic resection for PDAC. All study sites provided tissue microarrays (TMAs) containing tumour samples from patients who underwent surgical resection. A summary of relevant information is provided in Supplementary Tables 1 and 2. Four cohorts were included from Charité University Hospital[29], Klinikum rechts der Isar, TU München, Germany[30], Regensburg and Jena Institutes of Pathology [31], and the ESPAC-3 trial[32]. Information on the COMPASS study is provided in ref XX. Ethics Committee approval information is provided in the Supplementary Material.

Survival analyses

The primary outcome was overall survival (OS); estimates of OS were obtained using Kaplan Meier curves. Univariable and multivariable analyses were performed. Association with response to chemotherapy was assessed in patients from the COMPASS trial (ref).

RESULTS

GATA4 and GATA6 expression is associated with the classical phenotype in PDAC

To assess GATA expression, we examined the transcriptomes of 228 human normal pancreas samples from the Genotype-Tissue Expression (GTEx) Project[33]. GATA4 and GATA6 are the two main GATA factors expressed (Figure 1A). Immunohistochemical analysis of using specific antibodies revealed that GATA4 is detected only in acinar cells whereas GATA6 is detected in acinar, ductal and islet cells (Figure 1B).

To compare GATA4 and GATA6 expression in pancreatic cancer, we took advantage of the transcriptome data from the PanCuRx study that used laser-microdissection to enrich epithelial cells from primary and metastatic tumours[12]. GATA4 and GATA6 are also the main GATA genes expressed in this PDAC series and their levels show a moderate positive correlation (r=0.35, p-value = 1.617e-08; Pearson's test) (Figure 1C,E). We reproduced these findings using transcriptomic data from the Australian ICGC study of non-microdissected PDAC samples[11] (r=0.39, p-value=8.829e-05) (Supplementary Figure 1A,B). Classification of tumour samples according to molecular subtypes revealed that basal tumours display significantly lower levels of GATA4 and GATA6 mRNAs than classical tumours (Figure 1D).

Next, we classified samples from the PanCuRx study according to GATA4 and GATA6 expression levels; thresholds were established according to the distribution of the data (Supplementary Figure 1C, see Methods). Tumours were classified in four groups resulting from dichotomizing each covariate, as follows: GATA4 high/GATA6 high ($G4^{Hi}/G6^{Hi}$) (n=189), GATA4 low/GATA6 high ($G4^{Lo}/G6^{Hi}$) (n=24), GATA4 high/ GATA6 low ($G4^{Hi}/G6^{Lo}$) (n=18), and GATA4 low/GATA6 low ($G4^{Lo}/G6^{Lo}$) (n=16) (Figure 1E, Supplementary Figure 1D).

Using the molecular classification proposed in the PanCuRx study, we examined the association of GATA gene expression with tumour subtypes (classical, hybrid, or basal). The $G4^{Hi}/G6^{Hi}$ group was highly enriched in the Classical or Hybrid subtypes (35% and 29%, respectively). Tumours showing selective down-regulation of GATA4 mRNA were also highly enriched in these subtypes, whereas those showing selective down-regulation of GATA6 mRNA were highly enriched in the Basal subtypes (83%). All tumours showing low expression of both GATA6 and GATA4 belonged to the Basal

subtype and 88% of them were classified as Basal-like A (Figure 1F). To refine these analyses, we examined the expression of classical and basal genes in the epithelial compartment using the signature reported by Moffit et al[10]: selective reduction of GATA6, but not of GATA4, was associated with a significantly lower Z-score of the classical signature. Tumours showing concomitant low GATA4 and GATA6 expression showed the lowest classical score. In contrast, the Z-score of the Basal gene signature was not significantly different between the $G4^{Hi}/G6^{Lo}$ and the $G4^{Lo}/G6^{Lo}$ groups (Figure 1G).

Next, we analysed the status or *KRAS* and *SMAD4* in the four groups using whole-genome sequencing (WGS) data from the PanCuRx study. Tumours with homologous recombination defects and DNA mismatch repair defects [24%] were excluded due to unique mutational signatures[12]. The prevalence of *KRAS*-WT tumours was similar across groups but low expression of GATA4 and/or GATA6 was associated with increased imbalance of the mutant vs wild type allele (Chi-square test, $P<0.0001$; $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$, two-sided Fisher's exact test $P=0.076$) (Figure 1H). Interestingly, 80% of the $G4^{Lo}/G6^{Lo}$ tumours had an intact copy of *SMAD4* (Chi-square test, $P<0.0001$; $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$, two-sided Fisher's exact test $P<0.0001$) (Figure 1I).

In sum, GATA4 is highly expressed in normal human pancreas and in human PDAC, where it is down-regulated in basal tumours. Low expression of GATA4 on its own does not have a major impact on basal genes, suggesting that it participates in the maintenance of the canonical pancreatic differentiation programme when GATA6 expression is lost.

Low GATA4 expression is associated with enhanced transcriptomic effects of GATA6 down-regulation

To identify the mechanisms that might explain the differential contribution of GATA4 and GATA6 to the maintenance of the classical phenotype, we first performed a differential gene expression analysis using genome wide RNA-Seq data from the PanCuRx study. Taking the $G4^{Hi}/G6^{Hi}$ group as reference, 318 genes (87 up- and 231 down-) were significantly differentially expressed in tumours showing low expression of GATA4 alone ($G4^{Lo}/G6^{Hi}$) (q-value<0.05). In contrast, in tumours showing only low GATA6 ($G4^{Hi}/G6^{Lo}$), 3914 genes (1871 up- and 2043 down-) were significantly dysregulated. Low expression of both genes ($G4^{Lo}/G6^{Lo}$) was associated with the most dramatic phenotype, with 6096 differentially expressed genes (2933 up- and 3163 down-) (Fig 2A).

To acquire insight into the direct or indirect effects of down-regulation of GATA4 and GATA6 we performed weighted gene correlation network analysis (WGCNA). This method identifies discrete groups of co-expressed genes and classifies them in modules (i.e. genetic programmes) by using the TOM algorithm[34]. These modules are often enriched for genes that share biological functions[35]. All tumours (n=247) clustered based on their Euclidean distance (Supplementary Figure 2A), excluding the presence of outliers. $G4^{Hi}/G6^{Hi}$ and $G4^{Lo}/G6^{Hi}$ samples distributed evenly across the dendrogram, whereas $G4^{Hi}/G6^{Lo}$ and $G4^{Lo}/G6^{Lo}$ tumours clustered together, in accordance with the results of the differential expression analysis.

To build a weighted gene network it is critical to choose a proper soft thresholding power (β) value. Thus, we screened for different soft thresholding powers and selected β = 4 for later analysis (Supplementary Figure 2B). Next, gene co-expression networks were generated corresponding to 33 co-expression modules (Supplementary Figure 2C,D). Modules were named after colors following WGCNA conventions[34].

GATA4-dependent programmes. We examined modules that were significantly dysregulated between $G4^{Lo}/G6^{Hi}$ and $G4^{Hi}/G6^{Hi}$ tumours by comparing the mean expression (normalized by Z-score) of the genes in each module (Figure 2B). Functional annotation of the grey60 module, selectively up-regulated in the $G4^{Lo}/G6^{Hi}$ group, revealed enrichment of genes involved in keratinization and development. The darkturquoise module, down-regulated in $G4^{Lo}/G6^{Hi}$ samples, contained genes related to apoptosis and/or regulated by p53.

GATA6-dependent programmes. We examined modules that were significantly dysregulated when comparing $G4^{Hi}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$ (describing the effect of low GATA6 expression alone). Only one gene programme (red) was identified, composed of 785 genes accounting for cell cycle, mitosis and DNA repair processes. Other GATA6-dependent modules following the same trend that did not reach significance are the magenta module, with genes involved in WNT signaling (e.g. *WNT7A*, *WNT7B*, *WNT10A*, *WNT11*, *FZD7*, *FZD2*), and the lightgreen module enriched in “Reactome_Interferon_Signaling”.

GATA4-dependent programmes in tumours with low GATA6 expression. We examined modules that were significantly dysregulated when comparing the $G4^{Hi}/G6^{Lo}$ group (GATA6 down-only) with the $G4^{Lo}/G6^{Lo}$ group. Functional annotation of gene programmes consistently up-regulated in the $G4^{Lo}/G6^{Lo}$ group revealed enrichment in processes related to formation of the cornified envelope and sensory perception in the saddlebrown and darkorange modules, respectively (Figure 2B, Supplementary Tables 3 and 4). Enrichment in processes related to the basal phenotype (i.e. "Formation of the

cornified envelope") was also found in the greenyellow module. Importantly, genes related to the ΔNP63 pathway were contained in this module and were consistently up-regulated in the $G4^{Lo}/G6^{Lo}$ group (Figure 2C). In addition, immune pathways such as "cytokine signaling" and "innate immune pathways" (e.g. *IL1A*, *IL1B*, *IL36A*) were enriched in the greenyellow module, also up-regulated in the $G4^{Lo}/G6^{Lo}$ group (Figure 2D). The yellow module was significantly down-regulated in $G4^{Lo}/G6^{Lo}$ samples and was defined by genes involved in lipid metabolism, matrisome/extracellular matrix-related pathways, xenobiotic metabolism, O-linked mucin glycosylation, and inflammation. "Reactome_Metabolism_of_lipids" was the first annotation on the list with 91 overlapping genes (FDR q-value: 3.22E-21): 75/91 genes were down-regulated and 16/91 were up-regulated in the $G4^{Lo}/G6^{Lo}$ group (Figure 2E). Among the former are those related to fatty acid metabolism (e.g. *FA2H*, *MGLL*), cholesterol biosynthesis (e.g. *FDFT1*), acetyl-choline metabolism (e.g. *ACHE*, *ACSL5*), and xenobiotic metabolism (e.g. *UGT8, CYP3A4*). *GM2A* and *CYP24A1* are the most significant lipid-related genes up-regulated in $G4^{Lo}/G6^{Lo}$ samples.

The genomic distribution of GATA4 overlaps partially with that of GATA6

To assess the genome wide distribution of GATA4 and compare it to that of GATA6, we applied cleavage under targets and release using nuclease (Cut&Run) on PaTu8988S cells, because the available antibodies failed to provide reliable results using ChIP-Seq (not shown). Peak number . The GATA motif showed the highest enrichment in the peaks called, supporting the validity of this experimental approach (Figure 3A). Other motifs significantly enriched are those corresponding to HNF and FOXA, which are bound by well-established endodermal transcription factors involved in pancreatic differentiation (XX), AP-1, and ELF. Seven % and 87% of the peaks localized to transcription start sites and intergenic/intronic regions, respectively (Figure 3B). Functional annotation analysis (DAVID suite)[36] on 1426 GATA4 peaks located in promoter-annotated regions (-1kb, +100bp from TSS) revealed pathways similar to those identified when analysing GATA6-promoter bound regions[3] such as "axon guidance", "Wnt signaling pathway", "MAPK signaling pathway" although the P value was of borderline significance, probably due to the smaller number of genes included in the analysis (Supplementary Table 5). "Glycerophospholipid metabolism" was GATA4-specific since it was uniquely enriched in GATA4 promoter-bound genes.

We compared the genomic distribution of GATA4 with that of GATA6, using data from our previously published work also using PaTu8988S cells (xx). Considering all genomic locations, only 20% of binding regions are shared between GATA4 and GATA6.

In contrast, 78% of those corresponding to promoters are shared. Integration of DNA binding in promoters with transcriptomic data revealed that 127/137 of the up-regulated and 159/200 of the down-regulated genes in $G4^{Lo}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$ samples bound and regulated by GATA4 were also bound and regulated by GATA6, suggesting that GATA4 does not have a unique set of regulated genes when compared to GATA6 (Figure 3C, E). Functional annotation of bound and up-regulated genes revealed that most significantly enriched pathways were related to cell cycle (Figure 3D). Sterol and lipid metabolism pathways were among the processes participated by genes bound by both factors and differentially down-regulated in $G4^{Lo}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$ samples (Figure 3F).

We next compared GATA4 and GATA6 binding to genes included in the classical and basal signatures reported by Moffitt *et al.* In accordance with its role as a regulator of the classical programme, we found GATA6 binding sites in 23/23 genes from the classical signature and 17 of them were also bound by GATA4. Interestingly, only two genes (*TFF2*, *VSIG2*) were bound by both factors at promoter regions, suggesting a cooperation through gene-specific proximal and distal regulatory elements (Figure 3G). In line with the finding that GATA4 plays a lesser role in the regulation of the basal programme (Figure 1G), we did not find GATA4 binding to the promoter of genes in the basal signature (Figure 3H). This is in contrast with GATA6, which binds the promoters of 6/XX basal genes (*FAM83A*, *KRT15*, *S100A2*, *SLC2A1*, *TNS4* and *LEMD1*) (Figure 3H). These results indicate that GATA4 mainly regulates PDAC cell differentiation by directly acting on classical, and not basal, genes.

GATA4 knockdown has a subtle impact on the classical phenotype

To determine the effects of GATA4 down-regulation, we used lentiviral-mediated knockdown in two PDAC cell lines (DANG and PaTu8988S) expressing GATA4 and GATA6 and showing classical features. Upon efficient knockdown, the morphological effects of GATA4 down-regulation were subtler than those reported for GATA6 (Figure 4A and 4E). There were no significant/consistent changes in GATA6 mRNA or protein expression but expression of FOXA1/2 was significantly lower in both cell lines (Figure 4B-D, 4F-H), an effect that is similar to that previously reported for GATA6[3]. Consistently, the Cut&Run experiments showed GATA4 binding to the promoter of both *FOXA1* and *FOXA2* (Supplementary Figure 3A,B). In line with the morphological observations, no consistent changes were observed in CDH1 mRNA expression (Figure 4D,4H). Altogether, these findings indicate that at the phenotypic level loss of GATA4 also has less impact than loss of GATA6 [3].

Low expression of GATA4 and GATA6 is associated with the worst patient outcome

To interrogate whether the molecular cooperation described above has a clinical translation, we evaluated GATA4 and GATA6 protein expression in a multicenter TMA study of PDAC samples from patients undergoing tumour resection (n=745). This TMA series includes both patients enrolled in a clinical trial of adjuvant chemotherapy (ESPAC-3 cohort)[32] and a more diverse group of patients receiving standard therapy[29–31]. Table 1 summarizes the characteristics of the patients included in the study. We first compared the clinical features of the various patient series, focusing on outcome. Median overall survival was highest for patients from the ESPAC-3 study (Cohort 4) (23.5 months vs 14.5, 17.6, and 12.4 for the other series) (Supplementary Figure 4A). Considering all patients together, median survival was higher for patients who had received adjuvant chemotherapy: 5-FU (23.5 months), gemcitabine (22.8 months), or unknown (22.0 months) (differences not being significant) vs those who had not (14.6 months) (Supplementary Figure 4B). Similar survival was observed among patients from the ESPAC-3 trial and those receiving adjuvant therapy in the conventional setting (not shown).

GATA4 and GATA6 protein expression levels were also dichotomized as described for mRNA expression. GATA6 histoscores were distributed more evenly across samples than GATA4 histoscores, the distribution of which was skewed towards lower values (Figure 5A). A moderate, significant, correlation between the expression of both proteins was observed (rho=0.31, p-value = <2.2e-16; Spearman's test) (Figure 5B). Similar results were obtained when testing each individual cohort separately (Supplementary Figure 5A-H). Representative immunohistochemical results are shown in Figure 5C.

Survival analyses showed that GATA4 expression, as a dichotomous variable, was not associated with overall survival [hi vs low, 19.4 (95% CI 16.6-22.1) vs 17.5 (14.6-19.3)] (Figure 5D). In contrast, GATA6 expression was significantly associated with a shorter overall survival [hi vs low, 19.7 (17.4-21.6) vs 13.0 (10.9-17.6), p value<0.001] (Figure 5E). Patients with tumours showing low expression of both GATA4 and GATA6 had the worst outcome ($G4^{Lo}/G6^{Lo}$ vs all others, 12.2 (9.3-19.9) vs 19.3 (17.3-21.0), p value<0.001) (Figure 5F). When the results of each series were analysed separately, the same trend was observed but the differences were not statistically significant, likely due to the smaller sample size (Supplementary Figure 6A-D). Similar results were obtained when the same strategy was applied to the transcriptomic data from the Australian Pancreatic Cancer Initiative (Figure 5G).

Univariable analysis of all factors included in the dataset was performed (Table 1). Age, stage, grade, lymph node involvement, positive resection margins, adjuvant therapy, and GATA6 as a single marker were all significantly associated with outcome. For the multivariable analysis, all the clinical/pathological terms were considered in the final model, selected using a backwards stepwise selection. In addition, the combined expression of GATA4/GATA6 was included. Grade, lymph node involvement, adjuvant therapy, resection margins, and low expression of both GATA proteins were significantly and independently associated with survival. The HR for low expression of both GATAs, when compared against any other configuration of protein expression, was 1.59 (95%CI 1.22-2.10; p<0.001) (Table 2). Exploratory multivariable analyses suggest that the impact of GATA4/6 is consistent across all levels of the adjuvant therapy covariate (not shown); the interaction term between GATA4/6 status and adjuvant therapy was non-significant (Supplementary Table 5 and Supplementary Figure 7). These findings reveal that combined GATA4/GATA6 expression identifies the group of resectable PDAC patients with the worst outcome.

GATA6 expression has been proposed as a predictor of response to FOLFIRINOX chemotherapy in the COMPASS trial [16]. In comparison with GATA6, GATA4 has a lesser predictive effect in the same patient series. Nevertheless, the combination of GATA4 and GATA6 improved the predictive value of either alone, despite the smaller sample size of the $G4^{Lo}/G6^{Lo}$ group (Supplementary figure 8A,B). These results suggest that patients with $G4^{Lo}/G6^{Lo}$ tumours are might benefit more from other therapies.

Low expression of GATA4 and GATA6 is associated with increased liver metastasis

To assess GATA4 and GATA6 expression during tumour progression, we analyzed their expression as well as a panel of well-established classical and basal markers - in a small cohort of 6 matched pairs of primary PDAC and distant metastasis (see methods), as well as 2 additional metastases. We classified samples in four groups according to GATA4 and GATA6 median histoscore. All primary tumours tested along with their paired metastasis had a high expression of CDH-1 and most of them lacked the expression of the tested basal markers (Figure 6A).

Leveraging on the PanCuRx dataset, which includes a large number of metastatic samples, we compared expression of GATA4, GATA6, and the basal and classical markers in primary and metastatic tumours. We observed a significant positive correlation between GATA6 and CDH1 in metastatic samples while no significant

correlation was observed with GATA4 (Supplementary Figure 9A). In the case of the basal markers KRT5/6, KRT14 and TP63, we observed a consistent significant negative correlation with both GATA factors in metastatic samples, also less significant in primary tumours (Supplementary Figure 9B-D). These findings confirm that GATA4 and GATA6 expression is inversely associated with the basal phenotype in metastatic samples. Expression of GATA4 and GATA6 was significantly lower in metastases than in primary tumours (Figure 6B). Interestingly, 50% of samples in the $G4^{Hi}/G6^{Lo}$ group and 80% of the $G4^{Lo}/G6^{Lo}$ samples were liver metastases (Figure 6C).

DISCUSSION

In PDAC, GATA6 is a marker and a master regulator of the classical subtype. Here, we provide strong evidence that GATA4 plays an important role in the absence of GATA6. The effect of GATA4 on PDAC biology is reflected on patient outcome.

GATA transcription factors bind with high affinity to consensus GATA sites and have very similar binding specificities. A challenge in the understanding of their biology is the dissection of their overlapping and unique functions. In some cases, selective expression of one vs other GATA genes accounts for tissue-specific functions. This is the case of GATA6 in peritoneal macrophages[37] or of GATA4 in senescence in fibroblasts[38]. When multiple GATA proteins are expressed in a given cell, dissecting gene-specific functions is more challenging. Genetic mouse models have shown that deletion of *Gata4* or *Gata6* does not impact on pancreatic development, while deletion of both genes blunts pancreatic proliferation and differentiation[23,24]. A study using human pluripotent stem cells has reported the relevance of GATA4 dosage in a GATA6 heterozygous context, highlighting the interactions between both factors[39]. In the intestine, inactivation of *Gata4* and *Gata6* unveils a time- and space-constrained redundancy[40]. The complexity underlying their cooperation has been revealed in murine colorectal cancer, where *Gata6* deletion in adenomas results in increased BMP levels and a blockade of tumour stem cell self-renewal[41]. Genetically-engineered expression of one of the proteins under the regulatory elements of the paralogue gene has shown that GATA4 is required for liver and heart development[42].

The cooperativity between GATA proteins in cancer has been poorly explored. In PDAC, GATA6 has been widely studied but there are few, inconclusive, reports on GATA4 expression and activity[43–46]. *GATA6* is amplified and overexpressed in a subset of pancreatic and hepatobiliary tumours[11,12,47,48]. In contrast, *GATA4* copy number changes are infrequent. We hypothesized that the combined analysis of GATA4 and GATA6 expression might reveal specific biological functions and relevant clinical

outcome associations. Of note, GATA6 mRNA levels have been shown to accurately correlate with protein levels in human PDAC[16] but there is no published information regarding the correlation for GATA4. RNA-Seq analysis of tumours from the PanCuRx series using an agnostic module identification strategy revealed that GATA6 down-regulation has the major transcriptomic impact; tumours displaying only low GATA4 levels had a molecular profile similar to those with preserved expression of both transcription factors. Dependency on GATA4 and GATA6 has been reported in gastric cancer *in vitro* upon knockdown of one or both genes, either directly or indirectly through other transcription factors, including CDX2[49].

Modules with higher activity in tumours with combined down-regulation of GATA4 and GATA6 were enriched in "basal-ness"-related genes and in the ΔNp63 pathway. GATA6 down-regulation is necessary, but not sufficient, to activate the expression of the ΔNp63 programme and the basal phenotype both in mouse and patient tumours. The cooperative effect resulting from the absence of both GATA proteins is similar to that observed when GATA6 is down-regulated concomitantly with HNF1A or HNF4[50]. These pathways likely contribute to the aggressiveness of the Basal tumours by favouring an epithelial-to-epithelial transition that precedes an EMT. Genes in the cytokine signaling and immune pathways were also enriched in $G4^{Lo}/G6^{Lo}$ tumours, possibly contributing to immune escape and metastasis[50]. GATA4 and GATA6 levels were lower in liver metastases and they negatively correlate with basal markers. Comparison of $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$ metastatic samples identified CDX2 among the top 5 down-regulated genes in $G4^{Lo}/G6^{Lo}$ samples (not shown).

$G4^{Lo}/G6^{Lo}$ tumours also showed a down-regulation in genes involved in cholesterol and fatty acid biosynthesis. Metabolic profiling and bioinformatic analyses have revealed two main PDAC metabolic subtypes: cholesterogenic and glycolytic. The former is associated with the classical subtype and with a better prognosis, while the latter is associated with the basal subtype[51,52]. Accordingly, using previously described signatures[52], we found a significant up-regulation of the glycolysis pathway and down-regulation of the cholesterol biosynthesis pathway in the $G4^{Lo}/G6^{Lo}$ tumours (not shown). Disruption of cholesterol synthesis favors a switch to a basal phenotype that is mediated by up-regulation of TGF-beta[53]. Several membrane small metabolite transporters are expressed at higher levels in classical than in basal tumours; among them is the cholesterol transporter NPC1L1, the pharmacological and genetic inhibition of which can suppress the growth of PDAC cells in vitro and in mice[54]. The transcriptional cooperation resulting from the down-regulation of both GATA proteins supports the acquisition of aggressive tumour features.

To acquire a better understanding of the role of GATA4 in PDAC, we used Cut&Run to identify its genome wide binding profile, and compare it with that of GATA6. Together with the transcriptome analyses, the findings support largely overlapping roles in the maintenance of the classical programme. Our data also suggest that GATA6 has a broader role by also affecting the basal programme. However, this conclusion should be tempered by the fact that different techniques were used to assess the genomic localization of the two proteins. Knockdown of GATA4 in PDAC cells also had lesser effects than we previously reported for GATA6 and we did not find evidence that GATA4 knockdown impacts on GATA6 expression. These findings suggest that stable levels of GATA6 upon GATA4 knockdown reduce the impact of the latter.

Our study was prompted by the clinical value of GATA6 as a marker of outcome and therapeutic stratification. In the largest patient series reported until now, we show that patients with resectable tumours showing low levels of GATA6, but not of GATA4, have poor outcome but patients with tumours expressing low levels of both proteins had the worse survival. The association with outcome was significant as an independent predictor in the multivariable analysis and the magnitude of the effects - while modest - is similar to that of the other important outcome predictors such as lymph node involvement, administration of adjuvant therapy, and positive resection margins. The association was consistent across patient cohorts involving both patients selected for trials and those treated in the standard clinical setting. The large sample size of our study supports the robustness of the observations, though the findings are restricted to patients undergoing surgical resection of the primary tumour. Regarding therapeutic stratification, our study suggests that patients with $G4^{Lo}/G6^{Lo}$ tumours are not likely to benefit from FOLFIRINOX therapy.

Future studies should consider assessing the correlation between GATA4 mRNA and protein expression and the use of full sections instead of TMA cores to account for cellular heterogeneity. *GATA6* and *GATA4* are among the genes with higher 5hmC density in circulating cell free DNA in patients with PDAC[55]. 5hmC is associated with transcription and its cfDNA profiles mirror the hydroxy-methylation profile in tumour samples, suggesting that it may be possible to subtype patients using non-invasive methods to select optimal therapy.

Overall, this work shows that GATA4 has a less critical role than GATA6 in the maintenance of the classical PDAC programme. In the absence of GATA6, the contribution of GATA4 becomes relevant, as shown by the fact that loss of expression of both genes/proteins is associated with the most dramatic transcriptomic and clinical phenotypes. The robustness of the IHC assays, already confirmed by independent

investigators in the case of GATA6[56], calls for validation of our clinical associations in prospective studies and the extension to patients with advanced PDAC. It will also be important to confirm whether the combined assessment of GATA6 and GATA4 is useful for therapeutic stratification.

Acknowledgements

We thank Peter Bailey, Eduard Batlle, Raquel Benítez, Adrià Caballe, Mark Kalisz, Bernhard Kloesch, Miriam Marqués, Juan Méndez, Sara Rodríguez, and Camille Stephan-Otto Atolini for valuable contributions and the Histopathology Unit of CNIO for technical help.

Funding

The work was supported, in part, by the following grants: RTI2018-101071-B-I00 (Ministerio de Ciencia, Innovación y Universidades-MCIU, Madrid, Spain) to FXR, PI18/01347 (ISCIII-FIS, Madrid, Spain) to NM, PRECODE (EU-MSCA: 861196) to CP. MPdA was supported by a Ph.D. Fellowship from MCIU. CNIO is supported by MCIU as a Centro de Excelencia Severo Ochoa (grant SEV-2015-0510).

Competing interests: The authors declare no conflict of interest.

Author contributions

MPdA and FXR, together with PM, designed the study, analysed the results, and wrote the manuscript. MPdA performed the majority of the experiments; IF carried out the Cut&Run experiments and SZ performed the functional *in vitro* studies. The bioinformatics analyses were done by MPdA and JMdV. RJ performed the statistical analysis with contributions from NM. CP, AMS, TK, PR, and WW provided samples and clinical information. EC, WG, PG, DP, MB, TH, and JPN were involved in the ESPAC-3 trial and provided samples and clinical information. GHJ, SG, and FN obtained data and analysed the associations with treatment response in the COMPASS trial. All authors provided comments to the manuscript. FXR supervised the overall conduct of the study and obtained funds.

References

1 Siegel RL, Miller KD, Fuchs HE, *et al.* Cancer Statistics, 2021. *CA Cancer J Clin* 2021;**71**:7–33.

2 Kleeff J, Korc M, Apte M, *et al.* Pancreatic cancer. *Nat Rev Dis Prim* 2016;**2**:1–23.

3 Martinelli P, Carrillo-De Santa Pau E, Cox T, *et al.* GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. *Gut* 2017;**66**:1665–76.

4 Conroy T, Hammel P, Hebbar M, *et al.* FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. *N Engl J Med* 2018;**379**:2395–406.

5 Neoptolemos JP, Palmer DH, Ghaneh P, *et al.* Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. *Lancet* 2017;**389**:1011–24.

6 Conroy T, Desseigne F, Ychou M, *et al.* FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *N Engl J Med* 2011;**364**:1817–25.

7 Von Hoff DD, Ervin T, Arena FP, *et al.* Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. *N Engl J Med* 2013;**369**:1691–703.

8 Strobel O, Neoptolemos J, Jäger D, *et al.* Optimizing the outcomes of pancreatic cancer surgery. *Nat Rev Clin Oncol* 2019;**16**:11–26.

9 Collisson EA, Sadanandam A, Olson P, *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* 2011;**17**:500–3.

10 Moffitt RA, Marayati R, Flate EL, *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* 2015;**47**:1168–78.

11 Bailey P, Chang DK, Nones K, *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;**531**:47–52.

12 Chan-Seng-Yue M, Kim JC, Wilson GW, *et al.* Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat Genet* 2020;**52**:231–40.

13 Juiz N, Elkaoutari A, Bigonnet M, *et al.* Basal-like and classical cells coexist in pancreatic cancer revealed by single-cell analysis on biopsy-derived pancreatic cancer organoids from the classical subtype. *FASEB J* 2020;**34**:12214–28.

14 Martinelli P, Cañamero M, Del Pozo N, *et al.* Gata6 is required for complete

acinar differentiation and maintenance of the exocrine pancreas in adult mice. *Gut* 2013;**62**:1481–8.

15 Lomberk G, Blum Y, Nicolle R, *et al.* Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. *Nat Commun* 2018;**9**:1–10.

16 O'Kane GM, Grünwald BT, Jang G-H, *et al.* GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. *Clin Cancer Res* 2020;**26**:4901–10.

17 Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in development and disease. *Development* 2018;**145**.

18 Gong M, Simaite D, Kühnen P, *et al.* Two novel GATA6 Mutations cause childhood-onset diabetes mellitus, pancreas malformation and congenital heart disease. *Horm Res Paediatr* 2013;**79**:250–6.

19 Chao CS, McKnight KD, Cox KL, *et al.* Novel GATA6 mutations in patients with pancreatic agenesis and congenital heart malformations. *PLoS One* 2015;**10**:1–9.

20 Ang YS, Rivas RN, Ribeiro AJS, *et al.* Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* 2016;**167**:1734-1749.e22.

21 Zhao L, Xu JH, Xu WJ, *et al.* A novel GATA4 loss-of-function mutation responsible for familial dilated cardiomyopathy. *Int J Mol Med* 2014;**33**:654–60.

22 Shaw-Smith C, De Franco E, Allen HL, *et al.* GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes* 2014;**63**:2888–94.

23 Carrasco M, Delgado I, Soria B, *et al.* GATA4 and GATA6 control mouse pancreas organogenesis. *J Clin Invest* 2012;**122**:3504–15.

24 Xuan S, Borok MJ, Decker KJ, *et al.* Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J Clin Invest* 2012;**122**:3516–28.

25 Rodríguez-Seguel E, Villamayor L, Arroyo N, *et al.* Loss of GATA4 causes ectopic pancreas in the stomach. *J Pathol* 2020;**250**:362–73.

26 Gould J, Getz G, Monti S, *et al.* Comparative gene marker selection suite. *Bioinformatics* 2006;**22**:1924–5.

27 Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**.

28 Liberzon A, Birger C, Thorvaldsdóttir H, *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* 2015;**1**:417–25.

29 Muckenhuber A, Berger AK, Schlitter AM, *et al.* Pancreatic Ductal Adenocarcinoma Subtyping Using the Biomarkers Hepatocyte Nuclear Factor-

1A and Cytokeratin-81 Correlates with Outcome and Treatment Response. *Clin Cancer Res* 2018;**24**:351 LP – 359.

30 Schlitter AM, Segler A, Steiger K, *et al.* Molecular, morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes. *Sci Rep* 2017;**7**:1–12.

31 Pérez-Mancera PA, Rust AG, Van Der Weyden L, *et al.* The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature* 2012;**486**:266–70.

32 Neoptolemos JP, Stocken DD, Bassi C, *et al.* Adjuvant Chemotherapy With Fluorouracil Plus Folinic Acid vs Gemcitabine Following Pancreatic Cancer Resection A Randomized Controlled Trial. *JAMA* 2010;**304**:1073–81.

33 GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 2017;**550**:204–13.

34 Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;**4**.

35 van Dam S, Võsa U, van der Graaf A, *et al.* Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform* 2018;**19**:575–92.

36 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57.

37 Rosas M, Davies LC, Giles PJ, *et al.* The transcription factor Gata6 links tissue macrophage phenotype and proliferative renewal. *Science (80-)* 2014;**344**:645–8.

38 Kang C, Xu Q, Martin TD, *et al.* The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4. *Science (80-)* 2015;**349**:aaa5612.

39 Shi ZD, Lee K, Yang D, *et al.* Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development. *Cell Stem Cell* 2017;**20**:675-688.e6.

40 Walker EM, Thompson CA, Kohlnhofer BM, *et al.* Characterization of the developing small intestine in the absence of either GATA4 or GATA6. *BMC Res Notes* 2014;**7**:1–12.

41 Whissell G, Montagni E, Martinelli P, *et al.* The transcription factor GATA6 enables self-renewal of colon adenoma stem cells by repressing BMP gene expression. *Nat Cell Biol* 2014;**16**:695–707.

42 Borok MJ, Papaioannou VE, Sussel L. Unique functions of Gata4 in mouse liver induction and heart development. *Dev Biol* 2016;**410**:213–22.
43 Jiang W, Chen C, Huang L, *et al.* GATA4 Regulates Inflammation-Driven Pancreatic Ductal Adenocarcinoma Progression. 2021;**9**:1–16.
44 Gong Y, Zhang L, Zhang A, *et al.* GATA4 inhibits cell differentiation and proliferation in pancreatic cancer. *PLoS One* 2018;**13**:e0202449.
45 Karafin MS, Cummings CT, Fu B, *et al.* The developmental transcription factor GATA4 is overexpressed in pancreatic ductal adenocarcinoma. *Int J Clin Exp Pathol* 2010;**3**:47–55.
46 Fu B, Guo M, Wang S, *et al.* Evaluation of GATA-4 and GATA-5 methylation profiles in human pancreatic cancers indicate promoter methylation patterns distinct from other human tumor types. *Cancer Biol Ther* 2007;**6**:1546–52.
47 The Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 2017;**32**:185-203.e13.
48 Kwei KA, Bashyam MD, Kao J, *et al.* Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreatobiliary cancer. *PLoS Genet* 2008;**4**.
49 Sulahian R, Casey F, Shen J, *et al.* An integrative analysis reveals functional targets of GATA6 transcriptional regulation in gastric cancer. *Oncogene* 2014;**33**:5637–48.
50 Kloesch B, Ionasz V, Paliwal S, *et al.* A GATA6-centred gene regulatory network involving HNFs and ΔNp63 controls plasticity and immune escape in pancreatic cancer. *Gut* 2021;:1–12.
51 Daemen A, Peterson D, Sahu N, *et al.* Metabolite profiling stratifies pancreatic ductal adenocarcinomas into subtypes with distinct sensitivities to metabolic inhibitors. *Proc Natl Acad Sci U S A* 2015;**112**:E4410–7.
52 Karasinska JM, Topham JT, Kalloger SE, *et al.* Altered gene expression along the glycolysis–cholesterol synthesis axis is associated with outcome in pancreatic cancer. *Clin Cancer Res* 2020;**26**:135–46.
53 Gabitova-Cornell L, Surumbayeva A, Peri S, *et al.* Cholesterol Pathway Inhibition Induces TGF-β Signaling to Promote Basal Differentiation in Pancreatic Cancer. *Cancer Cell* 2020;**38**:567-583.e11.
54 Nicolle R, Blum Y, Marisa L, *et al.* Pancreatic Adenocarcinoma Therapeutic Targets Revealed by Tumor-Stroma Cross-Talk Analyses in Patient-Derived Xenografts. *Cell Rep* 2017;**21**:2458–70.
55 Guler GD, Ning Y, Ku CJ, *et al.* Detection of early stage pancreatic cancer using

5-hydroxymethylcytosine signatures in circulating cell free DNA. *Nat Commun* 2020;**11**:1–12.

56 Duan K, Jang G-H, Grant RC, *et al.* The value of GATA6 immunohistochemistry and computer-assisted diagnosis to predict clinical outcome in advanced pancreatic cancer. *Sci Rep* 2021;**11**:14951.

Figure legends

Figure 1. GATA4 and GATA6 are expressed in pancreatic cancer and are down-regulated in the basal subtype. (A) mRNA expression of GATA1-6 in normal human pancreas. Data was extracted from GTEx (see methods). (B) Representative IHC images of the expression of GATA4 and GATA6 in normal human pancreas. GATA6 is expressed in acinar, islet (dashed encircled area) and ductal cells (black arrow) whereas GATA4 is only detected in acinar cells. Scale bar: 20uM. (C) mRNA expression of GATA1-6 in the PanCuRx PDAC dataset. (D) Gene expression levels of GATA4 and GATA6 among PDAC subtypes from the PanCuRx dataset (boxplots are annotated by a Kruskal-Wallis P value). (E) Scatter plot showing a positive correlation between GATA4 and GATA6 mRNA expression (Pearson correlation) and sample classification after GATA4 and GATA6 data dichotomization in the PanCuRx dataset (see methods). (F) Stacked bar plot representing PDAC subtype distribution according to GATA4/GATA6 expression categories (Chi-square test, $P<0.0001$). (G) Z-score of the "Classical" and "Basal" signatures[10] for each of the samples according to the GATA4/GATA6 expression categories; Z-score was calculated as the mean expression (adjusted by Z-score) of the genes in the signature (Kruskal-Wallis $P<0.0001$, two-sided Mann–Whitney test $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$). (H, I) Stacked bar plot representing *KRAS* status or *SMAD4* status according to the GATA4/GATA6 expression categories in the PanCuRx samples (Chi-square test, $P<0.0001$).

Figure 2. GATA4 down-regulation is associated with enhancement of the transcriptomic effect of GATA6 down-regulation in a module-dependent fashion. (A) Stacked bar plot representing the number of differentially regulated genes (q-value<0.05) (Chi-square test, $P<0.0001$). (B) Heatmap representing mean expression normalized by Z-score of genes contained in each module. Each gene programme is designated by color, numbers of genes per gene programme are shown in parentheses. Grey dot, square or triangle denote $P<0.05$ in $G4^{Lo}/G6^{Hi}$ vs $G4^{Hi}/G6^{Hi}$; $G4^{Hi}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$ and $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$ comparisons respectively (two-sided Mann–Whitney test). (C) Heatmap indicating expression of genes that overlap in the ΔNP63 pathway (PID) and the greenyellow module. (D) Violin plot and heatmap indicating expression of genes that overlap in Innate Immune System and Cytokine Signaling (REACTOME) signatures and the greenyellow module. (E) Violin plot and heatmap indicating expression of genes that overlap in the Lipid Metabolism (REACTOME)

signature and the yellow module. Asterisk indicates *P*<0.05 when comparing $G4^{Hi}/G6^{Lo}$ vs $G4^{Lo}/G6^{Lo}$ samples (two-sided Mann–Whitney test).

Figure 3. Genome wide GATA4 binding and integration with transcriptome data reveal cell cycle and lipid metabolism as the main processes directly controlled by GATA4 and GATA6. (A) Top motifs found in GATA4 peaks using the HOMER known motifs tool, sorted by P-value. (B) Pie chart representing the genomic distribution of GATA4 binding regions in PaTu8988S cell line. (TTS: transcription termination site; TSS: transcription start site). (C) Venn diagram showing the overlap of GATA4- and GATA6-bound and up-regulated genes in $G4^{Lo}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$ (D) Bar plot representing top functional annotations of 127 commonly bound and up-regulated genes. (E) Venn diagram showing the overlap of GATA4- and GATA6-bound and down-regulated genes in $G4^{Lo}/G6^{Lo}$ vs $G4^{Hi}/G6^{Hi}$. (F) Bar plot representing top functional annotations of 159 commonly bound and down-regulated genes. (G) Venn diagram depicting genes from the classical signature from Moffitt *et al.* bound by GATA4 and GATA6 including all genomic locations. Genes with a peak in their promoter are highlighted in boxes. (H) Similar analysis using the basal signature from Moffitt *et al.*

Figure 4. GATA4 knockdown has subtle effects on the PDAC classical phenotype. Experiments were performed using PaTu8988S (A-D) and DANG (E-H) PDAC cells. (A,E) Phase contrast microphotographs of cells infected with either shCtrl or two different GATA4-targeting shRNAs (shG4-1 and shG4-2). Scale bar: 100 µM. (B,F) GATA4 and GATA6 expression detected by western blotting in total lysates from PaTu8988S cells infected with the indicated constructs. (C,G) Bar plot representing western blot signal quantification by densitometry. (D,H) RT-qPCR analysis of GATA4, GATA6, FOXA1, FOXA2, and CDH1 transcripts in cells infected with the indicated constructs. Graphs show mean ± SD of at least 3 independent experiments (two-sided Mann–Whitney test).

Figure 5. Patients whose tumours express low levels of both GATA4 and GATA6 have the worse outcome. (A) Density plot showing GATA4 and GATA6 histoscore distribution in all cases (n=745). (B) Scatter plot showing a positive correlation between GATA4 and GATA6 histoscore across all samples (Spearman correlation). (C) Representative images showing GATA4 and GATA6 expression, detected by IHC, in tumour samples from each of the combined expression categories. Scale bar: 50uM. Magnification, scale bar: 5uM. (D) Kaplan-Meier plot of overall survival by GATA4 IHC status (n=745). (E) Kaplan-Meier plot of overall survival by GATA6 IHC status (n=745).

(F) Kaplan-Meier plot showing the association of combined GATA4 and GATA6 IHC status with survival (n=745). (G) Kaplan-Meier plot comparing survival of patients from ICGC-Bailey et al.[11] according to the expression of GATA4 and GATA6 mRNA (n=96).

Figure 6. Down-regulation of GATA4 and GATA6 is associated with increased liver metastasis. (A) Heatmap representing relative histoscore of the indicated genes per paired sample. P= primary tumour, M=metastatic sample. (B) GATA4 and GATA6 expression in primary and metastatic samples from the PanCuRx dataset (Two-sided Mann–Whitney test). (C) Stacked bar plot representing the distribution of primary tumours and metastatic samples according to GATA4/GATA6 expression categories (Chi-square test, $P<0.0001$).

Table 1. Baseline demographics, clinical-pathological characteristics and expression of GATA4 and GATA6 in tumours from patients included in the study. Results of the univariable analysis of survival.

Summary statistics			Univariable analysis			
COVARIATE		**TOTAL**	**N (events)**	**Median OS (95% CI)**	**HR (95% CI)**	**P-val**
Total		745				
Age	<65	331 (44%)	331 (252)	20.5 (16.99, 23.92)		
	>65	280 (38%)	280 (235)	16.39 (14.37, 19.55)	1.29 (1.08, 1.542)	0.005
	median (IQR)	64 (57, 70.614)			1.011 (1.001, 1.0221)	0.0272
Sex	Female	269 (36%)	269 (216)	17.72 (14.73, 21.22)		
	Male	343 (46%)	343 (272)	18.86 (16.59, 21.06)	0.913 (0.763, 1.092)	0.318
Stage	1	25 (3%)	25 (14)	32.76 (20.14, NA)		
	2	135 (18%)	135 (108)	24.44 (18.43, 28.62)	1.582 (0.907, 2.762)	
	3	536 (72%)	536 (428)	16.92 (15.31, 19.45)	2.039 (1.197, 3.475)	
	4	39 (5%)	39 (36)	8.83 (6.64, 15.41)	3.448 (1.858, 6.399)	< 0.001
Grade	1	44 (6%)	44 (37)	22.65 (19.25, 28.62)		
	2	387 (52%)	387 (299)	22.11 (20.14, 24.8)	0.915 (0.65, 1.288)	
	3	293 (39%)	293 (241)	13.8 (12.67, 15.24)	1.358 (0.96, 1.92)	< 0.001
Lymph nodes	Negative	181 (24%)	181 (121)	26.18 (22.32, 32.26)		
	Positive	555 (74%)	555 (466)	16.26 (14.73, 17.72)	1.73 (1.414, 2.115)	< 0.001
Resection Margins	R0	388 (52%)	388 (296)	20.37 (17.25, 23.43)		
	R1	200 (27%)	200 (173)	16.1 (13.67, 20.1)	1.301 (1.078, 1.571)	0.007
Adjuvant Therapy	Yes	375 (50%)	141 (109)	14.55 (11.6, 18.1)		
	No	141 (19%)	375 (301)	22.6 (19.81, 25.13)	0.632 (0.507, 0.788)	< 0.001
GATA 6	<150	322 (43%)	322 (263)	16.36 (14.47, 19.29)		
	>150	380 (51%)	380 (299)	22.05 (18.33, 24.48)	0.765 (0.648, 0.903)	0.002
	median (IQR)	156.7 (66.7, 223.3)			0.998 (0.997, 0.999)	<0.001
GATA 4	<30	194 (26%)	194 (159)	18.1 (15.67, 22.67)		
	>30	345 (46%)	345 (270)	19.48 (16.53, 22.74)	0.948 (0.779, 1.153)	0.594
	median (IQR)	30 (0, 70)			0.999 (0.998, 1.001)	0.474

Table 2. Multivariable modeling of overall survival in the patients included in the study.

Covariate	Level	Estimate (se)	HR (95% CI)	P-value
Grade	Poor			
	Moderate	-0.31 (0.093)	0.73 (0.612, 0.88)	0.001
	Well	-0.21 (0.181)	0.81 (0.569, 1.157)	0.248
Stage	1			
	2	0.26 (0.293)	1.3 (0.733, 2.313)	0.368
	3	0.33 (0.286)	1.39 (0.791, 2.427)	0.255
	4	0.85 (0.337)	2.33 (1.206, 4.514)	0.012
Nodes	Negative			
	Positive	0.49 (0.111)	1.63 (1.309, 2.022)	<0.001
Adjuvant therapy	Yes			
	No	0.46 (0.202)	1.58 (1.064, 2.348)	0.023
	Unknown	0.51 (0.202)	1.66 (1.12, 2.469)	0.012
Resection margins	No			
	Yes	0.29 (0.104)	1.34 (1.091, 1.638)	0.005
	Unknown	-0.11 (0.261)	0.9 (0.538, 1.499)	0.681
GATA 4/6	Other			
	Lo-Lo	0.47 (0.14)	1.59 (1.22, 2.10)	0.001

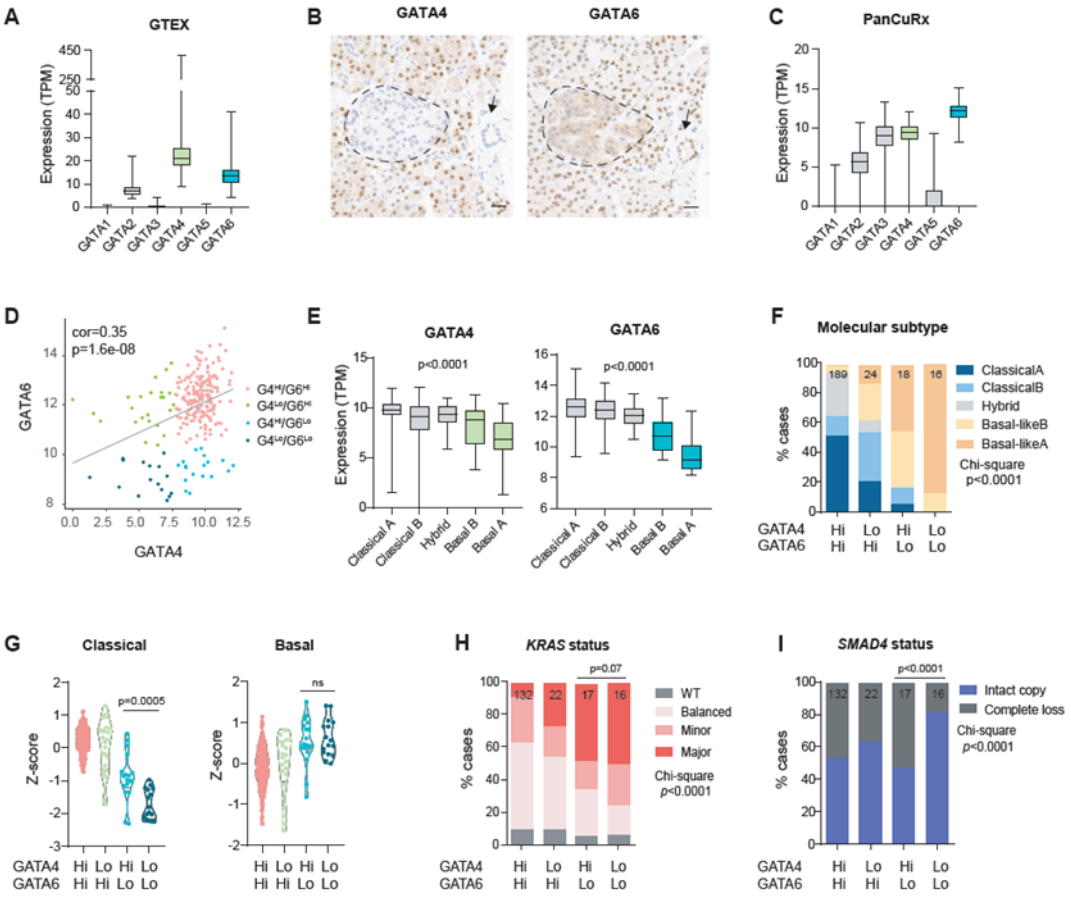

A
GTEX
Expression (TPM)
GATA1
GATA2
GATA3
GATA4
GATA5
GATA6
B
GATA4
GATA6
C
PanCuRx
Expression (TPM)
D
cor=0.35
p=1.6e-08
GATA6
GATA4
G4Hi/G6Hi
G4Lo/G6Hi
G4Hi/G6Lo
G4Lo/G6Lo
E
GATA4
GATA6
p<0.0001
Expression (TPM)
Classical A
Classical B
Hybrid
Basal B
Basal A
F
Molecular subtype
% cases
ClassicalA
ClassicalB
Hybrid
Basal-likeB
Basal-likeA
Chi-square
p<0.0001
GATA4
GATA6
G
Classical
Basal
p=0.0005
ns
Z-score
H
KRAS status
p=0.07
% cases
WT
Balanced
Minor
Major
Chi-square
p<0.0001
I
SMAD4 status
p<0.0001
% cases
Intact copy
Complete loss
Chi-square
p<0.0001

Figure 1

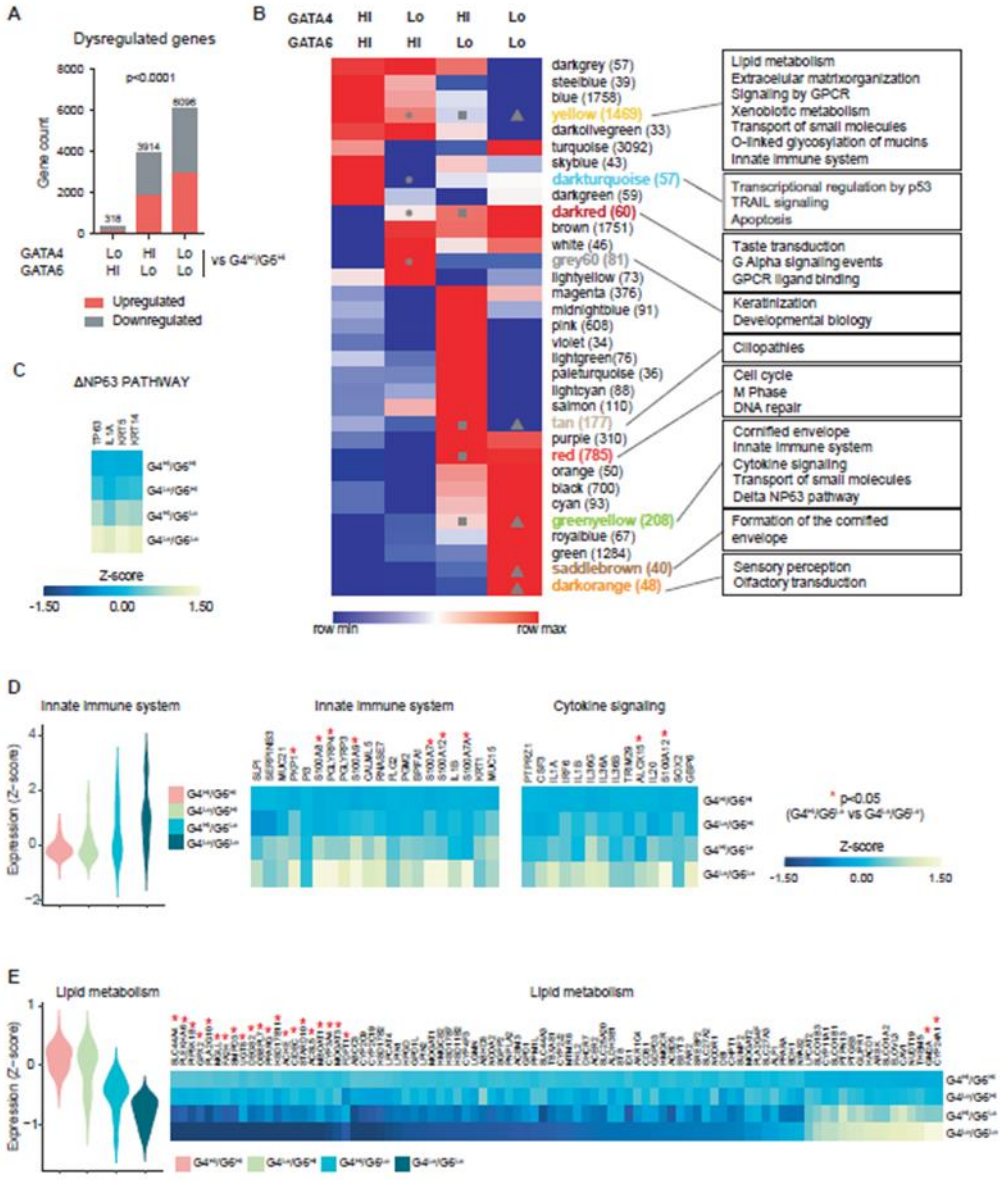

A
Dysregulated genes
p<0.0001
Gene count
8000
6000
4000
2000
0
318
3914
6096
GATA4 Lo HI Lo
GATA6 HI Lo Lo
vs G4HI/G6HI
Upregulated
Downregulated
B
GATA4 HI Lo HI Lo
GATA6 HI HI Lo Lo
darkgrey (57)
steelblue (39)
blue (1758)
yellow (1489)
darkolivegreen (33)
turquoise (3092)
skyblue (43)
darkturquoise (57)
darkgreen (59)
darkred (60)
brown (1751)
white (46)
grey60 (81)
lightyellow (73)
magenta (376)
midnightblue (91)
pink (608)
violet (34)
lightgreen(76)
paleturquoise (36)
lightcyan (88)
salmon (110)
tan (177)
purple (310)
red (785)
orange (50)
black (700)
cyan (93)
greenyellow (208)
royalblue (67)
green (1284)
saddlebrown (40)
darkorange (48)
Lipid metabolism
Extracelular matrixorganization
Signaling by GPCR
Xenobiotic metabolism
Transport of small molecules
O-linked glycosylation of mucins
Innate immune system
Transcriptional regulation by p53
TRAIL signaling
Apoptosis
Taste transduction
G Alpha signaling events
GPCR ligand binding
Keratinization
Developmental biology
Ciliopathies
Cell cycle
M Phase
DNA repair
Cornified envelope
Innate immune system
Cytokine signaling
Transport of small molecules
Delta NP63 pathway
Formation of the cornified envelope
Sensory perception
Olfactory transduction
row min
row max
C
ΔNP63 PATHWAY
TP63
ILIA
KRT5
KRT14
Z-score
-1.50
0.00
1.50
D
Innate immune system
Innate immune system
Cytokine signaling
Expression (Z-score)
* p<0.05
(G4HI/G6HI vs G4LO/G6LO)
Z-score
-1.50
0.00
1.50
E
Lipid metabolism
Lipid metabolism
Expression (Z-score)

Figure 2

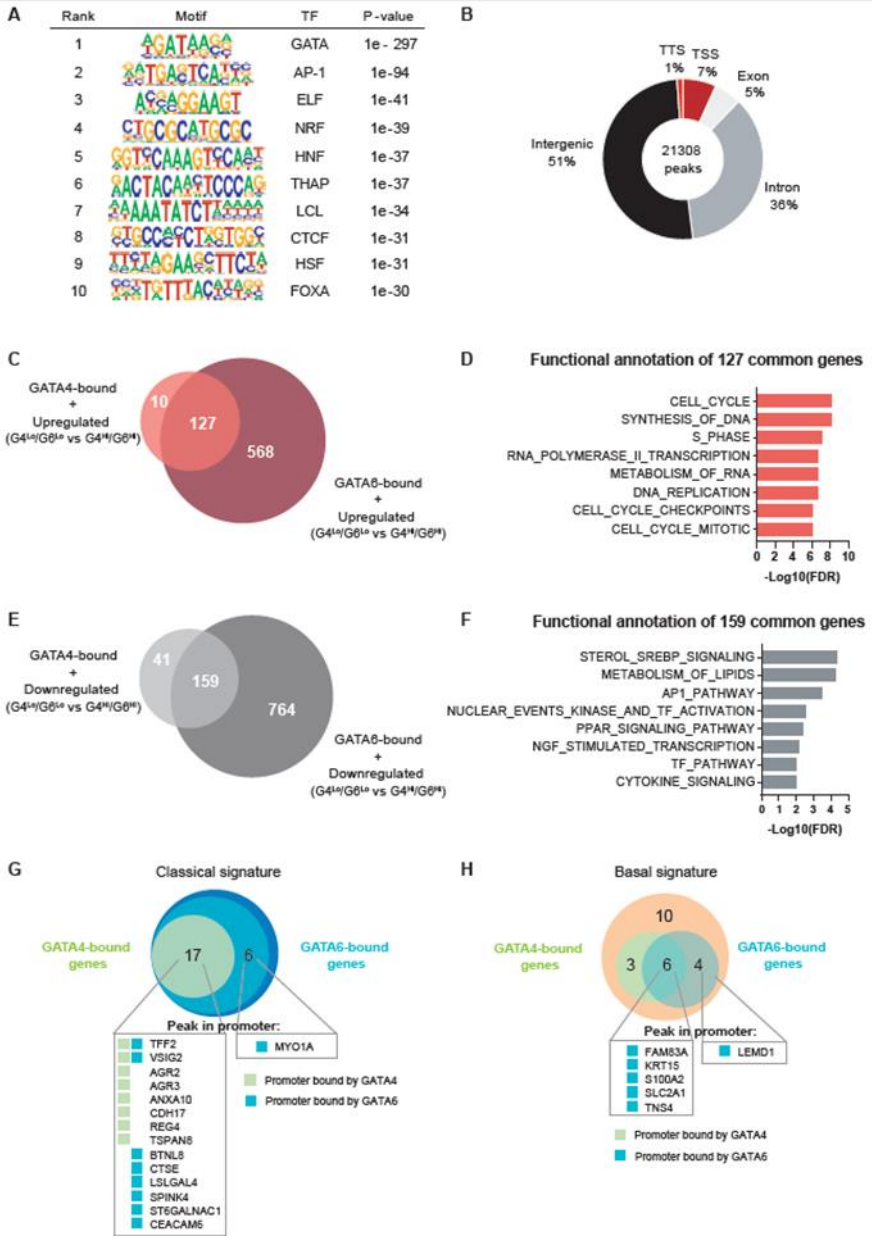

A
Rank
Motif
TF
P-value
1
GATA
1e-297
2
AP-1
1e-94
3
ELF
1e-41
4
NRF
1e-39
5
HNF
1e-37
6
THAP
1e-37
7
LCL
1e-34
8
CTCF
1e-31
9
HSF
1e-31
10
FOXA
1e-30
B
TTS 1%
TSS 7%
Exon 5%
Intergenic 51%
21308 peaks
Intron 36%
C
GATA4-bound + Upregulated (G4LA/G6LA vs G4HI/G6HI)
10
127
568
GATA6-bound + Upregulated (G4LA/G6LA vs G4HI/G6HI)
D
Functional annotation of 127 common genes
CELL_CYCLE
SYNTHESIS_OF_DNA
S_PHASE
RNA_POLYMERASE_II_TRANSCRIPTION
METABOLISM_OF_RNA
DNA_REPLICATION
CELL_CYCLE_CHECKPOINTS
CELL_CYCLE_MITOTIC
0 2 4 6 8 10
-Log10(FDR)
E
GATA4-bound + Downregulated (G4LA/G6LA vs G4HI/G6HI)
41
159
764
GATA6-bound + Downregulated (G4LA/G6LA vs G4HI/G6HI)
F
Functional annotation of 159 common genes
STEROL_SREBP_SIGNALING
METABOLISM_OF_LIPIDS
AP1_PATHWAY
NUCLEAR_EVENTS_KINASE_AND_TF_ACTIVATION
PPAR_SIGNALING_PATHWAY
NGF_STIMULATED_TRANSCRIPTION
TF_PATHWAY
CYTOKINE_SIGNALING
0 1 2 3 4 5
-Log10(FDR)
G
Classical signature
GATA4-bound genes
17
6
GATA6-bound genes
Peak in promoter:
TFF2
VSIG2
AGR2
AGR3
ANXA10
CDH17
REG4
TSPAN8
BTNL8
CTSE
LSLGAL4
SPINK4
ST6GALNAC1
CEACAM6
MYO1A
Promoter bound by GATA4
Promoter bound by GATA6
H
Basal signature
10
GATA4-bound genes
3
6
4
GATA6-bound genes
Peak in promoter:
FAM83A
KRT15
S100A2
SLC2A1
TNS4
LEMD1
Promoter bound by GATA4
Promoter bound by GATA6

Figure 3

Figure 4

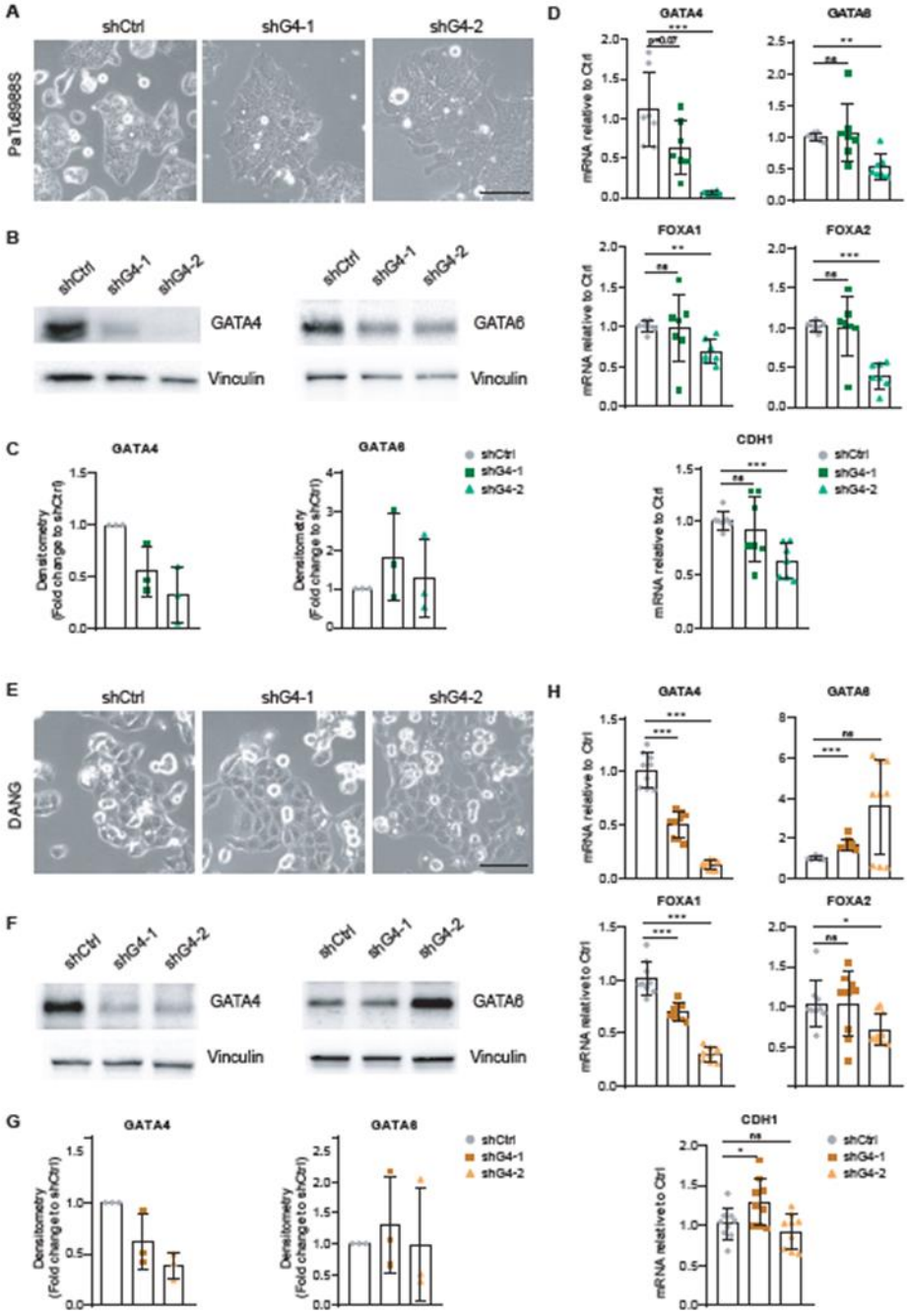


Figure 4

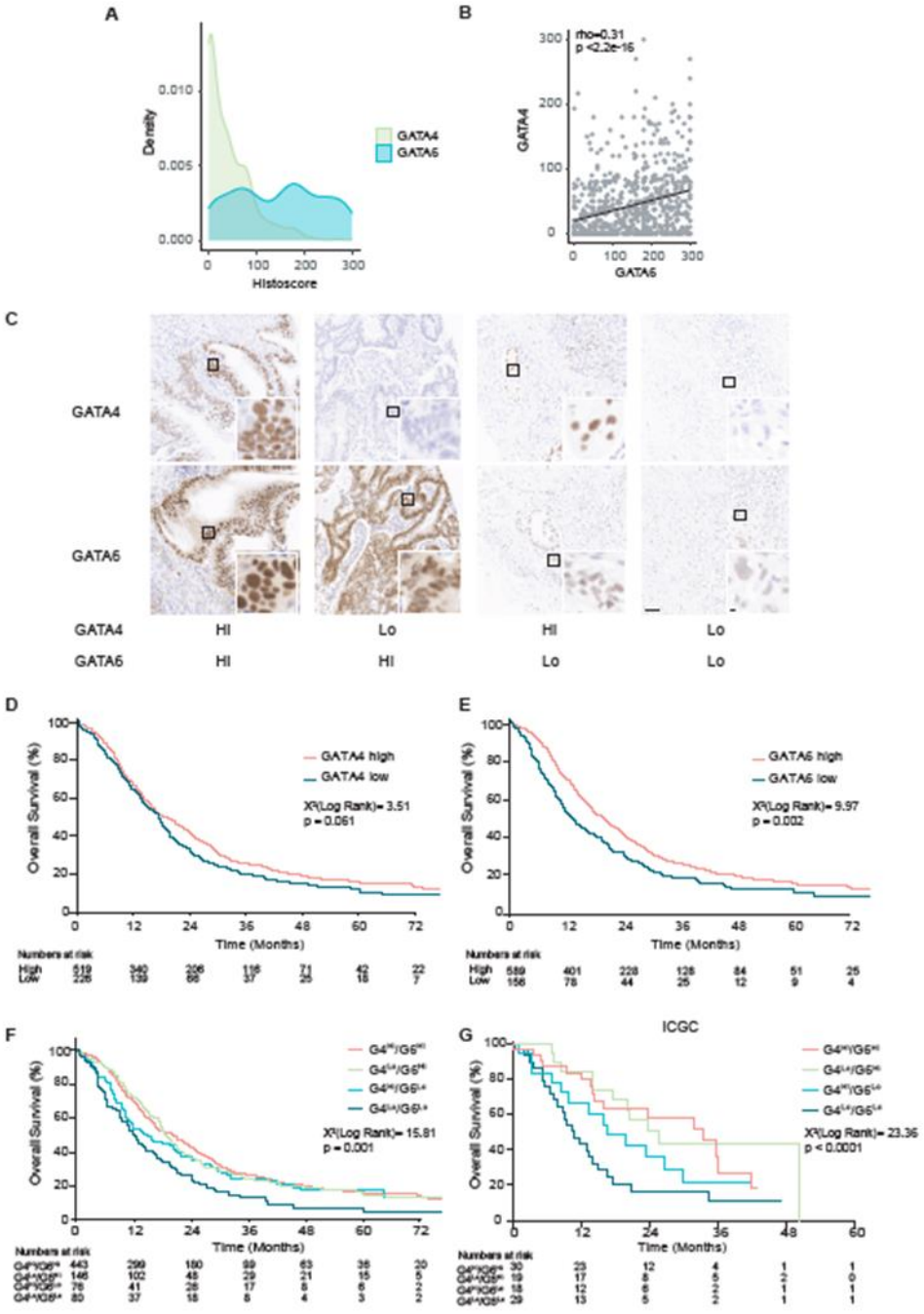

A
Density
0.010
0.005
0.000
0
100
200
300
Histoscore
GATA4
GATA6
B
rho=0.31
p <2.2e-16
GATA4
GATA6
C
GATA4
GATA6
GATA4 HI Lo HI Lo
GATA6 HI HI Lo Lo
D
Overall Survival (%)
GATA4 high
GATA4 low
X²(Log Rank)= 3.51
p = 0.061
Time (Months)
Numbers at risk
High 519 340 226 116 71 42 22
Low 226 139 66 37 25 18 7
E
Overall Survival (%)
GATA6 high
GATA6 low
X²(Log Rank)= 9.97
p = 0.002
Time (Months)
Numbers at risk
High 589 401 228 128 84 51 25
Low 156 78 44 25 12 9 4
F
Overall Survival (%)
X²(Log Rank)= 15.81
p = 0.001
Time (Months)
Numbers at risk
443 299 180 99 63 35 20
146 102 48 29 21 15 5
76 41 26 17 8 6 2
80 37 18 8 4 3 2
G
ICGC
Overall Survival (%)
X²(Log Rank)= 23.36
p < 0.0001
Time (Months)
Numbers at risk
30 23 12 4 1 1
19 17 8 5 2 0
18 12 6 2 1 1
29 13 5 2 1 1

Figure 5

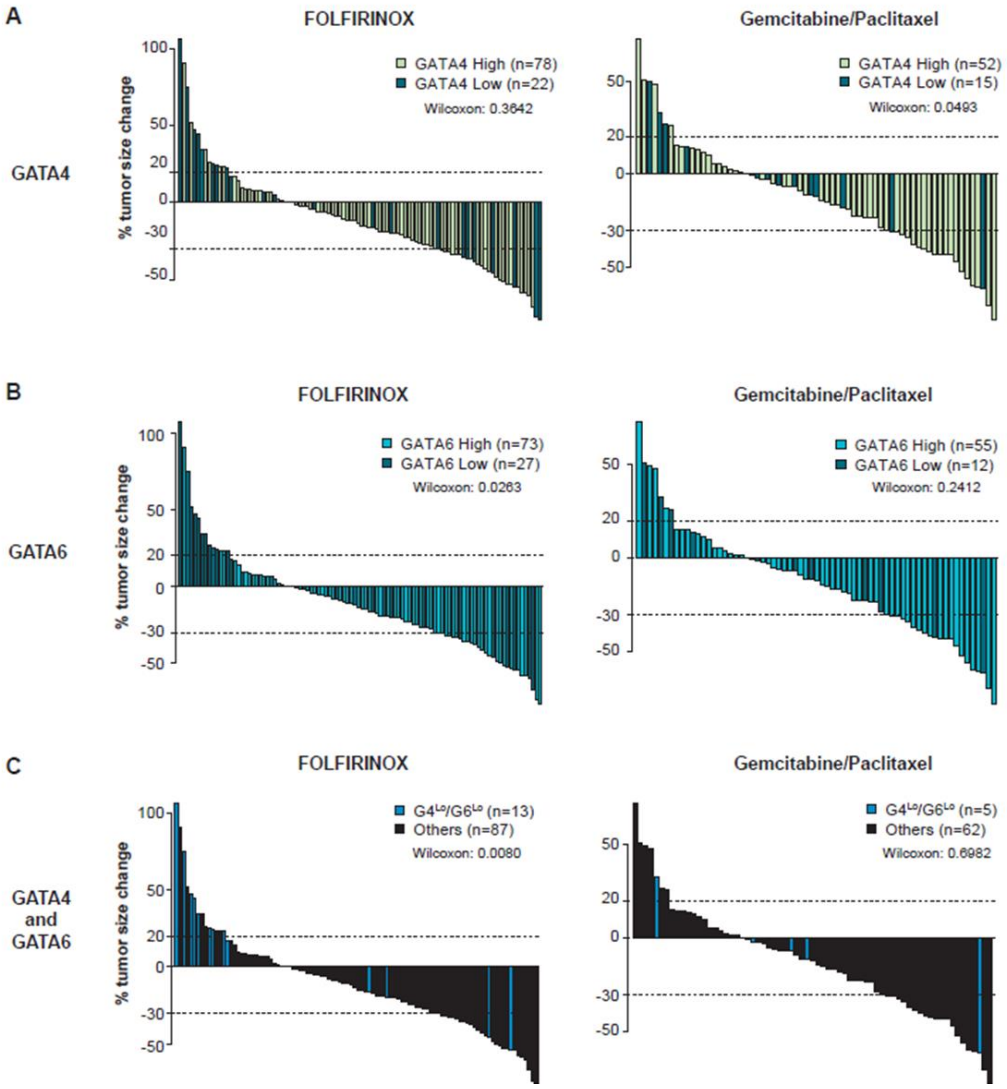


Figure 6

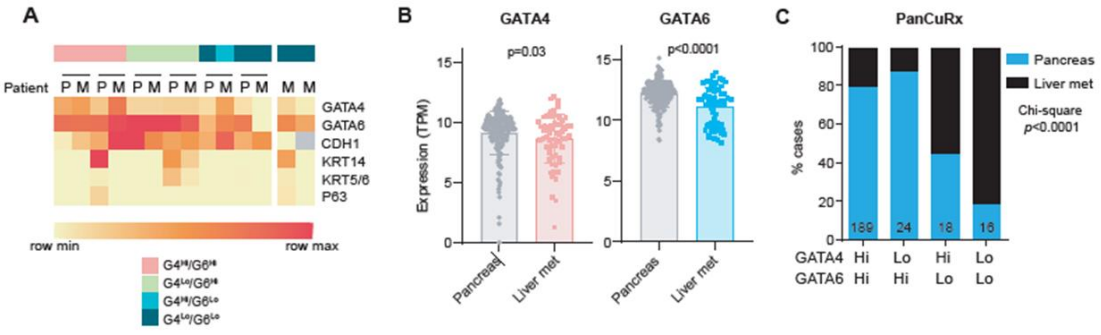
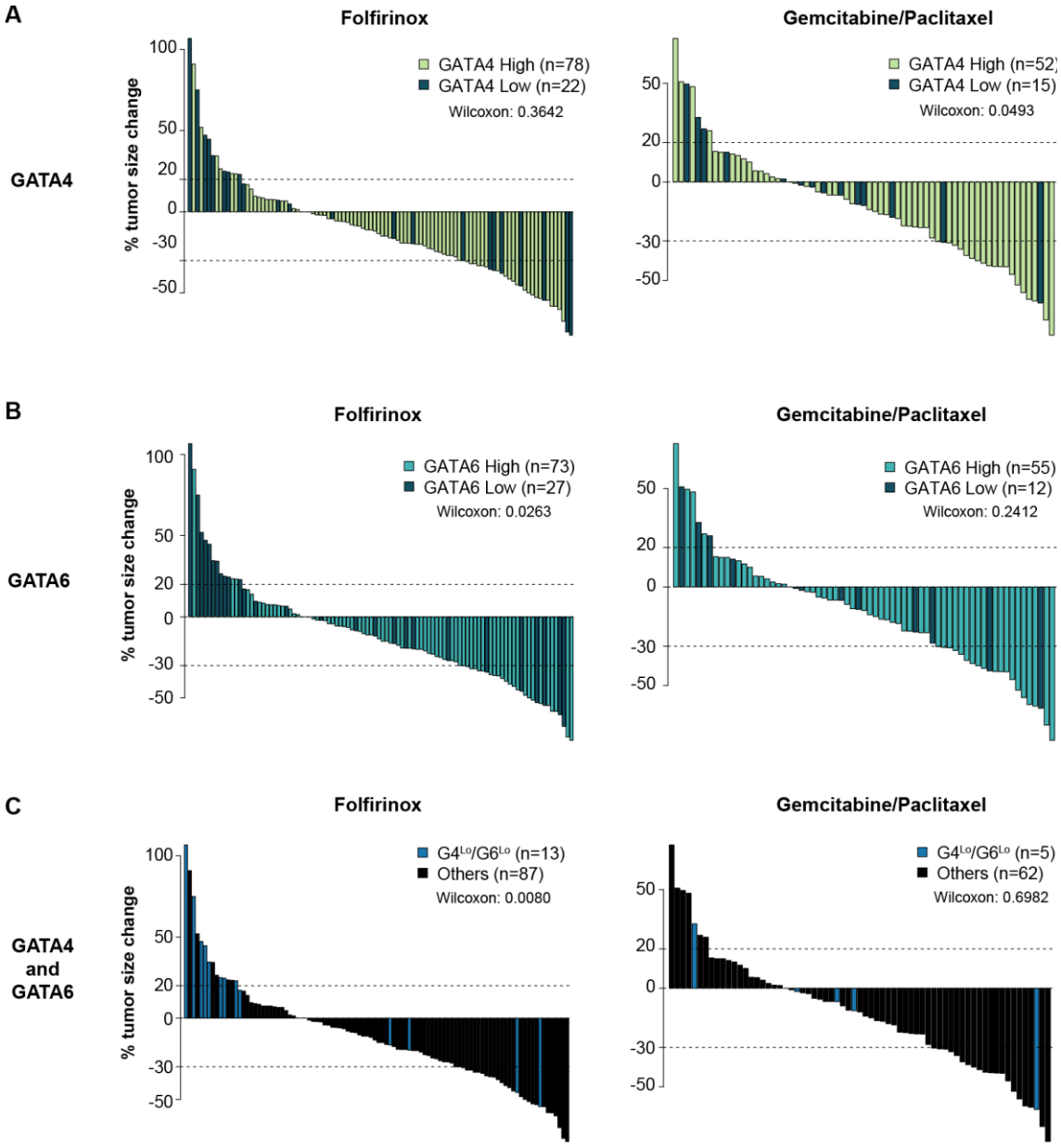


Figure 7



Supplementary Figure 8. Low GATA4 and GATA6 expression is associated with lack of response to FOLFIRINOX. (A-C) Waterfall plots of tumor response based on GATA4 (A), GATA6 (B) levels alone or in combination (C) from patients treated with FOLFIRINOX or Gemcitabine/Paclitaxel. P values were obtained using a two-sided Wilcoxon rank-sum test.