Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples from patients with pancreatic neoplasia

Jun Yu et al.

SUPPLEMENTARY MATERIALS, METHODS AND RESULTS

Cells and tissues

Twenty pancreatic cancer cell lines, including AsPC-1, BxPC3, Capan-2, CFPAC-1, HPAF-II, Mia PACA-2, Pa01C, Pa02C, Pa03C, Pa07C, Pa08C, Pa16C, Pa20C, Pa21C, Pa28C, Pa222C, Panc-1, PK8, PK9, and Su86.86, and three primary cultures of stromal fibroblasts as normal controls, including CAF19, CAF25, and SC3 were maintained in DMEM (4.5 mg/ml glucose, Invitrogen) supplemented with 10% fetal bovine serum (FBS) and antibiotics (100 µg/ml streptomycin and 100 units/ml penicillin). DNA was isolated where available from selected tumor tissues by macrodissecting formalin-fixed archived paraffin-embedded tissue sections as previously described[1] and subjected to NGS. For many cases only diagnostic fine needle aspirates were available, with insufficient and/or inadequate DNA for mutation analysis.

Genomic DNA was extracted using QIAamp DNA Micro Kit (Qiagen) and quantified by Quantifiler Human DNA Quantification kit (Applied Biosystems) before NGS library preparation.

Reference mutant DNA pools

Three reference pools of cancer/normal DNA were made from the 20 pancreatic cancer cell line DNA samples. These pools were generated by mixing cancer cell line DNA with fibroblast DNA

(SC3) to achieve mutation concentrations in the range of 0.5%, 0.1%, and 0.01%. To do this DNA from 20 pancreatic cancer cell lines were mixed together and then diluted in a 1/9 ratio with SC3 DNA for the "0.5% pool", in a 1/45 ratio (i.e. 2.2% of cancer cell line DNA mixed with 97.8% of SC3 DNA to generate the "0.1% pool" and a 1/450 ratio for the 0.01% pool. (Note, since the cancer cell lines are aneuploid, these represent estimated mutation concentrations). The estimated mutation concentrations in these pools and the concentrations measured by digital-NGS are provided in Table S3.

NGS sequencing

The Ampliseq Custom panel was chosen for digital NGS because it can multiplex amplify hundreds of amplicons simultaneously from very small quantities of input DNA. For the sequencing of cell lines and tumor samples, standard Ampliseq and sequencing conditions were used according to manufacturer's protocols. Thus, 20 nanograms of DNA (10ng for each primer pool) were used for Ampliseq PCR amplification, Fupa digestion, and P1 adaptor/Xpress barcode ligation. After library clean-up using Agencourt AMPure XP Reagent (Beckman Coulter) which uses magnetic separation of amplicons, the libraries were eluted into low TE buffer and subsequently quantified using the Ion Quantitation Kit (Life Technologies). Individual sample libraries were equalized to 15pM and pooled together. After introducing 20ul pooled libraries with emulsion PCR reagents into the Ion OneTouch2 system (Life Technologies) for 5 hours, the Ion Sphere Particles (ISPs) were cleaned and enriched in the Ion OneTouch ES (enrichment system) (Life Technologies). The enriched ISPs were loaded into a 318 chip for sequencing using an Ion Torrent Personal Genome Machine (Life Technologies). The post-sequencing raw fastq files were

launched in NextGENe (SoftGenetics, Chicago, IL) software for data analysis, including alignment to the hg19 human reference genome and SNV calling. Alignments were visually verified using the Integrative Genomics Viewer (IGV, v2.3, Broad Institute) and the NextGENeViewer.

Digital NGS

Pancreatic juice DNA samples were aliquoted into two 96 well plates (one for each primer pool) with each well containing 150 pg DNA (~40 genome equivalents). Each of the 96 aliquots of DNA was subjected to Ampliseq PCR followed by the subsequent steps of the NGS protocol. The average sequence depth of each amplicon sequenced on the 318 chip was ~500 reads for each of the 96 NGS reactions. The digital NGS assay was the same in the discovery and validation set with the exception that the *RNF43* primers were modified in the validation set to improve performance. Several juice samples from both cases and controls underwent an additional round of digital NGS sequencing to confirm results. For example, the juice mutations found by digital NGS in case #37 was confirmed by repeat digital NGS, and the juice sample from case #45 underwent a 2nd round of digital NGS to confirm the lack of any mutations.

We did also evaluate the performance of digital NGS using lower concentrations of input DNA by performing digital NGS on 96 NGS reactions with 66pg of cancer cell line pool DNA (at ~0.5% cancer cell line DNA concentrations), sequenced on a 318 chip, and although the results were very similar to those achieved with the 150 picogram digital NGS experiment on the same DNA pool (i.e. all mutations were detected apart from the frameshift mutations), the detection of mutations was somewhat less uniform, with some mutations detected in only one NGS reaction,

when they would have been expected to be detected in many more, even after accounting for the lower amount of input DNA. In addition, we also performed digital NGS with higher amounts of input DNA (400 picograms) and compared these results to those results obtained using 150 piograms of DNA. This digital NGS comparison was performed on several pancreatic juice samples. Digital NGS performed using 400 picograms per NGS reaction enabled more input DNA to be sequenced, and could identify hotspot mutations in *KRAS* and *GNAS* easily, but because using 400 picograms of input DNA would mean NGS reactions with true mutations would have mutation concentrations of ~0.8 - 1.0%, this version of the assay required using a lower variant concentration cut-off for calling mutations (0.8%) resulting in an unacceptably high false positive rate (data not shown).

Sequence Variant calling

Variants identified by the NextGENe software were filtered to select only variants with a nucleotide score of $\geq$30 (reference nucleotide score +mutant nucleotide score/indel score), and no strand bias. Synonymous somatic mutations by digital NGS were not considered deleterious and were not tabulated. The potential pathogenicity of somatic mutations identified by digital NGS was evaluated using ClinVar.

Droplet digital PCR methods

Digital droplet PCR involves isolating individual DNA molecules into thousands of nanoliter-sized droplets by emulsification. Individual PCR reactions occur in each droplet. Mutations can be detected using real-time PCR using specific fluorescent probes.

The mutational status of *KRAS* codon 12 was investigated with droplet digital PCR (ddPCR) using a BioRad GX200. Primers and probes were designed for *KRAS G12D, G12V* and *G12R* (these 3 KRAS mutations are found in ~ 90% of in pancreatic ductal neoplasms with KRAS mutations[2, 3]. The MGB probes were labeled with either FAM or VIC at 5'end and a non-fluorescent quencher (NFQ) at the 3' end. Each droplet of a PCR supermix was containing 4 µL (~25,000 copies) DNA, 2×ddPCR™ Supermix for probes (No dUTP) (Bio-Rad), 900nM primers, and 250nM probes (FAM for mutant and VIC for wild type) (Life Technologies), in a total volume of 20µL. The following primers were used for the *KRAS* assay: 5'-GCCTGCTGAAAATGACTGAATATAAACT-3' (Forward) and 5'-TTGTTGGATCATATTCGTCCAC-3' (Reverse). The following probes were used for *KRAS* assay: *KRAS WT* (VIC- TTGGAGCTGGTGGCGTA-MGBNFQ) and *KRAS* G12D (FAM-TTGGAGCTGATGGCGTA-MGBNFQ), *KRAS* G12V (FAM-TTGGAGCTGTTGGCGTA-MGBNFQ), or *KRAS* G12R (FAM-TTGGAGCTCGTGGCGTA-MGBNFQ). PCR cycling parameters: 95°C for 10 minutes, 39 cycles of 94°C for 30 seconds and 63°C for 60seconds, 98°C for 10 minutes for *KRAS* G12D, G12V and G12V assays. We obtained wild type DNA for *KRAS* codon 12 and 13 from a healthy control's peripheral blood. To evaluate the limit of detection of the assay, we performed 20 replicates of a healthy subject's peripheral blood lymphocyte DNA (~15000 copies /well) by ddPCR for each probe. No positive droplets were detected with the *KRAS* G12V, G12R probes, while the *KRAS* G12D had several wild-type DNA samples that had 1 or 2

positive droplets. Samples were considered positive for mutations if they had droplets above these background levels.

Statistics:

Because Case #24 had McCune Albright syndrome, it was not included in the usual IPMN group for statistical purposes.

SUPPLEMENTAL RESULTS

Digital NGS Results:

There was a good correlation between the estimated mutation concentration in the reference cancer DNA pools and the missense mutation concentrations determined by digital-NGS (r^2=0.58, p<0.0001). However, only 2 of 9 (22%) of the frameshift mutations in the DNA reference pools were detected even when these mutations were present at ~0.5% concentrations.

False positive SNVs in 0.5% and 0.1% pools

We sequenced DNA from 20 pancreatic cancer cell lines and three primary fibroblasts (CAF19, CAF25, and SC3) with our 9-gene panel using standard Ampliseq conditions on the Ion Torrent PGM (Tables S2) to identify somatic mutations in these cancer cell lines. To estimate the false positive rate of digital NGS in DNA pools the same filters were employed as were used for the juice analysis. A false positive SNV in one NGS reaction was defined as any SNV that passed all of our filters for a mutation in an NGS reaction (i.e. how often one or more 96 individual NGS

results of a digital-NGS assay would result in calling a sequence variant because it passed all filters used to detect mutation despite the variant not being present in the reference DNA pools). Based on the background error rate of the Ion Torrent we set a threshold mutation concentration of >1.5%, (3 standard deviations above the background error rate of this platform), for digital NGS reactions to be considered sufficient to consider it a candidate mutation. We did not find false positive *KRAS* or *GNAS* mutations in 96 digital-NGS reactions performed with fibroblast DNA. For hotspot mutation nucleotides in *KRAS* and *GNAS*, one positive NGS reaction was sufficient for calling the sample as positive for that mutation. .

For other variants we required that sequence variants be detected in three multiple independent NGS reactions before it was classified as a mutation present in a juice sample. This requirement to have multiple NGS assays with the same sequence variant was based the results of sequencing the reference DNA pools (three sets of 96 NGS reactions performed on the cancer DNA pools and one set of 96 NGS reactions on a fibroblast DNA pool). Although false positive calls in one NGS reaction were often made in the cancer cell line pools, the same false positive SNV was detected in more than one well only twice, and the same false positive SNVs was not detected in 3 NGS reactions. There were no false positive mutations detected in any of the hotspot mutations in *KRAS or GNAS* in a 96 reaction digital NGS run performed on fibroblast DNA and there was a high concordance between digital NGS and ddPCR results for *KRAS* mutations. Of the top 9 most common hot spot mutations mutated in *TP53* (R175H, R248Q, R248W, R273H, R273C, R282W, Y220C, R213*, R196*, 3 of these mutations were present in the mutant cancer cell lines pools. None of the other mutations were detected in more one than one NGS reaction to indicate that false positive mutations in the *TP53* hotspot mutations are common. For this reason, we called all *KRAS* mutations identified by digital NGS as positive, and set a threshold for requiring 2 digital NGS

reactions to call is hotspot mutations in *TP53* and 3 NGS reactions for all other mutations. We also reported digital NGS scores of 1 for mutations that were also found at higher levels in other juice samples from the same patient (e.g. *SMAD4* Q256X, *TP53* R175H; Table 3).

Digital-NGS addresses the problem of false-positive mutation calls arising due to sequencing errors of NGS assays by requiring that, in addition to the usual filters used for NGS variant calling, the same sequence variant be detected at the expected mutation concentration in multiple independent NGS reactions before considering it a true mutation[4]. Polymerase errors generated during initial rounds of PCR could create sequence variants in a digital-NGS reaction at sufficient concentrations to resemble those expected from sequencing a true mutation, but since most of these sequencing errors arise randomly, the same sequencing error is unlikely to arise in more than one of the 96 digital-NGS reactions. Requiring that a variant be detected in three independent NGS reactions reduces the chance that sequencing errors will be identified incorrectly as mutations and a higher requirement for positive calls could be used to further increase assay specificity over sensitivity.

Mutation detection using digital-NGS versus digital-droplet PCR

In the discovery set, digital-NGS was performed to identify low abundance mutations on 53 pncreatic juice samples. We compared these results to those obtained using ddPCR for the 3 most common *KRAS* mutations (G12D, G12V and G12R)[5]. Of 43 *KRAS* mutations detected in 52 juice samples using digital-NGS, ddPCR detected the corresponding mutation in the corresponding sample 39 times (90.7%). The 4 mutations not detected in the corresponding juice samples by ddPCR all had very low concentrations of mutant *KRAS* (~0.05% by digital-NGS). DdPCR

detected all mutations in juice samples that had a mutation score by digital-NGS of ≥3. Digital-NGS identified all the *KRAS* mutations in the corresponding juice samples at concentrations of ≥0.1% as measured by ddPCR. Three cases, one control (case#5, G12V), and two patients with PDAC (case#41, G12V and case#51, G12V and G12R), had low mutant *KRAS* juice concentrations (0.03% to 0.07% by ddPCR) below the limit of detection of the digital-NGS assay we employed for this study.

Discovery set results

In the discovery set, 20 of 22 (90.1%) patients with PDAC and 16 of 17 (94.1%) diagnosed with IPMN (without PDAC) had mutations detected in their pancreatic juice, compared vs. 5 of 13 (38.4%) controls (both p=0.002). *KRAS* mutations (detected by digital-NGS and ddPCR) were found in juice samples of 5 of 13 controls (38.4%), 13 of 17 (76.5%) juice samples from patients with IPMNs, and 18 of 22 (81.8%) patients with PDAC. Several patients, particularly those with PDAC, had multiple *KRAS* mutations detected in their juice samples[5]. Eighteen juice samples had *TP53* mutations, including 12 from cases with PDAC, and 6 with IPMN. All but one of these *TP53* mutations is judged deleterious in the IARC *TP53* mutation database (http://p53.iarc.fr/). Seven *SMAD4* mutations were detected in six patients, 5 with pancreatic cancer, one with IPMN, three truncating and 4 missense mutations. One missense mutation (P198S) is probably benign; the others are suspected to be deleterious. W524R has been identified in a gastric cancer[6], W524L causes juvenile polyposis[7], and the A457V and M543T mutations are located in regions where missense mutations are deleterious[8-10]. Numerous *SMAD4* missense mutations cause polyposis syndromes[7, 11]. Mutations in *TP53* and *SMAD4* were not detected in the juice samples of controls

but were found in 14 of 22 (63.6%) cases with PDAC (p<0.001). Cases diagnosed with IPMN (16 of 17) were more likely to have mutations detected in their pancreatic juice than controls (p<0.002). Fourteen (56%) of the 25 cases diagnosed as having IPMN (including 8 cases with PDAC and IPMN) had *GNAS* mutations detected in their pancreatic juice samples. Of the twelve cases that had *RNF43* mutations in their pancreatic juice, 8 also had a *GNAS* mutation, and 6 arose in patients diagnosed with IPMN. Cases diagnosed with IPMN were also more likely than controls to have mutations other than *KRAS* and *GNAS* detected in their pancreatic juice samples (11 of 17 vs 1 of 13) (p=0.002). Deleterious *TP53* and/or *SMAD4* mutations were found in the pancreatic juice of 6 of 17 patients diagnosed with IPMN without associated invasive cancer. Eight patients with IPMN underwent pancreatic resection, 4 with low-grade dysplasia (3 also had PanIN-2) and 4 with intermediate-grade dysplasia in their IPMN. Three of these 8 cases that underwent resection had *TP53* and/or *SMAD4* juice mutations and had intermediate-grade dysplasia in their IPMN and/or PanIN-2 in their resection specimen; the other cases with IPMN are still under surveillance without evidence of progression one or more years after their juice sample was obtained.

In the discovery set, pancreatic juice mutation concentrations were significantly higher in juice samples from patients with PDAC compared to controls (p<0.001), as were concentrations of mutant *KRAS* alone (p<0.001) and concentrations of mutant *TP53* and/or *SMAD4* (p<0.001). By ROC curve analysis, overall digital-NGS mutation scores could distinguish PDAC cases from controls with an AUC of 0.88 (p<0.001). Pancreatic juice concentrations of mutated *TP53* and/or *SMAD4* were higher among cases with PDAC than those with IPMN (Mann-Whitney, p=0.026). By ROC curve analysis, digital-NGS scores for mutant *TP53* and/or *SMAD4* could distinguish PDAC cases from IPMN cases without PDAC with 50% sensitivity and 82% specificity (AUC 0.71, p=0.0028), and from controls with an AUC of 0.79 (p=0.002). Among PDAC cases with

TP53 and/or *SMAD4* mutations, 7 of 14 had digital-NGS scores of ≥5 compared to 0 of 6 with IPMN (p=0.03). By ROC analysis, overall digital-NGS scores could also distinguish IPMN cases from controls with an AUC of 0.83 (p=0.001) and digital-NGS scores for mutant *TP53* and/or *SMAD4* could distinguish IPMN cases from controls with an AUC of 0.72 (p=0.046). Only one patient with an IPMN (case#20) had a *BRAF* mutation detected in their juice sample and one (case#53) had a *PIK3CA* mutation. Only deleterious variants were included in the analysis, tables and figures. One juice sample in the discovery set had a *TP53* mutation detected in their juice sample that is not in the IARC database (M340V, from a patient with an IPMN). All other *TP53* mutations have been reported to be deleterious in the IARC *TP53* mutation database (http://p53.iarc.fr/). One SMAD4 missense mutation (P198S) was found and judged to be benign.

Validation set results

In the validation set there were 62 cases (#54-#115), 11 cases with pancreatic cancer, 11 normal pancreas controls and 40 with IPMN. The results of the validation set were very similar to the discovery set and are summarized in Table 2, Figures S1, S2 and S3.

Further description of combined set results

SMAD4 missense mutations: W524R has been identified in a gastric cancer[6], W524L causes juvenile polyposis[7], and the A457V and M543T mutations are located in regions where missense mutations are deleterious[8-10].

The percentage of cases with IPMN that had *GNAS* mutations in their pancreatic juice in this study (53%) is similar to what we reported recently (51%)[5]. In our first report describing *GNAS* mutations[12], we reported that 64% of patients with pancreatic cysts diagnosed with IPMN had *GNAS* mutations detected in their pancreatic fluid of patients, but we also reported a lower percentage of patients with diminutive cysts had *GNAS* mutations (45%). At that time we considered diminutive cysts separately from larger cysts because we were not certain what the etiology of these diminutive cysts were. We now know that most of these diminutive cysts have GNAS mutations and are IPMNs or incipient IPMNs[13] so we no longer make this distinction.

REFERENCES

1. Kanda M, Sadakari Y, Borges M, et al. Mutant TP53 in Duodenal Samples of Pancreatic Juice From Patients With Pancreatic Cancer or High-Grade Dysplasia. Clin Gastroenterol Hepatol 2013;11:719-30.
2. Amato E, Molin MD, Mafficini A, et al. Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas. J Pathol 2014;233:217-27.
3. Kanda M, Matthaei H, Wu J, et al. Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. Gastroenterology 2012;142:730-733.e9.
4. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012;13:341.
5. Eshleman JR, Norris AL, Sadakari Y, et al. KRAS and guanine nucleotide-binding protein mutations in pancreatic juice collected from the duodenum of patients at high risk for neoplasia undergoing endoscopic ultrasound. Clin Gastroenterol Hepatol 2015;13:963-9 e4.
6. http://www.cbioportal.org/case.do?cancer_study_id=stad_tcga&case_id=TCGA-BR-8366.
7. http://www.arup.utah.edu/database/smad4/SMAD4_display.php.
8. Fleming NI, Jorissen RN, Mouradov D, et al. SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. Cancer Res 2013;73:725-35.
9. Hahn SA, Schutte M, Hoque AT, et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 1996;271:350-3.
10. De Bosscher K, Hill CS, Nicolas FJ. Molecular and functional consequences of Smad4 C-terminal missense mutations in colorectal tumour cells. Biochem J 2004;379:209-16.

11. Carr JC, Dahdaleh FS, Wang D, et al. Germline mutations in SMAD4 disrupt bone morphogenetic protein signaling. J Surg Res 2012;174:211-4.
12. Kanda M, Knight S, Topazian M, et al. Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts. Gut 2013;62:1024-33.
13. Basturk O, Hong SM, Wood LD, et al. A Revised Classification System and Recommendations From the Baltimore Consensus Meeting for Neoplastic Precursor Lesions in the Pancreas. Am J Surg Pathol 2015;39:1730-41.