Bioinformatic data analysis for NGS testing

After a successful sequencing reaction, the raw signal data were analyzed using Torrent Suite version 3.4.2. The pipeline includes signaling processing, base calling, quality score assignment, adapter trimming, PCR duplicate removal, read alignment to human genome 19 reference, quality control of mapping quality, coverage analysis, and variant calling. After completion of the primary data analysis, a list of detected sequence variants [single nucleotide variants (SNVs) and insertions or deletions (indels)] were compiled in a variant call file format and presented via the web-based user interface. Variant calls were further analyzed using an internally created customized software suite (SeqReporter; University of Pittsburgh Medical Center) that allowed variant filtering and annotation using COSMIC version 64, The Single Nucleotide Polymorphism Database (dbSNP) dbSNP build 137, the Database for Nonsynonymous SNPs' Functional Predictions (dbNSFP) light version 1.3, PolyPhen-2, SIFT, the University of California, Santa Cruz, genome browser and cBioportal, Memorial Sloan Kettering Cancer Center, IARC (WHO) TP53 mutation database, International Cancer Genome Consortium (ICGC) and My Cancer Genome. Amino acid predictions were carried out with in silico prediction algorithms SIFT[1, 2] and PolyPhen-2[3] to predict potential deleterious effect on protein function. Interpretation and reporting of clinical NGS detected somatic variants were performed in accordance with the *Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology and College of American Pathologists*.[4] Genetic alterations were reported if they consisted of variants of strong clinical significance (tier I) or variants of potential clinical significance (tier II); however, variants of unknown significance (tier III) and variants of known insignificance (tier IV) were not reported.[4]

Sanger sequencing for KRAS and GNAS

Between January 2013 and January 2014, 175 PCF specimens obtained by EUS-FNA were prospectively submitted for *KRAS* and *GNAS* Sanger sequencing as part of clinical care and within a 14-day turnaround within the CLIA-certified and CAP-accredited Molecular and & Genomic Pathology Laboratory at UPMC as described previously.[5] Genomic DNA was isolated from pancreatic cyst fluid obtained by EUS-FNA using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche, Indianapolis, IN) on Compact MagNA Pure (Roche, Indianapolis, IN). Extracted DNA was quantitated on the Qubit 2.0 Fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA). Among the 175 cases, 159 specimens were sufficient for molecular testing. PCF DNA was amplified with primers flanking *KRAS* exon 2 (5′-GGTGAGTTTGTATTAAAAGGTACTGG-3′ and 5′-TCCTGCACCAGTAATATGCA-3′) and exon 3 (5′-TGAAGTAAAAGGTGCACTG-3′ and 5′-GCATGGCATTAGCAAAGACTC-3′). Similarly, the detection of *GNAS* mutations was performed using primers flanking codon 201 at exon 8 (5′-TGACTATGTGCCGAGCGA-3′ and 5′-AACCATGATCTCTGTTATATAA-3′). PCR products were sequenced in both sense and antisense directions using BigDye Terminator v3.1 cycle sequencing kit on the ABI 3730. The sequence electropherograms were analyzed using Mutation Surveyor software. The limit of detection was approximately 10-20% of mutant alleles. Medical records were reviewed to document patient demographics, clinical presentation, EUS findings, fluid viscosity, CEA analysis and cytopathologic diagnoses. Pathology slides were also reviewed for each surgical specimen.

Additional patient cohort information

The clinical, pathologic and molecular findings of the study population are summarized in Tables 1 and 2. Patients ranged in age from 15 to 93 years (mean, 65.0 years) with a slight female predominance (341 of 595, 57%). All pancreatic cysts were initially identified by abdominal ultrasound (US), CT or MRI. For 198 (33%) patients, the indications for abdominal imaging included epigastric/abdominal pain (n = 106, 18%), pancreatitis (n = 75, 13%), weight loss (n = 11, 2%) and/or jaundice (n = 6, 1%). The remaining 397 (67%) patients were asymptomatic, and their pancreatic cyst(s) was detected incidentally.

References

1 Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012;**40**:W452-7.

2 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;**4**:1073-81.

3 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, *et al.* A method and server for predicting damaging missense mutations. Nat Methods 2010;**7**:248-9.

4 Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, *et al.* Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn 2017;**19**:4-23.

5 Singhi AD, Nikiforova MN, Fasanella KE, McGrath KM, Pai RK, Ohori NP, *et al.* Preoperative GNAS and KRAS testing in the diagnosis of pancreatic mucinous cysts. Clin Cancer Res 2014;**20**:4381-9.