JBUON 2015; 20(6): 1565-1575
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

Genome-wide pathway analysis in pancreatic cancer

Young Ho Lee, Gwan Gyu Song

Division of Rheumatology, Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea

Summary

Purpose: *The purpose of this study was to identify candidate single-nucleotide polymorphisms (SNPs) that might play a role in susceptibility to pancreatic cancer, elucidate their potential mechanisms, and generate SNP-to-gene-to-pathway hypotheses.*

Methods: *A genome-wide association study (GWAS) dataset of pancreatic cancer that included 496,959 SNPs from 3,851 pancreatic cancer patients and 3,934 control subjects of European descent was used in this study. The Identify candidate Causal SNPs and Pathways (ICSNPathway) method was applied to the GWAS dataset.*

Results: *ICSNPathway analysis identified 18 candidate SNPs, 11 genes (including HNF1A and HNF4G), and 30 pathways, which revealed 11 hypothetical biological mechanisms. The strongest hypothetical biological mechanism was one wherein rs2230739 alters the role of ADCY9 in various pathways and processes, including cyclase activity, phosphorus oxygen lyase activity, hsa04912, hsa04540, hsa04020, and hsa00230 ($0.010 \leq p < 0.001$; $0.038 \leq$ false discovery rate (FDR) ≤ 0.016). The second strongest mechanism was that rs16859886 modulates ADCY10 to affect its role in pathways including cyclase activity, phosphorus oxygen lyase activity, nucleobase, nucleoside, and nucleotide metabolic processing, and hsa00230 ($0.010 \leq p < 0.001$; $0.038 \leq FDR \leq 0.016$).*

Conclusions: *By using the ICSNPathway to analyze pancreatic cancer GWAS data, 18 candidate SNPs, 11 genes (including ADCY9, ADCY10, HNF1A, and HNF4G), and 30 pathways were identified that might contribute to the susceptibility of patients to pancreatic cancer.*

Key words: *genome-wide association study, pancreatic cancer, pathway-based analysis*

Introduction

Pancreatic cancer is a malignant neoplasm originating from transformed cells arising in the pancreas. Pancreatic cancer is the fourth most common cause of cancer-related deaths in the United States and the eighth worldwide [1], and has an extremely poor prognosis. The 5-year relative survival rate is 6%, all stages combined [1,2]. Although pancreatic cancer is complex and heterogeneous, and its etiology has not been determined, a genetic component of susceptibility to pancreatic cancer was established by case–control and family studies [3,4]. Previous studies have discovered that SNPs were highly associated with pancreatic cancer within genes such as *KRAS, BRCA1, 2, PALB2, STK11/LKB1, PALB2, PRSS1, SPINK1,* and *CDKN2A* [3,4].

GWASs offer a powerful means to search for genes that confer susceptibility to complex diseases [5]. As a result, an increasing number of GWASs are being reported, and this has led to the discovery and validation of novel disease genes [6]. Although large-scale GWASs have been carried out on complex diseases, including pancreatic cancer, many genetic components that contribute to variation in pancreatic cancer remain unexplained.

Available data suggest that individual genes and genetic variants make small-risk contributions to pancreatic cancer susceptibility by interacting with each other. Although some genetic

Correspondence to: OYoung Ho Lee, MD, PhD. Division of Rheumatology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, 126-1 5 ga, Anam-dong, Seongbuk-gu, Seoul 136-705, Korea. Tel: +822 920 5645, Fax: +822 922 5974, E-mail: lyhcgh@korea.ac.kr
Received: 12/05/2015; Accepted: 28/05/2015

signals have been examined at the single marker level in pancreatic cancer GWASs, the biological mechanisms identified remain controversial [7]. A key challenge in interpreting GWAS data is identifying causative SNPs and providing evidence for the hypothetical mechanisms that could be responsible for the observed traits [8-10]. Thus, we hypothesized that using a novel method to study existing GWAS datasets could provide additional insights and identify new candidate genes. The ICSNPathway method was developed to identify candidate SNPs and their corresponding candidate pathways using GWAS data together with integrated linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) [11].

We applied ICSNPathway analysis to a pancreatic cancer GWAS dataset to identify candidate causal SNPs and mechanisms of pancreatic cancer susceptibility, and to generate SNPto-gene-to-pathway hypotheses.

Methods

Study population

We used a publicly available pancreatic cancer GWAS dataset from NCBI dbGap (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000206.v3.p2), which included the genotypes of 543,436 SNPs obtained from the Illumina 610 Quad Infinium genotyping assay. The dataset comprised 3,851 pancreatic cancer patients and 3,934 controls of European descent. Pancreatic cancer cases were drawn from 12 cohort studies and 8 case–control studies. The dataset was filtered to remove individuals with p<0.001 for Hardy–Weinberg violation, and a call rate of <98% to reduce the effect of genotyping errors. In all, 496,959 SNPs passed the quality control filters.

Identification of candidate causal SNPs and pathways

ICSNPathway analysis was carried out in 2 stages [11]. The first stage involved the pre-selection of candidate causal SNPs by LD analysis and functional SNP annotation based on the most significant SNPs. The second stage involved annotating the biological mechanisms to preselected candidate causal SNPs using the PBA algorithm i-GSEA (improved gene-set enrichment analysis) [11].

A full list of pancreatic cancer GWAS SNP p values was entered into the ICSNPathway analysis. One concept applied in the ICSNPathway analysis is LD analysis, which searches for the most significant SNPs in LD within a GWAS dataset to identify more possible candidate causal SNPs based on an extended dataset, including HapMap data [12]. The other method involves the use of functional SNPs. ICSNPathway analysis pre-selects candidate causal SNPs based on functional SNPs, which are important for understanding the underlying genetics of human health. Functional SNPs are defined as SNPs that may alter protein or gene expression, or the role of a protein in the context of a pathway. They include deleterious and non-deleterious non-synonymous SNPs, SNPs that cause the gain or loss of a stop codon, those resulting in a frame shift, and SNPs located in essential splice sites or in regulatory regions.

The ICSNPathway server applies the *i*-GSEA PBA algorithm to the full list of GWAS SNP p values to detect the pathways associated with individual traits. Briefly, the process is as follows. (1): Each SNP is mapped to its nearest gene according to the localization of the SNP and the gene in the Ensembl 61 database (http://www.ensembl.org/biomart/martview), and the maximum $t = -\log(p\text{-value})$ values of SNPs mapped to genes are assigned to represent those genes. Then, all genes are ranked by decreasing representative t values. (2): For each pathway S, the enrichment score (*ES*, i.e., a Kolmogorov–Smirnov-like running-sum statistic with weight [a]) is calculated, which measures the tendency for genes of a pathway to be located at the top of the ranked gene list. (3): The *ES* is then converted to a significant-proportion-based *ES* (*SPES*) by multiplying the *ES* by m_1/m_2, where m_1 is the proportion of significant genes for pathway *S* (defined as genes mapped with at least one SHLP in the top 5% of the most significant SNPs in the GWAS), and m_2 is the proportion of significant genes for all genes in the GWAS. (4): SNP label permutation and normalization are used to generate the distribution of *SPES* and to correct for gene variation (bias caused by different genes with different numbers of mapped SNPs) and pathway variation (bias due to different pathways with different numbers of genes). (5): Based on the distribution of *SPES* values generated by the permutation, a nominal p value is calculated, and a false discovery rate (FDR) is computed for multiple testing correction.

The term "the most significant SNP" refers to SNPs with a p value below a certain threshold, which can be specified from the GWAS SNP p values. The ICSNPathway was used to analyze the significant pathways from the original GWAS when we chose the p value threshold ($<1\text{x}10^{-3}$) used in the study. Two parameters were set for the analysis. The first was "within gene," meaning that only p values of SNPs located within genes were used in the PBA algorithm. The second was an FDR cutoff (0.05) for multiple testing corrections. Control of the FDR is preferred for large-scale testing. Defined as the expected proportion of false-positives among all significant tests, it allows researchers to identify a set of "candidate positives," a high proportion of which are likely to be true-positives. The FDR, a permutation-based approach for multiple comparisons, was used to identify statistically significant genes. There were no specific criteria used to select the number of genes. We used cut-offs of a minimum of 2 and maximum of 500 to avoid very narrow or very broad

functional categories. We discarded pathways that contained more than 500 genes to avoid stochastic bias and the inclusion of a general biological process. Out of the several options available for pathway annotation, we selected 4 pathway databases: the Kyoto Encyclopedia of Genes and Genomes (KEGG) [13], BioCarta, gene ontology (GO) biological process [14], and GO molecular function. This ensured comprehensive coverage of pathways and high-quality information for well-defined pathways.

SNAP was developed to identify and annotate nearby SNPs in LD (proxies) by HapMap (http://www.broadinstitute.org/mpg/snap/). In this study, we used SNAP to (i) find proxy SNPs, (ii) determine whether SNP proxies were present in genes, (iii) resolve whether associations from multiple SNPs represented similar associations, (iv) plot regional views of associations or LD structures, and (v) retrieve annotations for SNPs [15].

Results

Candidate SNPs and pathways identified from the pancreatic cancer GWAS

Using the 496,959 GWAS SNP p values as the input and the most significant SNPs ($p < 1x10^{-3}$), ICSNPathway analysis identified 18 candidate SNPs, 11 genes, and 30 pathways (Tables 1-3; Figure 1). The top 5 candidate SNPs were rs2230739 ($-\log_{10}[p] = 3.066$), rs16859886 ($-\log_{10}[p] = 3.837$), rs1169288 ($-\log_{10}[p] = 3.757$), rs2464195 ($-\log_{10}[p] = 3.742$), and rs2464196 ($-\log_{10}[p] = 3.997$). Two of the 5 candidate SNPs (excluding rs16859886, rs1169288, and rs2464195) were not in LD with any other SNP. SNP rs16859886, which was in LD with rs203848 (r^2=0.805), was not present in the original GWAS dataset. Similarly, rs1169288, which was in LD with rs2650000 (r^2=0.911),

Table 1. Candidate causal SNPs

Candidate causal SNP	*Functional class*	*Gene*	*Candidate causal pathway**	*$-log_{10}(p)^{\dagger}$*	*In LD with*	*r^2*	*D'*	*$-log_{10}(p)^{\S}$*
rs2230739	Non-synonymous coding	ADCY9	1 2 20 23 26 30	3.066	rs2230739	–	–	3.066
rs16859886	Non-synonymous coding	ADCY10	1 2 18 30	–	rs203848	0.805	0.897	3.837
rs1169288	Non-synonymous coding	HNF1A	3 7 9 11 19 25 28	–	rs2650000	0.911	1.0	3.757
rs2464195	Non-synonymous coding	HNF1A	3 7 9 11 19 25 28	–	rs735396	1.0	1.0	3.742
rs2464196	Non-synonymous coding	HNF1A	3 7 9 11 19 25 28	3.997	rs2464196	–	–	3.997
rs2943549	Non-synonymous coding & splice site	HNF4G	3 4 7 8 9 11	–	rs2943547	0.821	1.0	3.133
rs2272669	Regulatory region	HNF4G	3 4 7 8 9 11	–	rs1805100	1.0	1.0	4.119
rs1805100	Regulatory region	HNF4G	3 4 7 8 9 11	4.119	rs1805100	–	–	4.119
rs6895902	Non-synonymous coding	MAML1	4 7 8 11 28	2.580	rs7734102	1.0	1.0	3.028
rs7178777	Non-synonymous coding (deleterious)	NUSAP1	5	3.027	rs7178777	–	–	3.027
rs7178634	Non-synonymous coding (deleterious)	NUSAP1	5	3.001	rs7178634	–	–	3.001
rs11239430	Regulatory region	OR13A1	6	–	rs7098434	1.0	1.0	3.004
rs2215530	Non-synonymous coding (deleterious)	OR1L4	6	–	rs7046355	1.0	1.0	3.227
rs10985760	Non-synonymous coding (deleterious)	OR1L6	6	–	rs7046355	1.0	1.0	3.227
rs10818740	Non-synonymous coding	OR1L6	6	–	rs7046355	1.0	1.0	3.227
rs773902	Non-synonymous coding	F2RL3	10 12 13 14 15 16 17 19 21 22 27 29	3.401	rs773902	–	–	3.401
rs1805082	Non-synonymous coding	NPC1	19 24 25	–	rs11663558	0.925	1.0	3.363
rs1805081	Non-synonymous coding	NPC1	19 24 25	3.250	rs1805081	–	–	3.250

SNP: single-nucleotide polymorphism, LD: linkage disequilibrium. *Numbers indicate the indices of pathways (listed in Table 3) ranked by significance (false discovery rate). †–log10(p) values of candidate causal SNPs in the original genome-wide association studies (GWASs). "–" denotes that this SNP was not represented in the original GWAS. §–log10(p) values of SNPs in LD with candidate causal SNPs in the original GWAS.

Table 2. Functional and association study of genes identified by GWAS pathway analysis

Gene	*Function*
ADCY9	May play a fundamental role in situations where the fine interplay between intracellular calcium and cAMP determines the cellular function. May be a physiologically relevant docking site for calcineurin (by similarity).
ADCY10	Soluble adenylyl cyclase that has a critical role in mammalian spermatogenesis. Produces cAMP, which mediates, in part, the cAMP-responsive nuclear factors indispensable for maturation of sperm in the epididymis. Induces capacitation, the maturational process that sperms undergo prior to fertilization. May be the bicarbonate sensor.
HNF1A	Transcriptional activator that regulates tissue-specific expression of multiple genes, especially in the pancreatic islet cells and liver. Required for the expression of several liver-specific genes. Binds to the inverted palindrome 5'-GTTAATNATTAAC-3'.
HNF4G	Transcription factor. Has a lower transcription activation potential than HNF4-alpha.
MAML1	Acts as a transcriptional coactivator for NOTCH proteins. Has been shown to amplify NOTCH-induced transcription of HES1. Enhances phosphorylation and proteolytic turnover of the NOTCH intracellular domain in the nucleus through interaction with CDK8. Binds to CREBBP/CBP, which promotes nucleosome acetylation at NOTCH enhancers and activates transcription. Induces phosphorylation and localization of CREBBP to nuclear foci. Plays a role in hematopoietic development by regulating NOTCH-mediated lymphoid cell fate decisions.
NUSAP1	Microtubule-associated protein with the capacity to bundle and stabilize microtubules (by similarity). May associate with chromosomes and promote the organization of mitotic spindle microtubules around them.
OR13A1	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G-protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome.
OR1L4	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G-protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome.
OR1L6	Olfactory receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G-protein-mediated transduction of odorant signals. The olfactory receptor gene family is the largest in the genome.
F2RL3	Receptor for activated thrombin or trypsin coupled to G proteins that stimulate phosphoinositide hydrolysis. May play a role in platelet activation.
NPC1	Intracellular cholesterol transporter, which acts in concert with NPC2 and plays an important role in the egress of cholesterol from the endosomal/lysosomal compartment. Both NPC1 and NPC2 function as the cellular 'tag team duo' to catalyze the mobilization of cholesterol within the multivesicular environment of the late endosome (LE) to effect egress through the limiting bilayer of the LE. NPC2 binds unesterified cholesterol that has been released from LDLs in the lumen of the late endosomes/lysosomes and transfers it to the cholesterol-binding pocket of the *N*-terminal domain of NPC1. Cholesterol binds to NPC1 with the hydroxyl group buried in the binding pocket and is exported from the limiting membrane of late endosomes/lysosomes to the ER and plasma membrane by an unknown mechanism. Binds oxysterol with higher affinity than cholesterol. May play a role in vesicular trafficking in the glia, a process that may be crucial for maintaining the structural and functional integrity of nerve terminals.

and rs2464195, which was in LD with rs735396 (r^2=1.0), were not present in the original GWAS dataset.

The biological mechanisms suggest that the candidate SNPs may alter the role of the corresponding genes or proteins in the context of pathway(s) associated with traits. The 18 candidate SNPs were included in 30 candidate pathways with roles in 11 hypothetical biological mechanisms. The strongest hypothetical biological mechanism was that rs2230739 alters the role of ADCY9 in various pathways and processes, including cyclase activity, phosphorus oxygen lyase activity, hsa04912, hsa04540, hsa04020, and hsa00230 ($0.010 \leq p<0.001$; $0.038 \leq FDR \leq 0.016$). The second strongest mechanism was that

Table 3. Candidate causal pathways

Index	*Candidate causal pathway*	*Description*	*p*	*FDR*
1	Cyclase activity	GO:0009975. Catalysis of a ring closure reaction	< 0.001	0.016
2	Phosphorus oxygen lyase activity	GO:0016849.	< 0.001	0.017
3	hsa04950	Maturity onset diabetes of the young	0.002	0.018
4	Regulation of transcription from RNA polymerase II promoter	GO:0006357. Any process that modulates the frequency, rate or extent of transcription from an RNA polymerase II promoter	< 0.001	0.018
5	DNA packaging	GO:0006323. Any process by which DNA and associated proteins are formed into a compact, orderly structure.	0.007	0.019
6	hsa04740	Olfactory transduction	0.001	0.019
7	Regulation of transcription DNA dependent_	GO:0006355. Any process that modulates the frequency, rate or extent of DNA-dependent transcription	0.001	0.020
8	Transcription from RNA polymerase II promoter	GO:0006366. The synthesis of RNA from a DNA template by RNA polymerase II (Pol II), originating at a Pol II-specific promoter. Includes transcription of messenger RNA (mRNA) and certain small nuclear RNAs (snRNAs).	< 0.001	0.021
9	Sequence-specific and DNA-binding-specific	GO:0043565. Interacting selectively with DNA of a specific nucleotide composition, e.g., GC-rich DNA binding, or with a specific sequence motif or type of DNA, e.g., promoter binding or rDNA binding.	< 0.001	0.021
10	Thrombin receptor activity	GO:0015057. Combining with thrombin to initiate a G-protein-mediated change in cell activity. A G-protein is a signal transduction molecule that alternates between an inactive GDP-bound and an active GTP-bound state.	0.008	0.022
11	Regulation of RNA metabolism	GO:0051252. Any process that modulates the frequency, rate, or extent of the chemical reactions and pathways involving RNA.	< 0.001	0.022
12	Regulation of body fluid levels	GO:0050878. Any process that modulates the levels of body fluids.	0.001	0.022
13	Hemostasis	GO:0007599. The stopping of bleeding (loss of body fluid) or the arrest of the circulation to an organ or part.	0.003	0.023
14	Blood coagulation	GO:0007596. The sequential process by which the multiple coagulation factors of the blood interact, ultimately resulting in the formation of an insoluble fibrin clot; it may be divided into three stages: stage 1, the formation of intrinsic and extrinsic prothrombin converting principle; stage 2, the formation of thrombin; stage 3, the formation of stable fibrin polymers.	0.003	0.023
15	Response to wounding	GO:0009611. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) because of a stimulus, indicating damage to the organism.	< 0.001	0.025
16	Coagulation	GO:0050817. The process by which a fluid solution, or part of it, changes into a solid or semisolid mass.	0.003	0.025
17	Response to external stimulus	GO:0009605. A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) because of an external stimulus.	0.002	0.026
18	Nucleobase, nucleoside, and nucleotide metabolic process	GO:0055086. The chemical reactions and pathways involving nucleobases, nucleosides, and nucleotides.	0.002	0.027
19	Regulation of biological quality	GO:0065008. Any process that modulates the frequency, rate or extent of a biological quality. A biological quality is a measurable attribute of an organism or part of an organism, such as size, mass, shape, color, etc.	< 0.001	0.029

Continued on next page

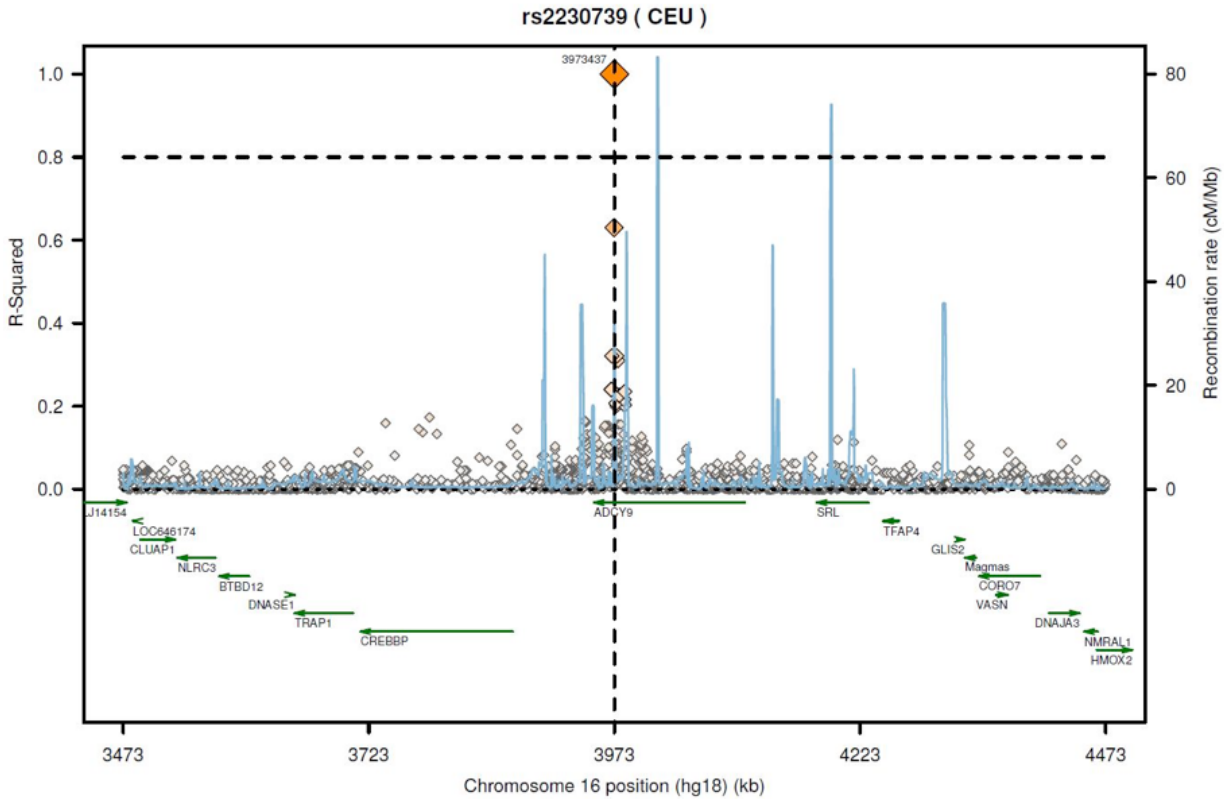
20	hsa04912	GnRH-signaling pathway	0.001	0.029
21	G protein signaling coupled to IP3 second messenger phospholipase C activation	GO:0007200. The series of molecular signals generated because of the binding of a G-protein-coupled receptor to its physiological ligand, followed by the activation of phospholipase C and the subsequent release of inositol trisphosphate.	0.003	0.030
22	hsa04080	Neuroactive ligand–receptor interaction	0.002	0.030
23	hsa04540	Gap junction	0.001	0.030
24	Lysosomal transport	GO:0007041. The directed movement of substances into, out of, or within a lysosome.	0.009	0.033
25	Chemical homeostasis	GO:0048878. The biological processes involved in the maintenance of an internal equilibrium of a chemical.	0.003	0.035
26	hsa04020	Calcium-signaling pathway	0.001	0.038
27	Second messenger-mediated signaling	GO:0019932. A series of molecular signals in which an ion or small molecule is formed or released into the cytosol, thereby helping relay the signal within the cell.	0.004	0.038
28	Positive regulation of macromolecule metabolic process	GO:0010604. Any process that increases the frequency, rate, or extent of chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass.	0.008	0.043
29	Receptor activity	GO:0004872. Combining with an extracellular or intracellular messenger to initiate a change in cell activity.	0.006	0.046
30	hsa00230	Purine metabolism	0.010	0.046

FDR: false discovery rate

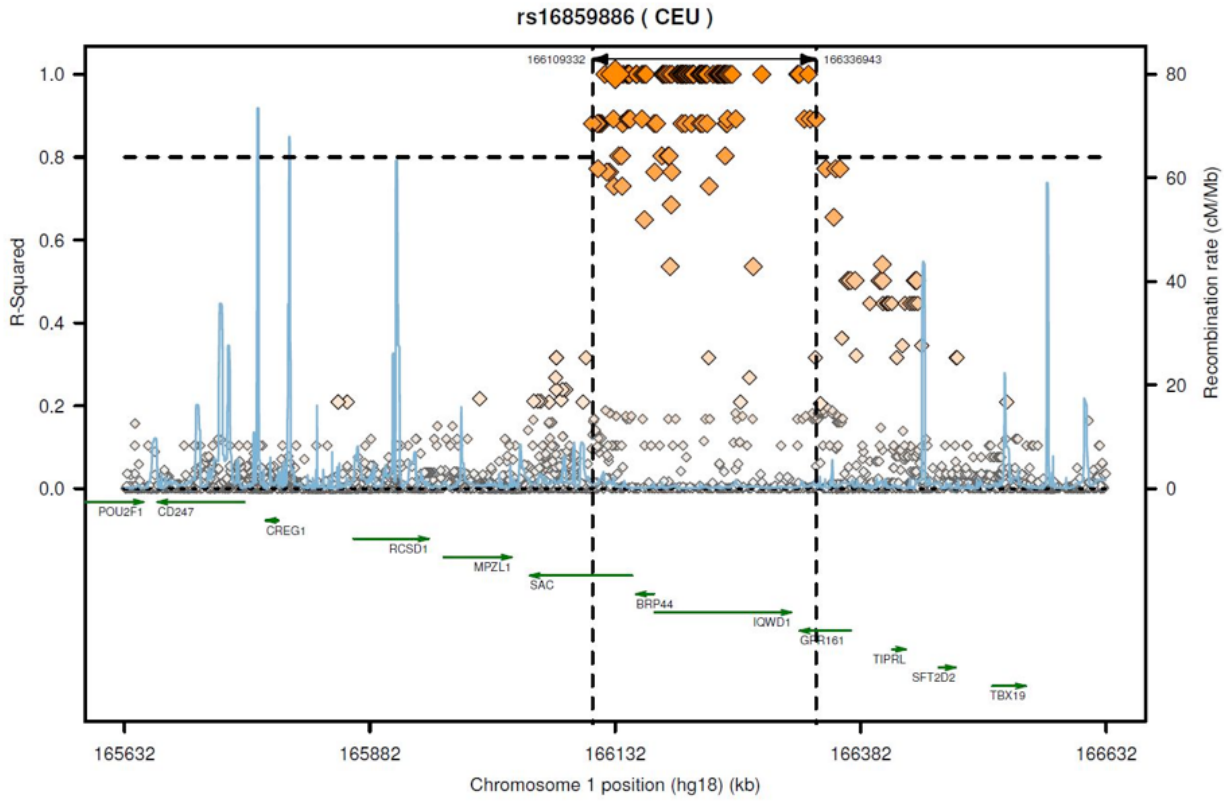
rs16859886 modulates ADCY10 to affect its role in various pathways and processes, including cyclase activity, phosphorus oxygen lyase activity, nucleobase, nucleoside, and nucleotide metabolic processing, and hsa00230 (0.010 ≤ p<0.001; 0.038 ≤ FDR ≤ 0.016). The third mechanism was that rs1169288, rs2464195, rs2464196 modulate HNF1A to regulate DNA-dependent, sequence-specific, and DNA–binding-specific transcription, regulation of RNA metabolic processing, regulation of biologic quality, chemical homeostasis, and positive regulation of macromolecule metabolic processing (0.008≤ p <0.002; 0.043≤ FDR ≤0.018). The fourth mechanism was that rs2943549, rs2272669, rs1805100→HNF4G→hsa04950, affect the regulation of transcription from RNA polymerase II promoter, regulation of DNA-dependent, sequence-specific, and DNA-binding-specific transcription, and regulation of RNA metabolic processing (0.002≤ p <0.001; 0.022≤ FDR ≤0.018). The fifth mechanism was that rs6895902→MAML1 modify the regulation of transcription from RNA polymerase II promoter, regulation of DNA-dependent transcription, regulation of RNA metabolic processing, and positive regulation of macromolecule metabolic processing (0.008≤ p <0.001; 0.043≤ FDR ≤0.018). The sixth mechanism was rs7178777, s7178777→NUSAP1→DNA packaging (p=0.007; FDR=0.019). The seventh mechanism was rs112394306→OR13A1 hsa04740 (p=0.001;FDR=0.019). The eighth mechanism was rs2215530→OR1L4→hsa04740 (p=0.001;FDR=0.019). The ninth mechanism was rs10985760 →OR1L6→hsa04740 (p=0.001; FDR=0.019). The tenth mechanism was→rs773902 F2RL3 to modify thrombin receptor activity, regulation of body fluid levels, hemostasis, blood coagulation, response to wounding, coagulation, response to external stimulus, regulation of biologic quality, G protein signaling coupled to IP3 second messenger, phospholipase C activation, hsa04080, second messenger-mediated signaling, and receptor activity (0.008≤ p <0.001; 0.043≤ FDR ≤0.022); and the eleventh mechanism was rs1805082, rs1805081 →NPC1 to modify regulation of biologic quality, lysosomal transport, and chemical homeostasis (0.009≤ p <0.001; 0.035≤ FDR ≤0.029).

We next investigated which genes play a role in the identified pathways using a pathwaybased

A


rs2230739 (CEU)
3973437
R-Squared
Recombination rate (cM/Mb)
1.0
0.8
0.6
0.4
0.2
0.0
80
60
40
20
0
LJ14154
LOC646174
CLUAP1
NLRC3
BTBD12
DNASE1
TRAP1
CREBBP
ADCY9
SRL
TFAP4
GLIS2
Magmas
CORO7
VASN
DNAJA3
NMRAL1
HMOX2
3473
3723
3973
4223
4473
Chromosome 16 position (hg18) (kb)

B


rs16859886 (CEU)
166109332
166336943
R-Squared
Recombination rate (cM/Mb)
1.0
0.8
0.6
0.4
0.2
0.0
80
60
40
20
0
POU2F1
CD247
CREG1
RCSD1
MPZL1
SAC
BRP44
IQWD1
GPR161
TIPRL
SFT2D2
TBX19
165632
165882
166132
166382
166632
Chromosome 1 position (hg18) (kb)

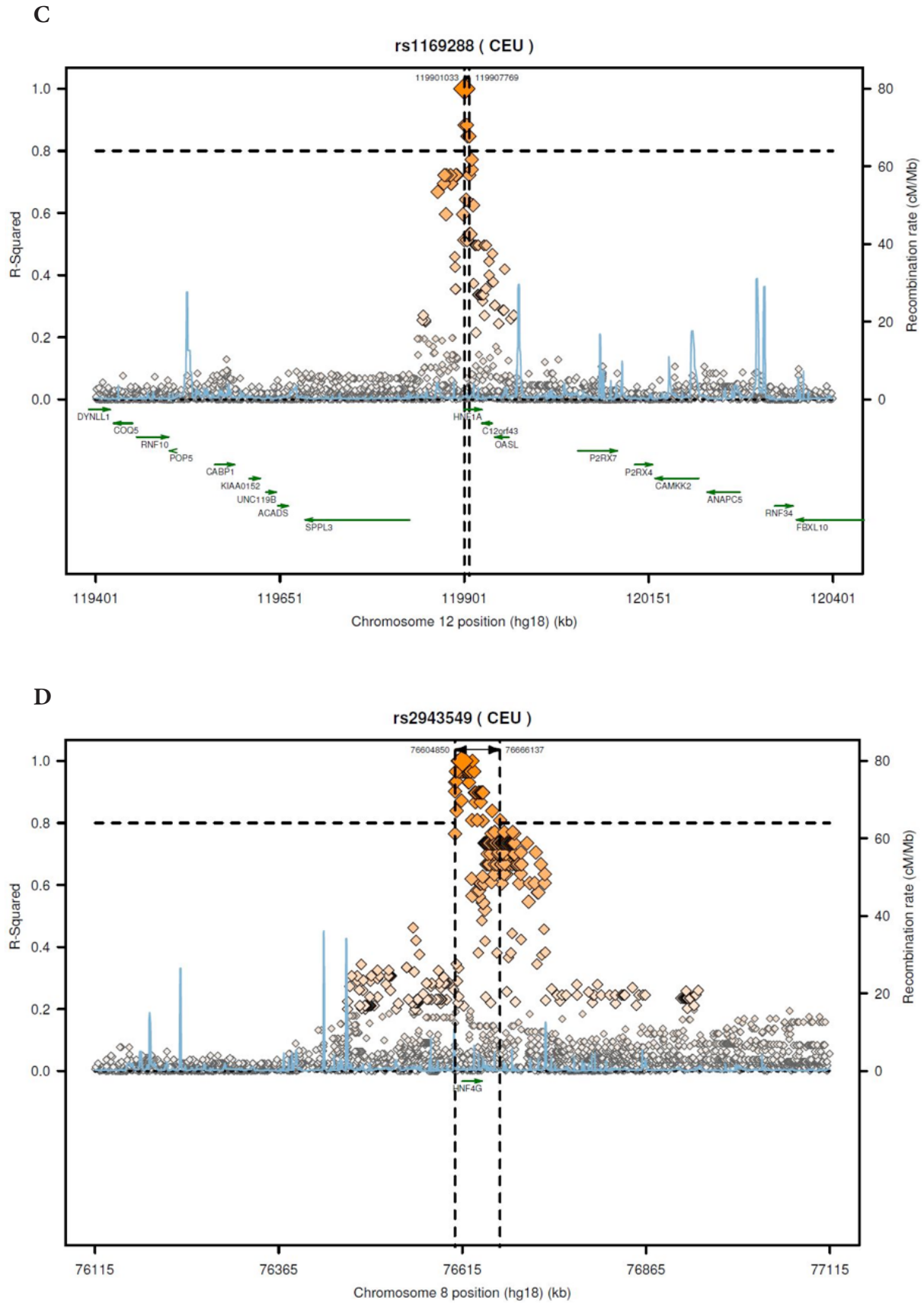


Figure 1. Regional linkage disequilibrium plots of rs2230739 (*ADCY9*) **(A)**, rs16859886 (*ADCY10*) **(B)**, rs1169288 (*HNF1A*) **(C)**, and rs2943549 (*HNF4G*) **(D)** single-nucleotide polymorphisms (SNPs). SNPs are plotted along with their proxies (based on 1000 genomes pilot 1 CEU) as a function of genomic location, and annotated by the recombination rate across the locus (light blue line). On the *y*-axis, pairwise r^2 values are provided for each proxy SNP by using color codes.

approach, and found a distinct clustering of genes involved in the candidate pancreatic cancer-causal pathways. The most significant pathway was the cyclase activity, specifically involving *ADCY10*, *GUCY1A2*, *ADCY8*, *ADCY9*, *GUCY1A3*, *GUCY2D*, *GUCY2C*, *GUCY1B3*, *ADCY7*, and *RTCD1* ($p<0.05$). The second most significant pathway was the phosphorus oxygen lyase activity, which involved *ADCY10*, *GUCY1A2*, *ADCY8*, *ADCY9*, *GUCY1A3*, *GUCY2D*, *GUCY2C*, *GUCY1B3*, and *ADCY7* ($p<0.05$).

Discussion

Multiple related genes in a pathway likely work together to confer disease susceptibility, and some of these genes may not reach genome-wide significance in any single GWAS. Therefore, PBA is required to identify new loci associated with susceptibility to complex diseases [16,17]. A complex molecular network and various cellular pathways may play key roles in the development of pancreatic cancer. If a specific pathway was relevant to disease susceptibility, association signals would be expected to be overrepresented for SNPs in that pathway [18]. Given the limited power of GWAS to detect single SNP associations, we adopted a pathwaybased approach to take into account the biological interplay between genes, and provide insights into how multiple genes might contribute to the pathogenesis of pancreatic cancer [19].

In the present pathway analysis, we used ICSNPathway analysis to identify 18 candidate SNPs, 11 genes, and 30 pathways, which provided 11 hypothetical biological mechanisms.

Our analysis identified 11 candidate genes (*ADCY9*, *ADCY10*, *HNF1A*, *HNF4G*, *MAML1*, *NUSAP1*, *OR13A1*, *OR1L4*, *OR1L6*, *F2RL3*, and *NPC1*) that could contribute to pancreatic cancer risk. The most significant SNP-to-gene-to-effect hypothesis was that rs2230739 alters the role of ADCY9 in various pathways and processes, including cyclase activity, phosphorus oxygen lyase activity, hsa04912, hsa04540, hsa04020, and hsa00230. ADCY9 may play a fundamental role in situations where the fine interplay between intracellular calcium and cAMP determines the cellular function, and may be a physiologically relevant docking site for calcineurin [20]. Elevations in intracellular cAMP trigger bicarbonate and fluid secretion from pancreatic duct cells, and known stimulants of pancreatic ductal secretion (VIP and βadrenergic agonists) stimulated and increased the expression of adenylate cyclase activity in human pancreatic adenocarcinoma cell lines [21]. Abnormality in ductal function may lead to pancreatic pathology, and about 90% of pancreatic cancer is of ductal origin [22]. However, the association between ADCY9 and pancreatic cancer remains unclear. The second strongest mechanism was that rs16859886 modulates ADCY10 to affect its role in various pathways and processes, including cyclase activity and phosphorus oxygen lyase activity. The third mechanism was that rs1169288, rs2464195, and rs2464196 modulate HNF1A to regulate DNA-dependent, sequence-specific, and DNA-binding-specific transcription, regulation of RNA metabolic processing, regulation of biologic quality, chemical homeostasis, and positive regulation of macromolecule metabolic processing. HNF1A is a transcriptional activator that regulates the tissue specific expression of multiple genes, especially in pancreatic islet cells and liver [23]. HNF1A is one of the critical regulatory transcription factors in the developing and mature pancreas]23]. The fourth mechanism was that rs2943549, rs2272669, rs1805100→HNF4G →hsa04950 affect the regulation of transcription from RNA polymerase II promoter, regulation of DNA-dependent, sequence-specific, and DNA-binding-specific transcription, and regulation of RNA metabolic processing. HNF4G is a transcription factor with a lower transcription activation potential than HNF4-alpha [24,25]. GWASs have revealed that HNF1A and HNF4G play key roles as susceptibility genes in the development of pancreatic cancer [25,26]. The fifth mechanism was rs6895902→MAML1 that affects the regulation of transcription from RNA polymerase II promoter, regulation of DNA-dependent transcription, regulation of RNA metabolic processing, and positive regulation of macromolecule metabolic processing. MAML1 acts as a transcriptional coactivator for NOTCH proteins, and enhances phosphorylation and proteolytic turnover of the NOTCH intracellular domain in the nucleus through interaction with CDK8 [27]. MAML1 plays a role in hematopoietic development by regulating NOTCH-mediated lymphoid cell fate decisions [27]. Disrupting a NOTCH signaling pathway may contribute to mucoepidermoid carcinoma, including pancreatic cancer [28].

Genes without a known immunological function are of particular interest because they could lead to the identification of novel mechanisms for susceptibility to pancreatic cancer. The present analysis suggests that the pathway and genes identified by PBA may contribute to susceptibility to pancreatic cancer.

Although these hypothetical mechanisms might contribute to pancreatic cancer susceptibility, the PBA test may also be susceptible to false-positive results, like singlemarker-based association tests. Therefore, these results should be viewed as preliminary and be used to generate new hypotheses, which should then be appropriately verified in independent studies. Thus, additional studies are needed to confirm the association between the candidate SNPs, genes, and pathways and pancreatic cancer. Nevertheless, pathway-based approaches play a complementary role in the identification of genes that confer disease susceptibility. Therefore, the results obtained in the present study may lead to the formulation of novel hypotheses for future investigation.

In summary, we examined a pancreatic cancer GWAS dataset to identify genetic associations with pancreatic cancer at both the SNP and pathway levels. By applying ICSNPathway analysis to the pancreatic cancer GWAS dataset, we identified 18 candidate SNPs and 11 genes (including *ADCY9*, *ADCY10*, *HNF1A*, *HNF4G*, and *MAML1*), 30 pathways, and 11 biological mechanisms that might contribute to pancreatic cancer susceptibility. However, further studies are needed to confirm and explore the genetic variations of the molecular pathways that might be associated with pancreatic cancer.

References

1. Hariharan D, Saied A, Kocher HM. Analysis of mortality rates for pancreatic cancer across the world. HPB (Oxford) 2008;10:58-62.
2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60:277-300.
3. Klein AP. Genetic susceptibility to pancreatic cancer. Mol Carcinog 2012;51:14-24.
4. Lin Y, Yagyu K, Egawa N et al. An overview of genetic polymorphisms and pancreatic cancer risk in molecular epidemiologic studies. J Epidemiol 2011;21:2-12.
5. Manolio TA. Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010;363:166-176.
6. Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. BMC Med Genet 2009;10:6.
7. Wei P, Tang H, Li D. Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. PLoS One 2012;7:e46887.
8. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 2010;11:843-854.
9. Lee YH, Song GG. Pathway analysis of genome-wide association studies on uric acid concentrations. Hum Immunol 2012;73:805-810.
10. Lee YH, Bae SC, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of genomewide association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol Biol Rep 2012;39:10627-10635.
11. Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. Nucleic Acids Res 2011;39 (Web Server issue):W437-443.
12. Altshuler DM, Gibbs RA, Peltonen L et al. International HapMap C integrating common and rare genetic variation in diverse human populations. Nature 2010;467:52-58.
13. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010;38 (Database issue):D355-360.
14. Ashburner M, Ball CA, Blake JA et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-29.
15. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008;24:2938-2939.
16. Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. Pharmacogenomics 2009;10:191-201.
17. Elbers CC, van Eijk KR, Franke L et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet Epidemiol 2009;33:419-431.
18. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. Eur J Hum Genet 2011;19:837-843.
19. Pedroso I, Breen G. Gene set analysis and network analysis for genome-wide association studies. Cold Spring Harb Protoc 2011;2011(9).
20. Cumbay MG, Watts VJ. Novel regulatory properties of human type 9 adenylate cyclase. J Pharmacol Exp Ther 2004;310:108-115.
21. al-Nakkash L, Simmons NL, Lingard JM, Argent BE. Adenylate cyclase activity in human pancreatic adenocarcinoma cell lines. Int J Pancreatol 1996;19:39-47.
22. Durie PR, Forstner GG. Pathophysiology of the exo-

crine pancreas in cystic fibrosis. J R Soc Med 1989;82 (Suppl 16):2-10.

23. Courtois G, Morgan JG, Campbell LA, Fourel G, Crabtree GR. Interaction of a liverspecific nuclear factor with the fibrinogen and alpha 1-antitrypsin promoters. Science 1987;238:688-692.
24. Drewes T, Senkel S, Holewa B, Ryffel GU. Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. Mol Cell Biol 1996;16:925-931.
25. Li D, Duell EJ, Yu K et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. Carcinogenesis 2012;33:1384-1390.
26. Pierce BL, Ahsan H. Genome-wide "pleiotropy scan" identifies HNF1A region as a novel pancreatic cancer susceptibility locus. Cancer Res 2011;71:4352-4358.
27. Wu L, Aster JC, Blacklow SC, Lake R, Artavanis-Tsakonas S, Griffin JD. MAML1, a human homologue of Drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. Nat Genet 2000;26:484-489.
28. Tonon G, Modi S, Wu L et al. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nat Genet 2003;33:208-213.