

論文の内容の要旨

論文題目 Genetic Mutation Analysis of Human Pancreatic Cancers using
Targeted Capture and Massively Parallel DNA Sequencing

(ターゲットゲノムシーケンス法を用いた膵臓癌における遺伝子変異解析)

氏名 王 凌华 (オウ リンカ)

Pancreatic cancer has become the 4th leading cause of cancer-related death in developed countries. It has proven to be one of the most devastating and lethal forms of all human cancers. Less than 5% of the patients can survive beyond 5 years after diagnosis. Despite the efforts of researchers and clinicians over the past 30 years, the survival rate of pancreatic cancer has not improved substantially. It is well known that cancer arises through accumulation of mutations in DNA that can confer selective advantage to the cells in which they arise. Over 90% of the pancreatic cancer cases were considered to be caused by somatically acquired mutations.

1) Purpose

The research purpose of this study is to get a better understanding of the biology of human pancreatic cancers through comprehensive scanning of somatic mutations using targeted whole-exome enrichment and next-generation sequencing technologies.

2) Materials and methods

We analyzed a total of 15 pancreatic tumor-normal pairs. The primary pancreatic cancer tissues contain a large fraction of contaminated non-neoplastic cells. To remove the contamination non-neoplastic cells and facilitate the detection of somatic mutations, we passaged the microdissected primary tumor tissues *in vitro* as cell lines and extracted the DNA and RNA for mutation analysis.

Targeted whole-exome enrichment was performed using Agilent SureSelect Human All Exon Kit V1.0. The sequencing libraries were prepared using a paired-end DNA sample prep kit from Illumina. The enriched genomic DNA was applied to Illumina flow cell, and paired-end 76-nucleotide-long reads were generated using the Illumina Genome Analyzer Ix Platform. The high-quality sequencing reads were mapped to the human reference genome (hg18) using BWA algorithm and the variants were called using SAMtools and Pindel algorithms. To pick out the high confident somatic variants, a

series of rigorous filters and rules were applied to the data set.

The cDNA sequencing libraries were prepared using a paired-end mRNA Sequencing Sample Prep Kit from Illumina. Paired-end 76-nucleotide-long reads were generated using Illumina Genome Analyzer IIX Platform. All high-quality reads were aligned to the human reference genome hg18 using TopHat and MapSplice aligner. The point mutations were called using SAMtools.

Genome-wide SNP genotyping was performed using the Affymetrix high-resolution Human SNP Array 6.0. The SNPs were genotyped using the Birdseed v2 module of the Affymetrix Genotyping Console software GTC 4.0.1. The CN status of each Affymetrix marker was assigned using the GIM algorithm.

The methylation status of MLH1 promoter of all pancreatic cancers was quantitatively measured using MassARRAY. 500ng genomic DNA was bisulfite converted using an EZ DNA Methylation Kit. Bisulfite-treated DNA was PCR amplified and the PCR product was transcribed by in vitro transcription prior to cleavage using RNase A.

Genomic DNA extracted from the tumor and matching normal samples were used to study MSI status using the consensus “Bethesda” panel of fluorescence-labeled markers. The output data files were analyzed by GeneMapper Software Version 4.0. Determination of MSI status was made according to the presence of mutant alleles in tumor DNA compared with matched normal DNA.

To validate the somatic mutations by capillary sequencing, the oligo primers were designed to amplify the genome fragments containing the candidate nucleotide mutations from tumor DNA and the matched normal DNA. PCR products were evaluated on a 2% agarose gel, purified and sequenced in both directions using Big Dye Terminator reactions and subsequently loaded on an ABI 3130xl capillary sequencer.

The statistical significance (P value) was calculated by Student's t-test when the data are normally distributed or by the nonparametric Wilcoxon signed-rank test when the data are not normally distributed. P values <0.05 were considered to be statistically significant.

3) Results

On average, 44.2 million (6.64 gigabases) of high-quality reads were generated per sample, and 88.3% of the reads were uniquely mapped to the human reference genome with expected insert size and proper orientations. 68.4% of the uniquely-mapped proper read pairs were mapped to the whole-exome targets. Per exome, 96.9% of the targeted bases were covered at least once, and 83.4% of the targeted bases were covered by at least 10 reads.

Using whole-exome sequencing, we identified a total of 1,520 somatic mutations in 1,359 unique protein-coding genes, including 39 nonsense substitutions, 836 missense substitutions, 423 synonymous substitutions, 49 substitutions in UTR regions, 137 small

and medium sized frame-shift insertions and deletions (indels), and 36 in-frame indels. Nearly 90% of the mutations were base substitutions and over 10% of them were indels. The length of indels varies from 1 base to 29 bases. Using genome-wide SNP 6.0 array, we identified 19 recurrent focal homozygous deletions. The genetic loci at CDKN2A and SMAD4 were frequently deleted in pancreatic cancers.

Among the 1,359 genes identified with somatic mutations, 56 genes shown to be recurrently mutated in two and more tumors. KRAS, CDKN2A, TP53 and SMAD4 were frequently mutated in pancreatic cancers. Over 70% of the genes carrying recurrent mutations suggested to be novel.

We found that the mutation rates of somatic substitutions and indels varied significantly among 15 tumor samples. Integrated analysis of whole-exome sequencing, DNA copy number alterations, gene expression levels and promoter methylation levels suggested that the increased mutation rates was correlated to the DNA copy number loss in MLH1, the essential member of DNA mismatch repair pathway. The tumors kept both alleles of MLH1 (MLH1-ROH) showed normal level of expression and modest level of somatic mutations, while the tumors lost one allele of MLH1 (MLH1-LOH) showed significant reduction of the gene expression and elevated mutation rate for somatic indels (10.6-fold, $P=0.005$), and a tumor lost both alleles of MLH1 (MLH1-HD) showed no expression of this gene and dramatically increased mutation rate for both indels (71.4-fold) and base substitutions. Some of the frame-shift indels were detected in well-characterized cancer-related genes, such as TP53, SMAD4, BRCA2 and TGFBR2.

In the MLH1-HD tumor, besides a dramatically higher mutation rate of each type of substitutions, we detected a characteristic mutation spectrum. Our data show that C:G to T:A transition is the predominant in this cancer type and markedly increased in MLH1-HD tumor, especially at non-CpG sites. We also observed a higher rate of A:T to G:C, C:G to A:T transitions and A:T to T:A transversions. However, the frequency of C:G to G:C and A:T to C:G transversions were quite low in all tumors analyzed regardless of the level of MLH1 expression.

Whole-exome sequencing shows advantages in evaluation of genomic instability over the conventional method, which failed to distinguish the MLH1-LOH tumors with “intermediately unstable” microsatellites from the MLH1-ROH tumors with “stable” microsatellites.

4) Discussions

Our analyses show that the mutation rates vary widely across tumors due to differential expression of MLH1, which caused by distinct copy number changes in this gene. In mammals, the MLH1 protein is an essential component of the MMR complex. MLH1 deficiency causes the accumulation of uncorrected indel mutations, particularly in highly repetitive DNA sequences, such as microsatellites, leading to microsatellite

instability.

In sporadic cancer, it was previously thought that mutations in DNA mismatch repair (MMR) genes are recessive, i.e. even when one allele of the MMR genes was inactivated by mutations, the remaining allele would still be able to maintain the genomic stability. In this study, however, a significantly increased mutation rate of somatic indels was identified in MLH1-LOH tumors using whole-exome sequencing technology. Our data suggests MLH1 haploinsufficiency, i.e. a single wild-type allele of MLH1 is not enough to maintain genomic stability.

An earlier study generated mice with a null mutation of MLH1 gene and measured the MMR activity *in vitro* using the cell-free extracts from the mouse embryo-derived fibroblast (MEF). They found that the errors in the reporter gene were repaired 2.3-fold less efficiently in MEF extracts of *mlh1*^{+/-} mice compared to that of *mlh1*^{+/+} mice. Although the MMR activity was measured *in vitro* using a single reporter gene, the observations could support our argument that loss of single allele of MLH1 may lead to genomic instability in cancer.

In this study, some of the frame-shift indels were detected in well-characterized cancer-related genes, such as frame-shift indels of TP53 were detected in two MLH1-LOH tumors and frame-shift indels of TGFBR2 were detected in another MLH1-LOH tumor. Frame-shift indel was also detected in BRCA2. These data suggested the allelic loss of MLH1 could be a driver mutation for pancreatic carcinogenesis.

5) Conclusions

In summary, we report the dataset of directly sequenced human pancreatic cancer exomes. Whole-exome sequencing identified a set of novel genes that recurrently mutated in the cancer samples. We clearly demonstrated the genomic instability and the characteristic mutation profiles in a genome-wide manner in the tumor with complete MLH1 deficiency. We notably found a significantly elevated mutation rate of coding indels in MLH1-LOH tumors and detected frame-shift indels in well-characterized cancer genes in these tumor samples, suggesting MLH1 haploinsufficiency and the potential contribution of MLH1 to pancreatic carcinogenesis.