Discovery Screen

(Japan, 38 cases and USA, 32 cases)
Ampullary Carcinoma Tissue (n = 60) Paired Normal Tissue (n = 60)
(and Duodenal Carcinoma Tissue, n = 10)

DNA

Whole Exome Sequencing

(Agilent SureSelect Human All Exon v4,
Illumina HiSeq2000, 188X mean coverage)

Array CGH Analysis

(Agilent SurePrint G3 CGH Microarray)

Primary Sequence Analysis

- Alignment to reference human genome (GRCh37)
- Remove probable PCR duplication

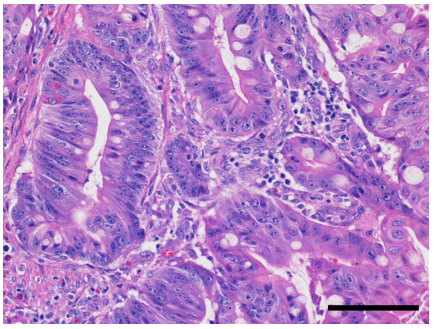
Identification of Somatic Copy Number Alterations

(GISTIC2.0)

Somatic Mutation Calling

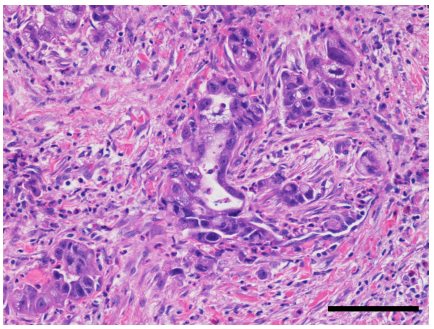
Subtype Classification

(Immunohistochemistry for CK20,
CDX2, MUC2, and MUC1)



< Intestinal-type >

Intestinal-type ampullary carcinoma showing columnar tumor cells with elongated cigar-shaped nuclei and nuclear stratification. Goblet cells are interspersed with the columnar cells.



< Pancreatobiliary-type >

Pancreatobiliary-type ampullary carcinoma showing simple tubular glands with cuboidal to low columnar cells and a single layer of round centrally placed nuclei with an abundant desmoplastic stroma. Scale bars = 100 μm.

Downstream Analysis

- Functional studies of *ELF3*
- Multi-region sequencing using a new technology (Glass Chip Macrodissection)
  → (Whole exome sequencing in 30 region)

Validation Screen; Targeted-Sequencing

Validation set including discovery set (n = 190)
to validate significantly mutated genes
(Agilent HaloPlex Custom Capture Kit,
Illumina HiSeq2500, 2,535X mean coverage)

Figure S1, related to Table 1. The flowchart of the entire analysis in the present ampullary carcinoma genomic study.

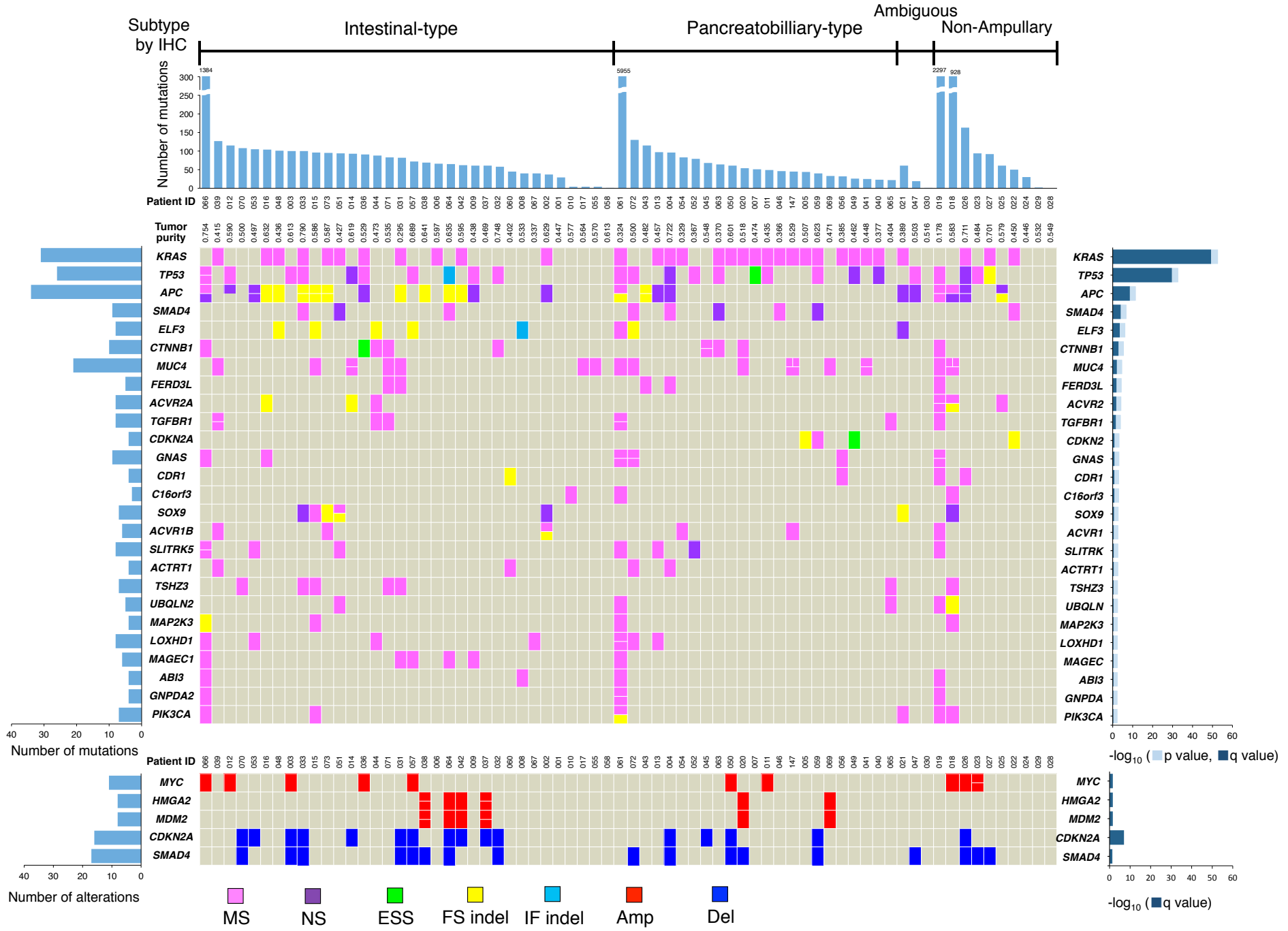

Subtype by IHC
Intestinal-type
Pancreatobilliary-type
Ambiguous
Non-Ampullary
Number of mutations
Patient ID
Tumor purity
KRAS
TP53
APC
SMAD4
ELF3
CTNNB1
MUC4
FERD3L
ACVR2A
TGFBR1
CDKN2A
GNAS
CDR1
C16orf3
SOX9
ACVR1B
SLITRK5
ACTRT1
TSHZ3
UBQLN2
MAP2K3
LOXHD1
MAGEC1
ABI3
GNPDA2
PIK3CA
Number of mutations
-log10 (p value, q value)
MYC
HMGA2
MDM2
CDKN2A
SMAD4
Number of alterations
-log10 (q value)
MS
NS
ESS
FS indel
IF indel
Amp
Del

Figure S2, related to Figure 1. OncoPlot summary of significantly ($p < 0.005$) mutated genes in 60 ampullary carcinoma and 10 duodenal carcinoma samples (discovery screen).

Ampullary carcinomas are immunohistchemically classified into intestinal-type and pancreatobiliary-type (or ambiguous-type). The top bar plot shows the number of somatic mutations from each tumor. The bottom middle plot shows amplified or deleted genes for each tumor based on copy number analysis (GISTIC2.0 algorithm). The bottom left plot shows the mutation count for each individual gene. The bottom right bar plot shows the significance of each gene: $-\log_{10}$ (p) values are shown in light blue, and $-\log_{10}$ (q) values in dark blue. MS, missense mutation; NS, nonsense mutation; ESS, essential splice-site mutation (the first or last 2 bp of an intron); FS indel, frameshift insertion or deletion; IF indel, in-frame insertion or deletion; Amp, amplification; Del, deletion.

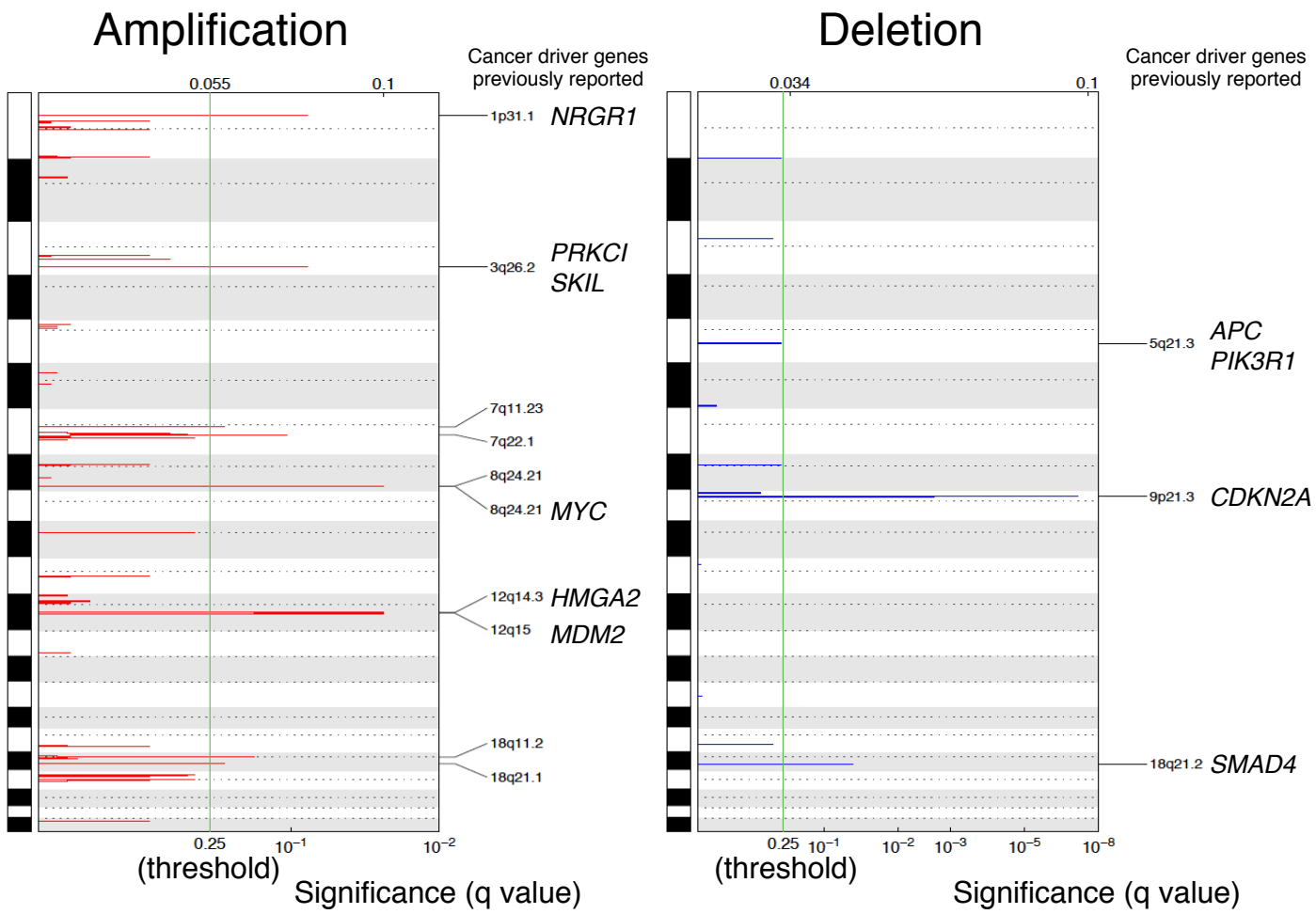
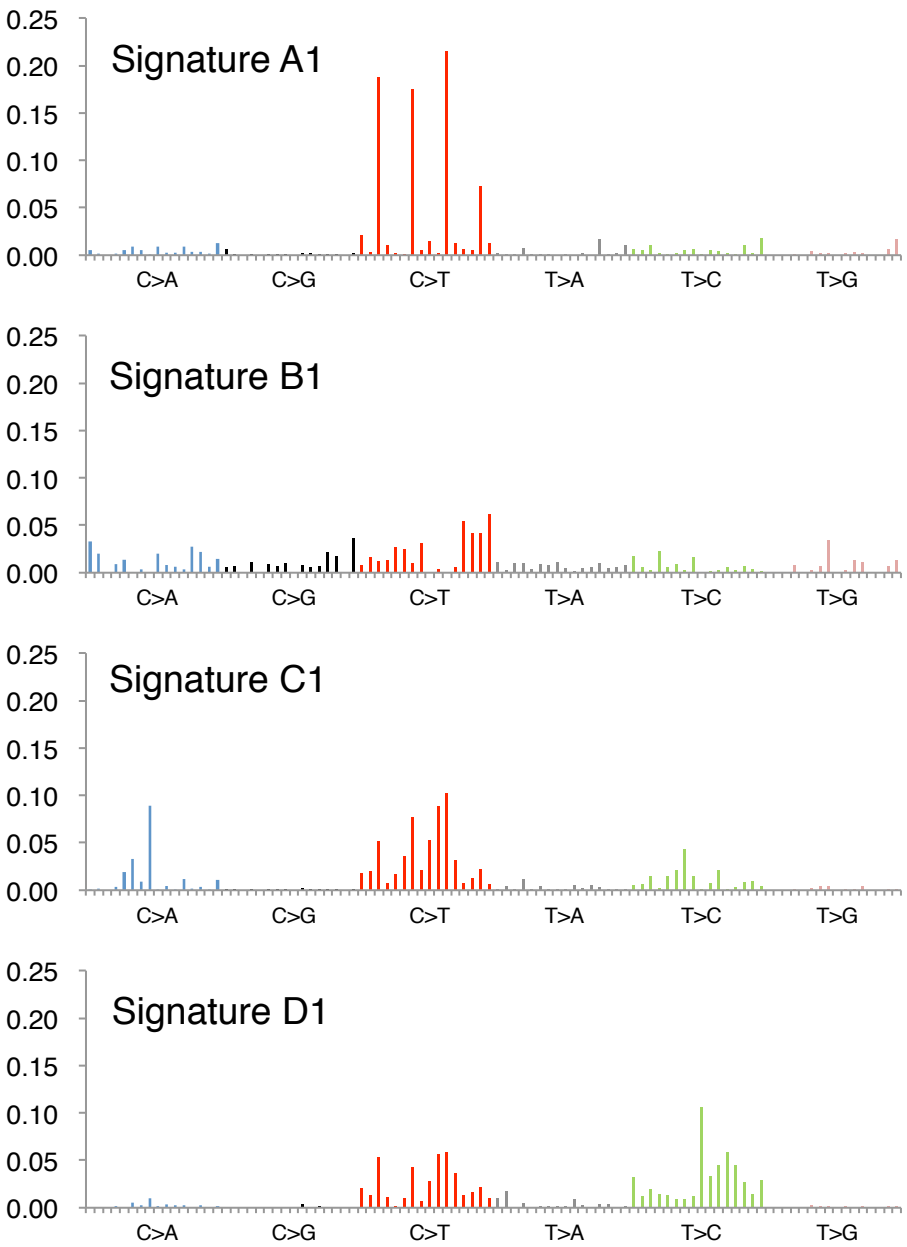

Amplification
Cancer driver genes previously reported
0.055
0.1
1p31.1
NRGR1
3q26.2
PRKCI
SKIL
7q11.23
7q22.1
8q24.21
8q24.21
MYC
12q14.3
HMGA2
12q15
MDM2
18q11.2
18q21.1
0.25
10^{-1}
10^{-2}
(threshold)
Significance (q value)
Deletion
Cancer driver genes previously reported
0.034
0.1
5q21.3
APC
PIK3R1
9p21.3
CDKN2A
18q21.2
SMAD4
0.25
10^{-1}
10^{-2}
10^{-3}
10^{-5}
10^{-8}
(threshold)
Significance (q value)

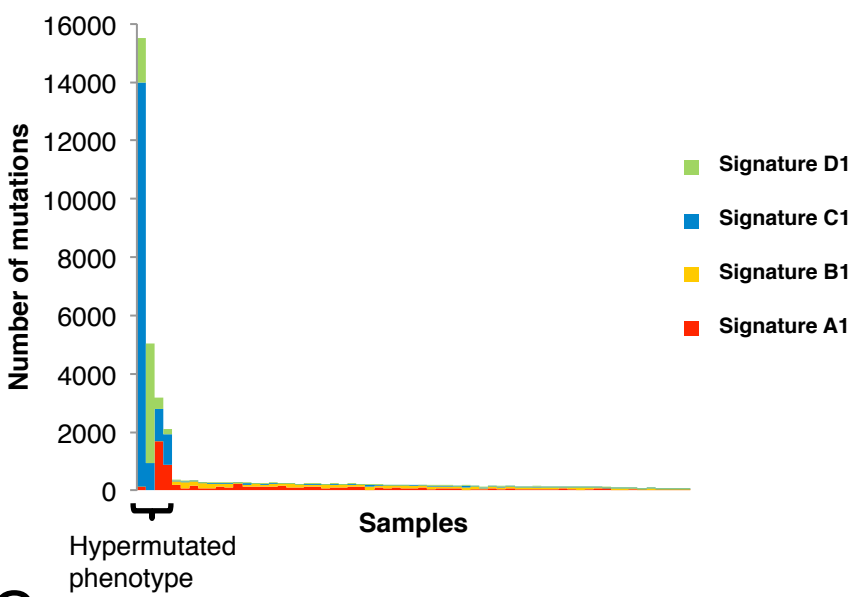
Figure S3, related to Figure 2. Significant broad and focal copy number alterations in the genome of 58 ampullary carcinomas and 10 duodenal adenocarcinomas.

GISTIC analysis of copy number changes in 58 ampullary carcinomas and 10 duodenal adenocarcinomas is shown. The statistical significance of the aberrations is displayed as FDR q values to account for multi-hypothesis testing. Chromosome positions are indicated along *y* axis with centromere positions indicated by dotted lines. Ten focal events (indicated by red bars for amplifications and blue bars for deletions) surpass the significant threshold (green line). The locations of the peak regions and the known cancer-related genes within those peaks are indicated to the right of each panel.

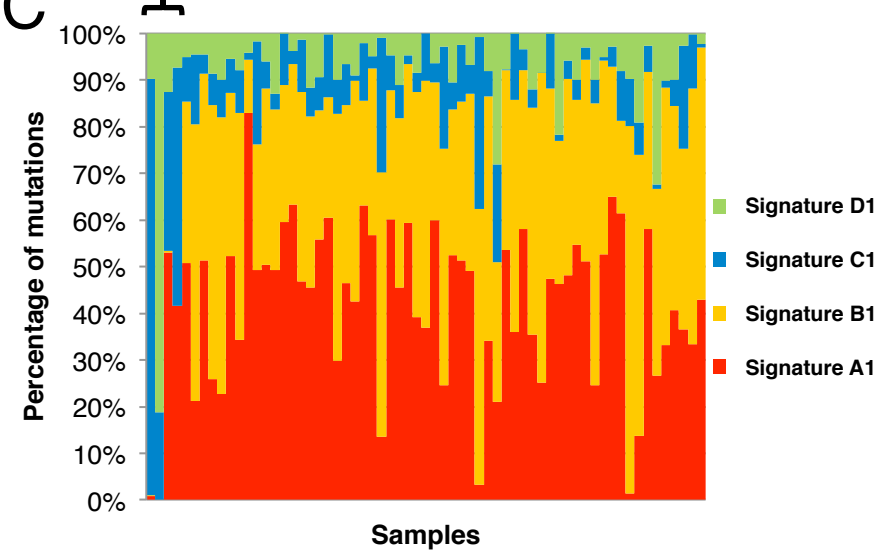
A



B



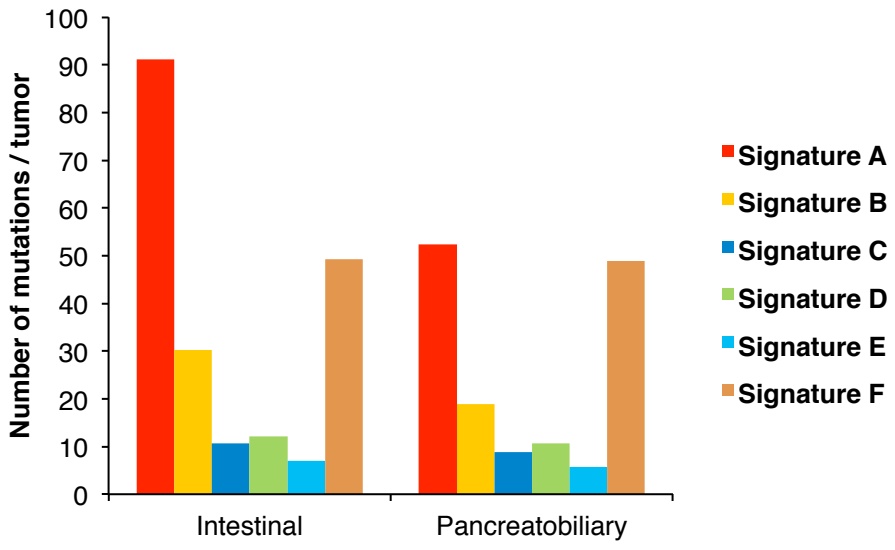
C



D

Signature	Probable association, p < 0.05 (* < 0.01)
A	Age*, Tumor size , Intestinal-type*
B	Intestinal-type
C	
D	Age*
E	
F	

E



F

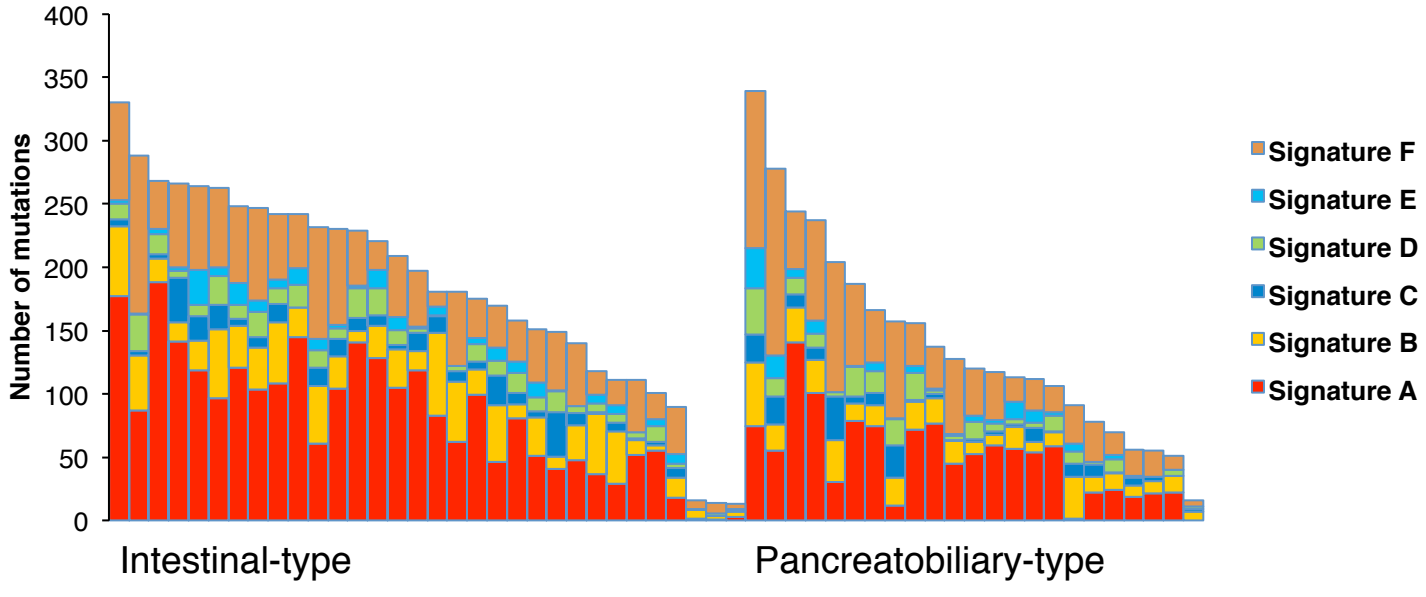


Figure S4, related to Figure 3.

(A) Four signatures were independently extracted from all samples by NMF.

(B) Distribution of four mutation signatures sorted by the total mutation number. Hypermutated cases are separated on the left.

(C) Distribution of four mutational signatures sorted by the percentage of four mutational signatures.

(D) The summary of probable association with each signature among clinicopathological factors (age, gender, nationality, tumor size and histological phenotypes).

(E) The comparison of number of mutations per tumor between intestinal- and pancreatobiliary-type tumors.

(F) The distribution of six mutational signatures to each tumor, sorted by the number of mutations in each histological phenotype.

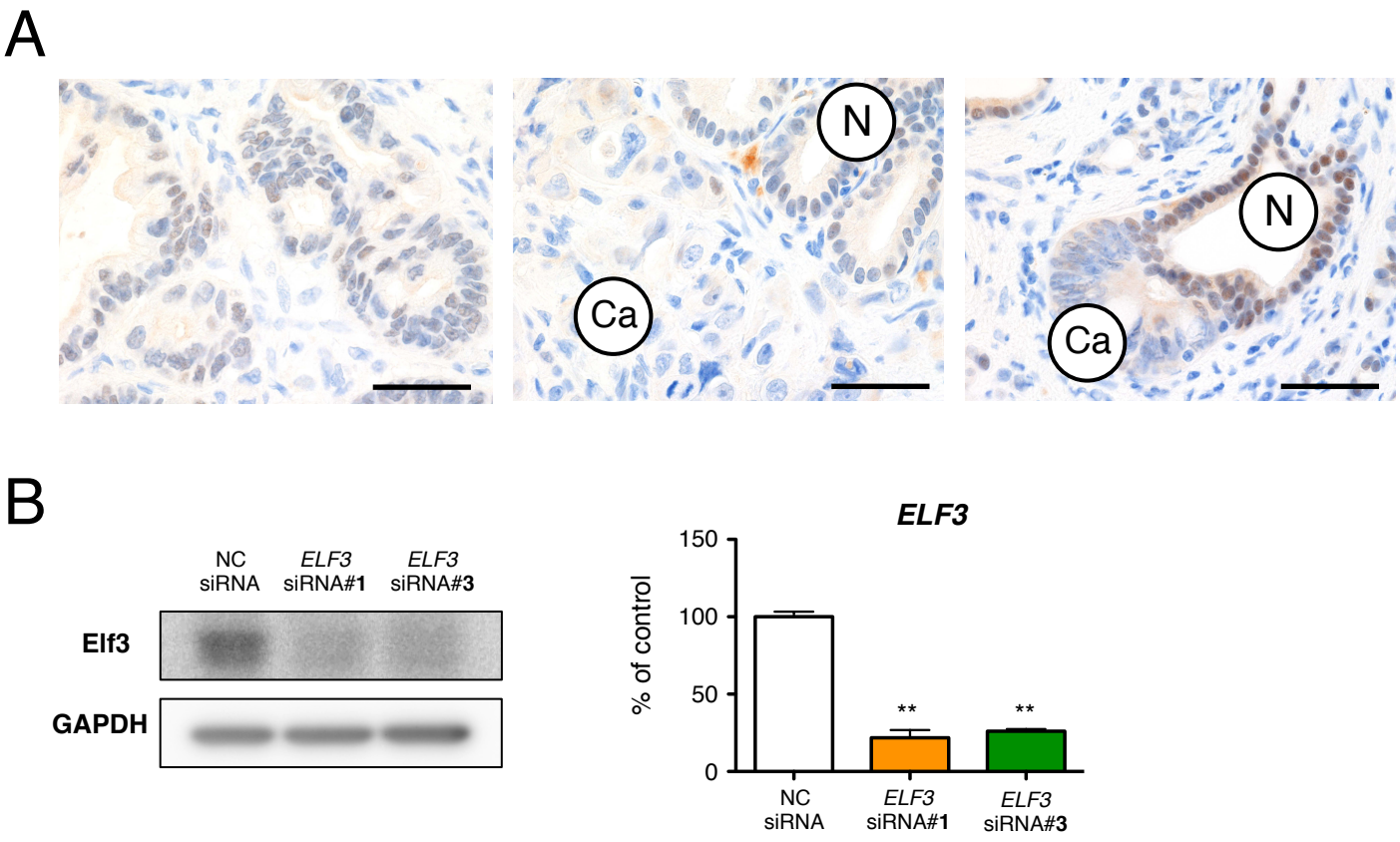

A
N
Ca
N
Ca
B
NC siRNA
ELF3 siRNA#1
ELF3 siRNA#3
Elf3
GAPDH
ELF3
% of control
150
100
50
0
**
**
NC siRNA
ELF3 siRNA#1
ELF3 siRNA#3
C
Proliferation
Absorbance (OD in nm)
0.8
0.6
0.4
0.2
0
NC siRNA
ELF3 siRNA#1
ELF3 siRNA#3
1
2
3
4
Days after plating (days)
Invasion
Number of cell invasion
200
150
100
50
0
NC siRNA
ELF3 siRNA#1
ELF3 siRNA#3
**
*
##
24 hr
48 hr
Migration
Number of cell migration
800
600
400
200
0
NC siRNA
ELF3 siRNA#1
ELF3 siRNA#3
**
##
24 hr
48 hr

Figure S5, related to Figure 4.

(A) Immunohistochemical labeling patterns for Elf3. **(Left)** Example of positive nuclear labeling for Elf3 in an ampullary carcinoma (no *ELF3* mutations). **(Middle)** Loss of nuclear labeling of Elf3 in an ampullary carcinoma (indicated by ***Ca***) with *ELF3* mutation. Normal accessory digestive glands of the ampulla of Vater are positive (indicated by ***N***). **(Right)** Loss of nuclear labeling of Elf3 in another ampullary carcinoma with *ELF3* mutation, showing neoplastic cells (indicated by ***Ca***) proliferate as though they are replacing normal accessory glands of the ampulla of Vater (indicated by ***N***) at a boundary between them. Scale bars = 100 μm.

(B) Representative western blot for the expression of Elf3 in HDuodEC3 cells treated with the specific siRNA (*ELF3* siRNA #1 and #3) and negative control siRNA (NC siRNA). Quantitative RT-PCR analysis for the expression of *ELF3* in control and *ELF3* knockdown cells. Total RNA was prepared 72 hr after transfection (mean ± SEM, n = 4 per group, $^{**}p < 0.01$ *versus* NC siRNA).

(C) Cellular proliferation and invasion/migration assays with control and *ELF3* knockdown cells. Cell growth was measured for 1, 2, 3 and 4 days and performed in triplicate (mean ± SEM). Cell invasion and migration were measured for 24 and 48 hr and performed in three times (mean ± SD, $^{**}p < 0.01$ *versus* 24 hr NC siRNA, $^{*}p < 0.05$ *versus* 24 hr NC siRNA, $^{\#\#}p < 0.01$ *versus* 48 hr NC siRNA).

Table S1, related to Table 1. Clinicopathologic features of ampullary carcinomas. Provided as an Excel file.

Table S2, related Figure 1. No-synonymous mutations in the discovery screen (n = 70). Provided as an Excel file.

Table S3, related Table 2. Significantly mutated genes of ampullary carcinomas in the discovery screen (n = 60). Provided as an Excel file.

Table S4, related to Figure 4. Adjusted variant allele frequencies for *ELF3* mutations by comparison to tumor percent/average variant allele frequencies in exome-sequenced carcinomas (discovery screen). Provided as an Excel file.

Table S5, related Figure 1. Significantly mutated genes of ampullary carcinomas in the validation screen (n = 172). Provided as an Excel file.

Table S6, related to Table 3. Significantly mutated genes in intestinal-type ampullary carcinomas (n = 93), pancreatobiliary-type ampullary carcinomas (n = 66) and non-ampullary duodenal carcinomas (n =18). Provided as an Excel file.

Table S7, related to Figure 4. Clinicopathologic features in patients with *ELF3* mutations. Provided as an Excel file.

Table S8, related to Figure 4. Relationship of *ELF3* mutations to immunohistochemical labeling for Elf3 protein. Provided as an Excel file.

Table S9, related to Figure 2. Summary of mutations in potential therapy target genes. Provided as an Excel file.

Table S10, related to Figure 1. Comparison of significantly mutated genes between Japanese and

American patients with ampullary carcinomas. Provided as an Excel file.

Table S11, related to Figure 5. Non-synonymous mutaions of multi-region exome sequencing.

Provided as an Excel file.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Patients and Tissue Samples

The tissues and clinical information used in this study were obtained under informed consent and approval of the institutional review boards of each institute. Samples used were retrospectively and prospectively acquired and restricted to primary, non-pretreated ampullary carcinomas and duodenal carcinomas. Frozen tissue samples of ampullary carcinoma and duodenal carcinoma, and paired normal tissues were obtained from individuals who underwent pancreatoduodenectomy (the Whipple procedure) in each institute. When a carcinoma is located in the duodenum and only extends peripherally to involve the ampulla, it is classified as a primary of the duodenum. Snap frozen tissue samples were embedded in OCT compound (Sakura Finetek), sectioned by a cryostat and stained by hematoxylin and eosin. We performed macrodissection to enrich the tumor fraction relative to the dominant stromal component and other normal cells.

DNA Preparation, DNA Capture and Sequencing

gDNA was extracted using QIAamp DNA Mini kit (Qiagen). One microgram of DNA per sample was sheared with a Covaris SS Ultrasonicator. Exome capture was performed with Agilent SureSelect Human All Exon Kit v4.0 (Agilent Technologies). Each sample was sequenced on an Illumina Hiseq 2000 instrument using a read length of 2 x 100 bp.

Mutation Calling

Paired-end reads were aligned to the human reference genome (GRCh37) using the Burrows-Wheeler Aligner (BWA) for both tumor and normal samples. Probable PCR duplications, for which paired-end reads aligned to the same genomic position, were removed, and pileup files generated using SAM tools (Li et al., 2009) and a program developed in house (Totoki et al., 2014). To find somatic point mutations and short indels, stringent confidence filtering conditions were applied. The details of our filtering conditions are reported previously (Totoki et al., 2014).

Processing the Significantly Mutated Genes

Significantly mutated genes were estimated by aggregating somatic substitutions and short indels. First, the expected number of each type of mutation in each gene was estimated as follows. The substitution rate was

estimated by dividing the number of synonymous mutations by the number of synonymous sites in the genome. For each gene, the expected number was calculated by multiplying the substitution rate by the number of non-synonymous sites and splice-sites in a gene. Since the substitution rate in CpG sites was much higher than in other regions, the substitution rates and the expected numbers in CpG- and non-CpG sites were estimated separately as follows.

n: The number of samples

M_{CG_i}: The number of synonymous mutations in CpG sites in *i*-th sample

M_{NCG_i}: The number of synonymous mutations in non-CpG sites in *i*-th sample

S_{CG}: The number of synonymous sites in CpG sites in the genome

S_{NCG}: The number of synonymous sites in non-CpG sites in the genome

N_{CG}: The number of non-synonymous sites and splice-sites in CpG sites in a gene

N_{NCG}: The number of non-synonymous sites and splices-sites in non-CpG sites in a gene

C_i: The fraction of sequence coverage in the genome in *i*-th sample (usually the fraction of coding regions which have more than 20X sequence depth for whole exome sequencing)

EN: The expected number of nonsynonymous and splice-site substitutions in a gene

$$EN = \sum_{i=1}^{n} \left(\frac{M_{CG_i} \times N_{CG}}{S_{CG} \times C_i} + \frac{M_{NCG_i} \times N_{NCG}}{S_{NCG} \times C_i} \right)$$

Coding indel rate was estimated by dividing the number of coding indels by the number of coding sites in the genome. For each gene, the expected number was calculated by multiplying the coding indel rate by the coding length in a gene as follows.

I_i: The number of coding indels in *i*-th sample

S: The number of coding sites in the genome

L: The coding length in a gene

EI: The expected number of coding indels in a gene

$$EI = \sum_{i=1}^{n} \frac{I_i \times L}{S \times C_i}$$

The expected number of protein-altering mutations was calculated by aggregating the expected number of nonsynonymous and splice-site substitutions in CpG and non-CpG site and coding indels as follows.

E: The expected number of protein-altering mutations in a gene

$$E = EN + EI$$

Tests of significance of each gene were performed by assuming a Poisson distribution. The adjustment by multiple testing was performed using the Benjamin and Hochberg's method (Benjamini and Hochberg, 1995).

Concurrence and Mutual Exclusion Analysis

We searched for genes of which the mutations concurrently or mutual-exclusively occurred with *ELF3* mutations. We used genes with q values < 0.1 (24 genes) and genes with 4 or more identical mutations (2 genes) as candidate genes. For each candidate gene, we performed a permutation test, where we permutated the states of presence and absence of mutations across all samples. We tabulated the numbers of samples with concurrent and exclusive mutations (Xc and Xe, respectively) against the observed mutations of *ELF3*. We repeated this randomized process for 10,000 times. We also tabulated these numbers (Nc and Ne) based on observed mutations for the candidate gene. We calculated the frequencies of Xc ≥ Nc and Xe ≥ Ne, respectively, and used them as empirical p values for each candidate gene.

Mutation Pattern Analysis

Cases with small number of mutations cannot represent accurate frequency of mutation patterns, so cases with mutations less than 40 were excluded for further mutation pattern analysis. The number of somatic 96-substitution types, C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G and T>G/A>C with the bases immediately 5' and 3' to each substitution in coding region, was counted for each sample. Dividing by the total substitution number, the frequencies were used for principal-component analysis (PCA). PCA was implemented using the R command prcomp with the scaling option on. The Wilks's test was used to evaluate significant mean vector differences between the two groups. Non-negative matrix factorization (NMF) was applied to the 96-substitution pattern using published software. Mutational data of colorectal carcinomas and pancreatic carcinomas were obtained from mutational catalogues of ICGC. After excluding hypermutated phenotypes, cases with mutations less than 40 and cases with mutations analyzed by whole genome sequencing, mutational data of 44 ampullary carcinomas (Intestinal-type, n = 29; pancreatobiliary-type, n = 22), 22 pancreatic carcinomas and 422 colorectal carcinomas were mixed and were re-analyzed together. We ran 1,000 iterations of NMF and each NMF iterated until convergence (10,000 iterations without change), or until the maximum number of 1,000,000 iterations was reached. Model selection of NMF for mutational signatures were performed as we previously reported (Totoki et al., 2014).

Gene Selection for Targeted Sequencing

To accurately evaluate the frequency and distribution of somatic mutations in the validation screen, targeted sequencing was performed with deeper read coverage. Ninety-two genes were selected that included recurrent mutated genes, especially significantly ($q < 0.1$) mutated genes and drug-targetable genes identified in the discovery screen, and driver genes identified in previous reports on colorectal cancer (Cancer Genome Atlas, 2012), bile duct cancer (Chan-On et al., 2013; Jiao et al., 2013; Simbolo et al., 2014) and pancreatic cancer (Waddell et al., 2015) exome studies (see below).

ACTRT1	*CMYA5*	*FLG*	*MN1*	*RNF43*
ACVR1B	*CNTN4*	*FLRT2*	*MSH2*	*ROBO1*
ACVR2A	*CTNNB1*	*GNAS*	*MSH3*	*ROBO2*
APC	*CTNND2*	*HIF1A*	*MSH6*	*SCN3A*
APOB	*DICER1*	*HOOK3*	*MYCN*	*SLITRK5*
ARID1A	*DIS3*	*IDH1*	*NF1*	*SMAD4*
ARID2	*DPP10*	*IDH2*	*NRAS*	*SOX9*
ASXL1	*DPYSL4*	*JAK3*	*PALB2*	*STK11*
ATM	*EBF3*	*KDM5C*	*PBRM1*	*TCF7L2*
BAP1	*ELF3*	*KMT2B*	*PIK3C2A*	*TGFBR1*
BRAF	*EPHA3*	*KRAS*	*PIK3C2G*	*TGFBR2*
BRCA1	*EPHA5*	*LOXHD1*	*PIK3CA*	*TP53*
BRCA2	*EPHA6*	*LRP1B*	*PMS1*	*TRIO*
CARD11	*ERBB2*	*MAGEC1*	*PMS2*	*TRRAP*
CDC42BPB	*ERBB3*	*MAP2K7*	*POLE*	*TSHZ3*
CDH10	*ERBB4*	*MLH1*	*PTEN*	*U2AF1*
CDKN2A	*FAM123B*	*MLH3*	*PTPRT*	
CLTC	*FBXW7*	*MLL2*	*RAB11FIP5*	
CLTCL1	*FERD3L*	*MLL3*	*RBM10*	

Mutation Confirmation Using Targeted Deep Sequencing

To validate somatic mutations in selected 92 genes, we determined the sequences using a target enrichment system (HaloPlex, Agilent Technologies) that differed from that in the exome sequencing. The expected coverage of the coding region based on the amplicon design was 99.6%. The HaloPlex target enrichment system relies on a tailor cocktail of restriction enzymes and customized probes to capture genomic regions of interest. We used a Bravo Automated Liquid Handling Platform (Agilent Technologies) for automated library construction. All libraries of targeted-enriched DNA were analyzed on a 2200 TapeStation (Agilent Technologies) to verify successful enrichment. All samples were sequenced on the Illimina HiSeq2500 platform with paired-end 150 bp reads according to the manufacture's instruction.

Copy Number Analysis

We profiled for somatic copy-number alterations (SCNAs) with Agilent CGH array (SurePrint G3 CGH Microarray, 1x1M), using their matched non-tumor tissues as a copy number reference. We applied the GISTIC2.0 algorithm to identify SCNAs that might be responsible for driver tumorigenesis.

Sanger Sequencing of the *ELF3* Gene

PCR amplification was carried out using 20 ng of gDNA for *ELF3* exons 1 to 8 using intronic primers flanking these exons (see below). PCR products were sequenced by use of a M13F primer (5'-GTAAAACGACGGCCAGT-3') or M13R primer (5'-CAGGAAACAGCTATGACC-3') that was incorporated into the forward and reverse primer of each primer pair, respectively. Sequencing data were analyzed with Sequencher 5.0.1 software (Gene Codes). Mutation analysis, confirmation and determination of somatic status were carried out using matched normal tissues from the same patient.

Exon	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
1	GGGAGTGTAAGGAGAGGACCC	M13R-CTGGAATTTGCCTAGAGACCC
2	CCAGCCTAGGTGACAGGAGTG	M13R-CCTACGGCCACACTGAACTC
3	GCTGAGTCGAGTTCAGTGTGG	M13R-TTGAGGGAGGAAGAAGTCTGG
4	M13F-AATTGCAGCAGGTCATCAGAC	CTGGCTCTCAGGACACACTTC
5	TCTGGCAGGAACAGGAACAG	M13R-GGGAGTTAGGGAGAGTAGCCC
6	M13F-GTCCCATTTAGCAATGCACAG	GAAGGGATACCTGCAACAACC

7	M13F-CTCTGGAGAGGCTTGCTGC	CTAAGGACTCAGCCCTGTTGC
8	M13F-AGGCTCAGCTTAGTCAGGCAG	CAGCTTCTCTCCACAGCACAG

Immunohistochemistry

Paraffin-embedded samples of the primary carcinomas from 172 patients were immunostained for MUC1, MUC2, CDX2 and Cytokeratin 20 (CK20). Immunohistochemical labeling was carried out using a Bond Max instrument (Leica Microsystems) as previously described (Oshima et al., 2013). An anti-human MUC1 mouse monoclonal antibody (clone Ma552, diluted 1:150, Leica Microsystems), an anti-human MUC2 mouse monoclonal antibody (clone Ccp58, diluted 1:300, Leica Microsystems), an anti-human CDX2 mouse monoclonal antibody (clone AMT28, diluted 1:50, Leica Microsystems), and an anti-human CK20 mouse monoclonal antibody (clone KS20.8, diluted 1:75, Leica Microsystems) were used. Normal acinar cells in each case served as an internal control for positive MUC1 immunolabeling. Normal intestinal mucosal cells in each case were served as internal controls for positive MUC2, CDX2 and CK20 immunolabeling. Immunohistochemical labeling for Elf3 was performed in a subset of *ELF3* mutant and *ELF3* wild-type ampullary carcinomas (n = 12 for each group) (**Figure S5** and **Table S8**). An anti-human Elf3 rabbit polyclonal antibody (HPA003316, diluted 1:30, Sigma-Aldrich) was used. Normal accessory digestive glands of the ampulla of Vater in each case served as an internal control for positive Elf3 immunolabeling. Negative controls for each of the antibodies were included using non-immune serum instead of the primary antibodies.

Classification of Subtypes Based on the Immunohistochemistry

Morphologically, intestinal-type ampullary carcinoma shows columnar tumor cells with elongated cigar-shaped nuclei and nuclear stratification (see **Figure S1**). Goblet cells are interspersed with the columnar cells. Pancreatobiliary-type ampullary carcinoma shows simple tubular glands with cuboidal to low columnar cells and a single layer of round centrally placed nuclei with an abundant desmoplastic stroma. In this study, immunohistochemical results were adopted to obtain objective evaluation for classifying the two phenotypes (Intestinal-type *versus* Pancreatobiliary-type). According to the recent paper (Ang et al., 2014), "intestinal-type" is defined as having (1) positive staining for CK20 or CDX2 or MUC2 and negative staining for MUC1, or (2) positive staining for CK20, CDX2, and MUC2, irrespective of the MUC1 result; and "pancreatobiliary-type" was defined as having positive staining for MUC1 and negative staining for CDX2

and MUC2, irrespective of CK20 results. Cases not fitting one of these three categories are regarded as "ambiguous".

Cell Culture

Since an immortalized normal epithelial cell line of ampulla has not been established, we used an immortalized normal epithelial cell line of common bile duct origin, designated HBDEC2-3H10 and an immortalized normal epithelial cell line of duodenal mucosa origin, designated HDuodEC3. These lines were selected for functional analyses because *ELF3* mutations have also been observed in 7/74 (9.5%) common bile duct carcinomas in our recent study (Nakamura et al., 2015) and 1/18 (5.6%) duodenal carcinomas in the present study. These cell lines were established by infecting lentiviruses CSII-CMV-CDK4R24C, CSII-CMV-cyclin D1 and CSII-CMV-hTERT (Inagawa et al., 2014) into primary cells isolated from common bile duct and duodenal mucosa, respectively. HBDEC2-3H10 is a clonal cell line established by limiting dilution, and HDuodEC3 is a pooled population with extended life span. HBDEC2-3H10 and HDuodEC3 were maintained in F-medium supplemented with 10 μM Y-27632 (Liu et al., 2012) without feeder cells. Before we use the HBDEC2-3H10 for functional studies, the karyotype analysis was carried out using standard G-banding by an outsourced service (LSI Medience Corporation). Chromosomal analysis of HBDEC2-3H10 in G-banding showed that the majority of the cells (19/20) were near diploid, 46, XY, add (1) (q32), i(8) (q10) [19].

Cell Transfection, Cell Proliferation, and Cell Invasion/Migration

siRNA triplex oligonucleotides against human *ELF3* (*ELF3* siRNA #1, #2 and #3) and a non-targeting negative control siRNA (Silencer Select Negative Control #1 siRNA) were synthesized by Life Technologies and Sigma (see below). Cells were plated into triplicate wells of 96-well plates (5×10^3 cells per well). At the same time, mixtures of siRNA and Lipofectamine RNAiMAX reagent (Invitrogen) were added to each well as 5 nM siRNA solutions. Cell proliferation was assessed after 1, 2, 3 and 4 days using the CellTiter-96 Aqueous One Solution Cell Proliferation Assay (Promega) according to the manufacture's instruction. Cell invasion/migration assays were performed using BD BioCoat Matrigel invasion chamber and control inserts (24 well, BD Biosciences) according to the manufacture's instruction. 2×10^4 HBDEC2-3H10 cells HDuodEC3 cells or transfected with siRNAs were seeded in the upper chamber, while the medium with 5%

fetal bovine serum was placed in the lower chamber. The cells on the lower side of the filters were counted at x100 magnification in five different fields of each filter.

RNA oligonucleoide sequences utilized in this study.

	Sense (5'-3')	Antisense (5'-3')
ELF3 siRNA #1	GAAGUGACGUGGACCUGGATT	UCCAGGUCCACGUCACUUCCA
ELF3 siRNA #2	GCCGAUGACUUGGUACUGATT	UCAGUACCAAGUCAUCGGCCC
ELF3 siRNA #3	GGACCAAACUCACGGACCATT	UGGUCCGUGAGUUUGGUCCTT

Western Blotting, Quantitative RT-PCR and Immunofluorescence Analysis

Cells were seeded in 6-well plates (1.5×10^5 cells per well) and mixtures of siRNA and Lipofectamine RNAiMAX reagent were added to each well as 5 nM siRNA solutions. Seventy-two hours after transfection, cells were lysed in RIPA buffer containing 1% protease inhibitor cocktail. The protein concentration of each sample was measured using a BCA protein assay kit (Thermo Scientific). Whole cell lysates were subjected to 5-25% gradient SDS-PAGE using standard protocols. The following antibodies were used: Elf3 (HPA003316, Sigma) and GAPDH (sc-32233, Santa Cruz Biochemistry). Membranes were probed with secondary anti-rabbit or anti-mouse horseradish peroxidase-conjugated antibodies, and developed using a chemiluminescence western blotting detection system SuperSignal (Thermo Fisher Scientific). Total RNA was extracted using the RNeasy Mini kit (Qiagen). Quantitative RT-PCR was performed with iTaq Universal SYBR Green Supermix (Bio-Rad), using primers for ELF3, MMP1, MMP9, Vimentin, CK19, ZEB1, ZEB2, TWIST1, E-cadherin, or GAPDH (see below). Values obtained in quantitative RT-PCR were normalized to those for GAPDH. Immunofluorescence was performed using a primary antibody of a rabbit anti-vimentin (LB-3010, diluted 1:500, LSL). The secondary antibody was a goat anti-rabbit Alexa Fluor 488 (Invitrogen). After transfection, cells were seeded in 8 well chamber slide (Thermo scientific). The medium was supplemented with 10 μM Y27632 for 48 hr. Then, the medium was changed into a medium without Y27632 and incubated for 48 hr before staining. Stained sections were viewed and photographed using fluorescence microscope (BZ-9000, KEYENCE).

Primer sequences used in real-time PCR.

Gene	Forward primer sequence	Reverse primer sequence

	(5'-3')	(5'-3')
ELF3	GCAACATGACCTACGAGAAGC	CGACTCTGGAGAACCTCTTCC
MMP1	GCTAACCTTTGATGCTATAACTACGA	GGATTTGTGCGCATGTAGAA
MMP9	GAACCAATCTCACCGACAGG	GCCACCCGAGTGTAACCATA
Vimentin	GAGAACTTTGCCGTTGAAGC	TCCAGCAGCTTCCTGTAGGT
CK19	GGTCAGTGTGGAGGTGGATT	TCAGTAACTCGGACCTGCT
ZEB1	AACTGCTGGGAGGATGACAC	TCCTGCTTCATCTGCCTGA
ZEB2	GACCTGGACGTGAAGGAAAA	GGCACTTGCAGAAACACAGA
TWIST1	AGCTACGCCTTCTCGGTCT	CCTTCTCTGGAAACAATGACATC
E-cadherin	TGGAGGAATTCTTGCTTTGC	CGCTCTCCTCCGAAGAAAC
GAPDH	GCTCTCTGCTCCTCCTGTTC	ACGACCAAATCCGTTGACTC

Time-lapse Images (3D Cell Invasion)

3D cell invasion assay of time-lapse images was acquired using IncuCyte (Essen Bioscience) according to the manufacture's instruction with some modification. Briefly, cells transfected with siRNAs were seeded in 96-well Matrigel matrix (BD Biosciences)-coated plate (8×10^3 cells per well). Four hours after cell plating, 50 μL per well of Matrigel matrix was then added to each well. The plate was incubated in a 37°C in 5% CO_2 incubator for 30 min to allow Matrigel matrix to gel, and then overlaid with a medium. Cells were incubated and captured in a 37°C in 5% CO_2 incubator for three days.

Glass Chip Macrodissection (GCM)

GCM is a macro-dissection method that enables breaking up a whole section into uniform squares. The method is carried out with the following steps: 1) a section is attached on the breakable glass substrate and stained; 2) the glass substrate is broken up into smaller squares along the breaking border and the section is cut along the breaking borders at the same time; 3) the broken substrates are collected to microtubes and the portion of the section on the substrates are processed for further analysis. To put the method into practice, a device termed chip sheet is used for the process. The chip sheet consists of a plastic sheet with light-sensitive adhesive on its front side and a breakable glass substrate attached on the adhesive. The glass substrate is a 22 mm square cover glass. The breaking borders are formed by laser light focused inside of the glass substrate as lines at 0.5 mm pitch to form 0.5 mm square chips (manufactured by DISCO). The

light-sensitive adhesive can be cured by ultra violet (wave length, 365 nm) lighting on the backside of the sheet before collecting glass chips from the sheet. The chip sheet is attached to aluminum square plate sized 75 mm with circular hole of 50 mm diameter at the center. Attachment of the frozen section is carried out with chip sheet as well as that of glass slide. Then the section is washed with distilled water and stained with 0.67% toluidine blue for 10 s. Excess dye was washed out with distilled water twice. Then the chip sheet was dried by blower wind at room temperature in 2 min. Breaking is carried out by scraping the back side of the chip sheet using thumb nail. The two directions of scraping are orthogonal to the breaking border respectively. Expansion and UV cure process are needed before chip collection. For uniform expansion, an expander is used. The expander consists of a cylinder on a jack and chip frame holder. The cylinder is raised by the jack and pushed up the back side of the chip sheet to expand. In this process, the broken chips are moved apart. The expanded chips and sheet are fixed with grip rings and the rings are cut from the aluminum plate. UV cure is carried out by 10 s illumination on the back side of the expanded chip sheet with ultra violet LED illuminator (NULED-102CT, NS-Lighting). In the finishing step, chips are collected in microtubes using a collection tool. The collection tool is equipped with a microtubes holder, a USB camera (Dino-lite AD7013MZT, AnMo Electronics), a microscope stage and a punching needle. The USB camera is connected to a laptop PC and used for observation of chips from under the chip sheet. The microtube holder shares the position under the chip sheet with the USB camera alternatively. The punching needle is placed above the chip sheet. The chip sheet is on a stage connected to microscope stage used for precise movement. To collect a chip, the microtube holder is placed under the chip sheet, then the punching needle is moved down to punch out the chip and the dropped chip is collected in the microtube.

Multi-region Exome Sequencing of an Ampullary Carcinoma

The quantity of DNA obtained from each section (regions No.1-No.32) was measured by Qubit (Invitrogen). In this study, five serial frozen sections were prepared and we dissected the tissue from 32 squares of 1 mm^2 units (thickness of 18 μm) (**Figure 5B**) and collected the same region from all five sections into the same tube. The amount of gDNA was 155.0 ng per region on averages (range 69.6-242.7 ng). We excluded the two regions since the amount of DNA was less than 100 ng (regions No.8 and No.9 in **Figure 5B**). The samples were subjected to SureSelect Human All Exon v4.0 (Agilent Technologies) based on exome sequencing. The DNA libraries were prepared according to the manufactory's protocol for the preparation of gDNA libraries from 200 ng DNA samples. The DNA libraries were sequenced on an Illumina Hiseq 2500

paired-read platform (a read length of 2 x 100 bp). The sequencing data of regions No.31 and No.32 were excluded since the tumor purities of cancer cells were low based on the genotyper (Karkinos) (Totoki et al., 2014). The data of region No.21 was also excluded since the number of mutations was extremely large probably due to the sequencing errors.

Phylogenetic Tree

The phylogenetic tree was constructed based on the neighbor-joining method (Saitou and Nei, 1987) and the Nei's genetic distance (Nei, 1978). The Nei's distance was calculated as:

$$D_{xy} = -\ln (I_{xy}),$$

$$I_{xy} = \frac{\sum_{i=1}^{M} p_{ix} p_{iy}}{\sqrt{\sum_{i=1}^{M} p_{ix}^2} \sqrt{\sum_{i=1}^{M} p_{iy}^2}}$$

,where p is the variant allele frequency of a synonymous mutation that was normalized by tumor purities. The subscripts x and y represent two different regions and M indicates the number of sites.

SUPPLEMENTAL REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B *57*, 289-300.

Chan-On, W., Nairismagi, M. L., Ong, C. K., Lim, W. K., Dima, S., Pairojkul, C., Lim, K. H., McPherson, J. R., Cutcutache, I., Heng, H. L.*, et al.* (2013). Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. Nature genetics *45*, 1474-1478.

Inagawa, Y., Yamada, K., Yugawa, T., Ohno, S., Hiraoka, N., Esaki, M., Shibata, T., Aoki, K., Saya, H., and Kiyono, T. (2014). A human cancer xenograft model utilizing normal pancreatic duct epithelial cells conditionally transformed with defined oncogenes. Carcinogenesis *35*, 1840-1846.

Jiao, Y., Pawlik, T. M., Anders, R. A., Selaru, F. M., Streppel, M. M., Lucas, D. J., Niknafs, N., Guthrie, V. B., Maitra, A., Argani, P.*, et al.* (2013). Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. Nature genetics *45*, 1470-1473.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O. A., Nealon, C., Dakic, A., Simic, V.*, et al.* (2012). ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. The American journal of pathology *180*, 599-607.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics *89*, 583-590.

Oshima, M., Okano, K., Muraki, S., Haba, R., Maeba, T., Suzuki, Y., and Yachida, S. (2013). Immunohistochemically Detected Expression of 3 Major Genes (CDKN2A/p16, TP53, and SMAD4/DPC4) Strongly Predicts Survival in Patients With Resectable Pancreatic Cancer. Ann Surg *258*, 336-346.

Totoki, Y., Tatsuno, K., Covington, K. R., Ueda, H., Creighton, C. J., Kato, M., Tsuji, S., Donehower, L. A., Slagle, B. L., Nakamura, H.*, et al.* (2014). Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nature genetics *46*, 1267-1273.

Waddell, N., Pajic, M., Patch, A. M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K.*, et al.* (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. Nature *518*, 495-501.