International Journal of Pharmacy and Biological Sciences-IJPBS™ (2019) 9 (1): 469-482
Online ISSN: 2230-7605, Print ISSN: 2321-3272
Research Article | Biological Sciences | Open Access | MCI Approved
UGC Approved Journal

Pathway Study of Gene Variants of Pancreatic Cancer Detected by Whole Exome Sequencing Data Analysis

M. L. Patil*, S. B. Madagi and J.Hoskeri

[1]Maheswari L Patil, Research Scholar, [2]Prof. Shivakumar B Madagi, Professor and [3]Dr. Joy Hoskeri, Asst. Professor, Department of Bioinformatics, Science Block, Jananashakti Campus, Akkamahadevi Women's University, Torvi Vijayapura-586108, Karnataka, INDIA.

Received: 8 Oct 2018/ Accepted: 10 Nov 2018/ Published online: 01Jan 2019
Corresponding Author Email: navi.maheswari@gmail.com

Abstract

Whole Exome sequencing, an application of next generation sequencing involves sequencing protein coding regions of genome. In current work WES data analysis was carried out on human pancreatic cancer genome using Ubuntu linux platform where the potential gene variants were identified. The human pancreatic genome sequence pairs from ENA database ERR232253, ERR232254 and ERR232254 with accession number were used for analysis where quality check of samples was performed using FastQC tool. Followed by the alignment of quality checked samples with reference genome hg38 using Bowtie2, resulting in SAM format file and was further converted to BAM format using SAMtools. The BAM file is sorted and the duplicates were removed using Picard tool, then generation of VCF file using BCF tools that predicts the possible gene variants found in the samples. The result revealed that gene HRNR showed average of 21 mutations indicating potential gene variant in pancreatic cancer. Also among the nonsynonymous mutations in samples 97 genes are found to be common, possessing 5 and more mutations. The work showed domain analysis, involvement of genes in biological process, pathway analysis and gene function. Investigation showed that 18 genes affected Protein Biosynthesis and 26 were specific to Nucleus tissue location.

Keywords

Biomarker, NGS, no synonymous mutation, pancreatic cancer, WES.

*****

INTRODUCTION

Pancreatic cancer is currently the fourth leading cause of cancer. Pancreatic cancer is common in age group 60 to 80; also there is increased incidence in diabetes [1-4] and chronic pancreatitis patients [5]. The malignant tumor of the pancreas is Ductal adenocarcinoma, it is often referred to as pancreatic

cancer [6]. In Pancreatic ductal adenocarcinoma (PDAC) the 5-year survival rate of around 7% and 1 year survival rate of 10-28% hence this is also called most lethal types of cancer [7]. Currently early diagnosis and prevention of the pancreatic cancer has become difficult, because pancreatic cancer patients rarely exhibit symptoms [8]. Based on the stage of the cancer, surgery is considered to be most effective treatment but it's only for 20% of patients who are fit for it [9]. After the surgery the most common treatment for pancreatic cancer is adjuvant chemotherapy with gemcitabine or S-1, an oral fluoropyrimidine derivative, or oxalipaltin; also if the patients who are not surgical candidate then FOLFIRINOX (fluorouracil, folinic acid [leucovorin], irinotecan and oxaliplatin) and gemcitabine plus nanoparticle albumin-bound paclitaxel (nab-paclitaxel) can be used for the treatments[9]. The genes V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), cyclin-dependent kinase Inhibitor 2A (CDKN2A), Tumor protein p53 (TP53) and SMAD family member n°4 (SMAD4) are four major genes in pancreatic cancer [9]. With the development of DNA sequencing, the most promising approach Next generation sequencing used for the analysis, where NGS providing information to develop non-invasive biomarkers and novel effective treatment for PC. NGS analysis on pancreatic cancer confirmed common mutations such as KRAS and TP53, and revealed genomic alterations that may be responsive to targeted agents [10].

DNA sequencing involves the process of figuring out the precise order of the four bases found in DNA. One of the high throughput applications is next-generation method of DNA sequencing. NGS reveals human genomic information; also helps to elucidate the function of the genome, in turn help to provide therapeutic remedies by developing personalized medicine [11]. NGS for DNA methylation sequencing solves problem with low sample throughput, short read length and low quantity [12]. The common application is Whole Exome Sequencing [13], other application involve microorganism sequencing [14], Targeted sequencing [15]. NGS provided way to find genomic alterations occurring within tumors also CNV (copy number variants) which is helpful for cancer-genome analysis to improve diagnosis and treatment. [16].

K-ras mutation is frequently detected mutation in pancreatic cancer [17, 18] and this K-*ras* mutation can be used as potential biomarker for the cancer early detection [19]. Plectin-1 expressed highly in the PanIN-3 tissues can be used as novel biomarkers [20]. Tissue factor pathway inhibitor (TFPI) can be used as candidate plasma biomarker [21]. Epigenetic factors ppENK [22, 23], RASSF1A [24], Cyclin D2 [25] can be used as biomarkers for PC. CD1D, KCNK12, CLEC11A, NDRG4, IKZF1, PKRCB and KRAS DNA are methylation markers strongly connected with pancreatic cancer [26]. miRNA can be used as non-invasive biomarker in early detection of PC [27-28]. CA19-9 most extensively biomarker for PC [29-30]

Most cancer cases begin with a mutation in the DNA. Cancer of the pancreas is a genetic disease which means that it is caused by changes (mutations) in DNA. The germline mutations in BRCA1 and BRCA2 affect the pancreatic cancer [31, 32] and mutation rates are around 5% [33]. The mutation of the KRAS2 oncogene is major mutation for pancreatic cancer, which is found in > 90% of patients [8, 34, 35]. TP53 is tumor suppressor gene mutated 70% in pancreatic cancer [36]. The p16/*CDKN2A*, also known as INK4A another tumor suppressor gene and there is 98% of mutations occurs in pancreatic cancer [37]. Mutations of SMAD4 a tumor suppressor gene occur in around 50% of pancreatic cancer, its mutation leads to loss of protein activated [38]. Also the ATM [39, 40], PALLD [41] and FANCC [42] mutations have been identified in pancreatic cancer in small number.

The present investigation focused of Exome analysis of the pancreatic cancer samples to mark out the potential gene variants. This identification is to forecast the biological process and function, pathway that may be affected from the mutations observed in variants of gene in pancreatic condition. The gene that has highest number of mutations was believed to be biomarker for pancreatic cancer.

MATERIALS AND METHODS

Raw data collection

The raw file of pancreatic cancer ERR232253, ERR232254 and ERR232255 in fastQ format were downloaded from ENA database, that were reported to be sequenced using the illumine sequencer and were paired end type. In present study, the WES pipeline used the hardware with minimum 8GB RAM and the platform Ubuntu operating system

Quality analysis of datasets

The raw Exome datasets downloaded were subjected to quality check using FastQC tool [43], as it provides the quality of the samples. Also it informs about any problems that may arise at downstream analysis of WES. The report of QC generated in HTML format include GC content, quality scores score per position, overrepresented sequences, read length distribution, sequence duplication level where the overrepresented sequences indicate the

contamination of adapter or primer that has to be trimmed.

Preprocessing of data

The preprocessing step is to remove the adapter sequences indicating contamination that may be originated during quality check of the samples. These are low quality base reads that were likely to be present at the end of reads. The trimming of such reads and adapter removal were performed for the overrepresented sequences.

Alignment of datasets to Reference Genome

Alignment of the datasets involves, downloading the reference genome hg38 *Homo sapiens* of size 3.5GB, downloaded from UCSC genome web browser. The index is build to this reference genome using Bowtie2 [44] index program, and generates 6 files. These 6 files are used for the alignment of the sample with reference genome. There are other tools for alignment to reference genome, but Bowtie2 is faster than the other tools [45], so the current research uses Bowtie2 for mapping reads to the reference genome. Sequence Alignment Mapping (SAM) format file is generated. The Summary of the alignment was written on the console

Post-alignment Processing

The SAM file generated in previous step was converted to BAM (Binary Alignment Mapping) format using call program of SAMtools [46]. The sorted BAM file is generated using sort program of SAMtools in turn generating index file. SAMtools manipulate the resulting alignment in SAM/BAM format with its utilities such as sort, merge, indexing. Next is to eliminate the duplicates those were supposed to be introduced during PCR uneven amplification of DNA fragments. Program MarkDuplicates of Picard tool [46] is used.

Variant calling

The SAMtools mpileup program and BCF tools [47] were used for the identification of gene variants indicating variant calling step of WES. The first part SAMtools mpileup program generates file with mpileup format stores likelihood of aligned reads of samples computed by mpileup program of SAMtools. The generated mpileup file is used for generation of Variant calling format (VCF) using BCF tools call program where gene variants were identified.

Annotation of Variants

SIFT 4G annotator [48] was used to call the variants in the VCF file, where SIFT predicts the affecting protein function by substitutions of amino acid. The resulting file generated in excel format and consists of details of gene. This file was used to in the research work for further study.

RESULTS

Quality Analysis of Samples

The quality check of the samples downloaded from the ENA database was done by using FastQC tool. The results of the tool are in HTML format that includes overrepresented sequences, box plots and total sequences %GC, sequence length. The results are as in Table 1-Table 6 explains the quality of the samples used. The results of the quality check don't show any overrepresented sequences in pancreatic samples, hence no trimming of sequence has performed.

Summary of Alignment

The samples of pancreatic cancer after quality check were aligned to the reference genome hg38 downloaded from UCSC genome browser. The result of the alignment of the samples of paired end reads are summarized as below,

ERR232253_1.fastq -2 ERR232253_2.fastq

24991550 reads; of these:
24991550 (100.00%) were paired; of these:
2633971 (10.54%) aligned concordantly 0 times
17701206 (70.83%) aligned concordantly exactly 1 time
4656373 (18.63%) aligned concordantly >1 times
2633971 pairs aligned concordantly 0 times; of these:
1226925 (46.58%) aligned discordantly 1 time
1407046 pairs aligned 0 times concordantly or discordantly; of these:
2814092 mates make up the pairs; of these:
1591136 (56.54%) aligned 0 times
471465 (16.75%) aligned exactly 1 time
751491 (26.70%) aligned >1 times
96.82% overall alignment rate

The alignment of above samples are summarized with 3 sections as Concordant alignment, Discordant alignment and concordant or discordant. Firstly, in concordant alignment 17701206+4656373 = 22357579 (70.83%+18.63% = 89.46 %) reads are aligned. Secondly, in disconcordant alignment remaining 2633971 reads in which 1226925 reads are aligned discordantly. This is 46.58% of reads aligned discordantly. Thirdly, remaining alignment is concordant or disconcordant which are aligned in paired end mode. The remaining reads align as single mate or one mate aligns and other mate unaligned or may not align at all. So the last part that is Total-(concordant+ discordant). This is calculated as 24991550- (22357579 +1226925) = 1407046 reads. This alignment is single end and for paired end, reads*2 means to calculate in mates. To know the overall alignment, for paired alignment summarized as (22357579 * 2 + 1226925 * 2) + 471465 + 751491=48391964 mates, are aligned to

24991550*2 mates which is 96.8166% nearly 96.82%.

ERR232254_1.fastq -2 ERR232254_2.fastq

24553677 reads; of these:
24553677 (100.00%) were paired; of these:
2595239 (10.57%) aligned concordantly 0 times
17383897 (70.80%) aligned concordantly exactly 1 time
4574541 (18.63%) aligned concordantly >1 times

2595239 pairs aligned concordantly 0 times; of these:
1202530 (46.34%) aligned discordantly 1 time

1392709 pairs aligned 0 times concordantly or discordantly; of these:
2785418 mates make up the pairs; of these:
1578189 (56.66%) aligned 0 times
468914 (16.83%) aligned exactly 1 time
738315 (26.51%) aligned >1 times
96.79% overall alignment rate

The alignment of above samples are summarized with 3 sections as Concordant alignment, Discordant alignment and concordant or discordant. Firstly, in concordant alignment 17383897+4574541 = 21958438 (70.80%+18.63% = 89.43 %) reads are aligned. Secondly, in disconcordant alignment remaining 2595239 reads in which 1202530 reads are aligned discordantly. This is 46.34% of reads aligned discordantly. Thirdly, remaining alignment is concordant or disconcordant which are aligned in paired end mode. The remaining reads align as single mate or one mate aligns and other mate unaligned or may not align at all. So the last part that is Total-(concordant+ discordant). This is calculated as 24553677- (21958438+1202530) = 1392709 reads. This alignment is single end and for paired end, reads*2 means to calculate in mates. To know the overall alignment, for paired alignment summarized as (21958438*2+1202530*2) + 468914+ 738315 = 47529165 mates, are aligned to 24553677*2 = 49107354 mates which is 96.786% nearly 96.79%.

ERR232255_1.fastq -2 ERR232255_2.fastq

25613837 reads; of these:
25613837 (100.00%) were paired; of these:
2594116 (10.13%) aligned concordantly 0 times
18258111 (71.28%) aligned concordantly exactly 1 time
4761610 (18.59%) aligned concordantly >1 times

2594116 pairs aligned concordantly 0 times; of these:
1209990 (46.64%) aligned discordantly 1 time

1384126 pairs aligned 0 times concordantly or discordantly; of these:
2768252 mates make up the pairs; of these:
1530473 (55.29%) aligned 0 times
491135 (17.74%) aligned exactly 1 time
746644 (26.97%) aligned >1 times
97.01% overall alignment rate

The alignment of above samples are summarized with 3 sections as Concordant alignment, Discordant alignment and concordant or discordant. Firstly, in concordant alignment 18258111+4761610 = 23019721 (71.28%+18.59% = 89.87%) reads are aligned. Secondly, in disconcordant alignment remaining 2594116 reads in which 1209990 reads are aligned discordantly. This is 46.64% of reads aligned discordantly. Thirdly, remaining alignment is concordant or disconcordant which are aligned in paired end mode. The remaining reads align as single mate or one mate aligns and other mate unaligned or may not align at all. So the last part that is Total-(concordant+ discordant). This is calculated as 25613837- (23019721+1209990) = 1384126 reads. This alignment is single end and for paired end, reads*2 means to calculate in mates. To know the overall alignment, for paired alignment summarized as (23019721 * 2 + 1209990 * 2) + 491135 + 746644 = 49697261 mates, are aligned to 25613837*2 = 51227674 mates which is 97.01%.

Analysis of Variants

The variant analysis part include the analysis of the xls file generated from the SIFT 4G annotator. The result file is used for the annotation and interpretation of gene variants involved in pancreatic cancer. The results of the generated file has chromosome name, position, reference allele, alternate allele, Transcript ID, Gene ID, Gene Name, Region, Variant type, Reference Amino Acid , alternate Amino Acid, Amino acid position, SIFT score, SIFT median, number of sequences, dbSNP and SIFT prediction. The variant type of genes involved will either be any of these types such as Non-synonymous, Non-coding, Frame shift Deletion, Frame shift Insertion, Synonymous, Substitution, Non-Frame shift Deletion, Non-Frame shift Insertion, Start Lost, Stop Loss and Stop Gain. There are novel genes involved in causing pc, where Sample1 (ERR232253_1 and ERR232253_2) has 6,633novel genes, Sample2 (ERR232254_1 and ERR232254_2) has 6,306 and Sample3 (ERR232255_1 and ERR232255_2) has 35582 novel genes. Further here the work revealed that the nonsynonymous genes in Sample1, Sample2 and Sample3 are 7380, 7330 and 7278 respectively. The results of nonsynonymous variant type with 5 and above mutations are

depicted in supplementary tables Table 1 to Table 3. The common genes of all samples are tabulated in Table 7. The annotation and interpretation of common genes available in all samples are tabulated as in supplementary Table 4 that summarizes the protein name, domain name, functions, location and biological Pathway Process. The biological process study is carried out to know the genes affecting the biological pathway and this is depicted in the Table 8. The locations of the gene products are in Table 9.

Table 1: Fast QC Report of ERR232253_1 (referring to first strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	ERR232253_1
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	24991550
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 2: Fast QC Report of ERR232253_2 (referring to second strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	**ERR232253_2**.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	24991550
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 3: Fast QC Report of ERR232254_1 (referring to first strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	ERR232254_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	24553677
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 4: Fast QC Report of ERR232254_2 (referring to second strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	ERR232254_2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	24553677
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 5: Fast QC Report of ERR232255_1 (referring to first strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	ERR232255_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2563837
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 6: Fast QC Report of ERR232255_2 (referring to second strand of pancreatic cancer genome data collected from ENA database)

Measure	Value
Filename	ERR232255_2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2563837
Sequences flagged as poor quality	0
Sequence length	74
%GC	44

Table 7: List of Common Genes with mutations 5 and above in samples ERR232253, ERR232254 and ERR232255 of Human Pancreatic Cancer

Common Genes of Pancreatic Cancer Samples			
ERR232253 Gene Name	**ERR232254 Gene Name**	**ERR232255 Gene Name**	**Common Gene Name in Pancreatic Cancer Samples**
ABCA4	ABCA4	ABCA4	ABCA4
AKAP13	AKAP13	AKAP13	AKAP13
ALPK1	ALPK1	ALPK2	ALPK2
ALPK2	ALPK2	ANKRD12	ANKRD12
ANKRD12	ANKRD12	ATP13A5	ATP13A5
ATP13A5	ATP13A5	ATP7B	ATP7B
ATP7B	ATP7B	BRCA1	BRCA1
BRCA1	BRCA1	C17orf80	C17orf80
C17orf80	C17orf80	C2orf16	C2orf16
C2orf16	C2orf16	CAP1	CAP1
CAP1	CAP1	CASC5	CASC5

Common Genes of Pancreatic Cancer Samples			
ERR232253 Gene Name	**ERR232254 Gene Name**	**ERR232255 Gene Name**	**Common Gene Name in Pancreatic Cancer Samples**
CASC5	CASC5	CCDC110	CCDC110
CCDC110	CCDC110	CCDC168	CEP192
CCDC168	CEP192	CCDC88B	CFH
CCDC183	CFH	CDH23	COL5A3
CCDC88B	COL5A3	CEP192	CPAMD8
CEP192	CPAMD8	CFAP54	CSNK2A3
CFAP54	CSNK2A3	CFH	CYP2A7
CFH	CTU2	CLCNKB	CYP4A22
CLCNKB	CUBN	COL27A1	DNAAF1
COL5A3	CYP2A7	COL5A3	DNAH5
CPAMD8	CYP4A22	CPAMD8	ERAP1
CSNK2A3	DNAAF1	CSNK2A3	EXO1
CUBN	DNAH5	CTU2	FAM208B
CYP2A7	DOPEY2	CYP2A7	FAM220A
CYP4A22	ERAP1	CYP4A22	FAT2
DCHS2	EXO1	DCHS2	FBN3
DNAAF1	EXOC3L4	DLC1	FCGBP
DNAH11	FAM208B	DNAAF1	FYCO1
DNAH5	FAM220A	DNAH5	GPRIN2
ERAP1	FAT2	DNHD1	HEATR1
EXO1	FBN3	ERAP1	HPS4
EXOC3L4	FCGBP	EXO1	HRNR
FAM208B	FYCO1	FAM208B	HSPG2
FAM220A	GPRIN2	FAM220A	IL1RL1
FANCA	HAP1	FANCA	INADL
FAT2	HEATR1	FAT2	KIAA1551
FBN3	HELZ2	FBN3	KIF20B
FCGBP	HPS4	FCGBP	KRT32
FSIP2	HRNR	FSIP2	KRTAP10-1
FYCO1	HSPG2	FYCO1	KRTAP10-10
GPRIN2	IGFN1	GPRIN2	LAMA5
HAP1	IL1RL1	HEATR1	LGALS8
HEATR1	INADL	HPS4	LINS
HPS4	KIAA1551	HRNR	MACF1
HRNR	KIF20B	HSPG2	MKI67
HSPG2	KRT32	IGFN1	MOCOS
IGSF22	KRTAP10-1	IL1RL1	MOV10L1
IL1RL1	KRTAP10-10	INADL	MUC16
INADL	KRTAP12-2	KIAA1551	MUC3A
KIAA1551	LAMA5	KIF20B	MXRA5
KIF20B	LAMC3	KRT32	MYOM3
KRT32	LGALS8	KRTAP10-1	MYPN
KRTAP10-1	LINS	KRTAP10-10	NEIL3
KRTAP10-10	MACF1	LAMA5	NLRP1
LAMA5	MAP2K3	LAMC3	NUP210
LGALS8	MERTK	LGALS8	OR1L6
LINS	MKI67	LINS	OR2W3
MACF1	MOCOS	MACF1	OR4L1
MAGEB16	MOV10L1	MAGEB16	OR52E6
MAP2K3	MUC16	MERTK	OR5H6
MKI67	MUC3A	MKI67	OR8D4

Common Genes of Pancreatic Cancer Samples			
ERR232253 Gene Name	**ERR232254 Gene Name**	**ERR232255 Gene Name**	**Common Gene Name in Pancreatic Cancer Samples**
MOCOS	MXRA5	MOCOS	PARP4
MOV10L1	MYOM3	MOV10L1	PCDHA4
MUC16	MYPN	MUC16	PCDHA9
MUC3A	NEIL3	MUC3A	PCNT
MXRA5	NLRP1	MXRA5	PIGQ
MYOM3	NSUN4	MYOM3	PIKFYVE
MYPN	NUP210	MYPN	PKD1L1
NEIL3	OBSCN	NEIL3	PLEKHG4B
NLRP1	OR1L6	NLRP1	PRAMEF2
NSUN4	OR2W3	NUP210	PSMD13
NUP210	OR4L1	OBSCN	PTX4
OR1L6	OR51A2	OR11G2	QRICH2
OR2L8	OR52E6	OR1L6	RAET1E
OR2W3	OR5H6	OR2W3	RP1
OR4L1	OR8D4	OR4L1	RP1L1
OR52E6	OVGP1	OR51A2	SIGLEC1
OR5H6	PARP4	OR51M1	SIGLEC12
OR8D4	PCDHA4	OR52E6	SLC22A14
PARP4	PCDHA9	OR5H6	SLC9C1
PCDHA4	PCNT	OR8D4	SNX19
PCDHA9	PIGQ	OVGP1	SPTB
PCNT	PIKFYVE	PARP4	SYNE1
PIGQ	PKD1L1	PCDHA4	SYNE2
PIKFYVE	PLEKHG4B	PCDHA9	TEP1
PKD1L1	PRAMEF2	PCNT	TG
PLEKHG4B	PREX1	PIGQ	TLR10
PRAMEF2	PSMD13	PIH1D1	TNC
PSMD13	PTX4	PIKFYVE	TTN
PTX4	QRICH2	PKD1L1	ULK4
QRICH2	RAET1E	PLEKHG4B	UMODL1
RAET1E	RP1	PRAMEF2	USH2A
RP1	RP1L1	PSMD13	WDR90
RP1L1	SIGLEC1	PTX4	ZNF208
SIGLEC1	SIGLEC12	QRICH2	ZNF469
SIGLEC12	SLC22A14	RAET1E	ZNF534
SLC22A14	SLC35G4	RNF213	ZNF717
SLC9C1	SLC39A4	RP1	
SNX19	SLC9C1	RP1L1	
SPTB	SNX19	SIGLEC1	
STAB1	SPTB	SIGLEC12	
SYNE1	SPTBN5	SLC22A14	
SYNE2	SYNE1	SLC39A4	
TEP1	SYNE2	SLC9C1	
TG	TEP1	SNX19	
TICRR	TG	SPTB	
TLR10	TLR10	SPTBN5	
TNC	TNC	SYNE1	
TTN	TTN	SYNE2	
ULK4	UGT2B28	TACC2	
UMODL1	ULK4	TEP1	
USH2A	UMODL1	TG	

Common Genes of Pancreatic Cancer Samples			
ERR232253 Gene Name	**ERR232254 Gene Name**	**ERR232255 Gene Name**	**Common Gene Name in Pancreatic Cancer Samples**
WDR90	USH2A	TLR10	
ZAN	WDR90	TNC	
ZNF208	ZNF208	TTN	
ZNF469	ZNF469	UGT2B28	
ZNF534	ZNF534	ULK4	
ZNF717	ZNF717	UMODL1	
ZNF804B	ZNF761	USH2A	
	ZNF804B	VCX2	
	ZNF880	WDR90	
.		ZNF208	
		ZNF469	
		ZNF534	
		ZNF717	
		ZNF804B	

Table 8: List of Biological pathways affected from the gene variants of pancreatic cancer samples

Pancreatic Cancer Biological Process				
Sl. No.	**Pathway Name**	**Number of Genes**	**Percentage of Genes**	**Genes Name**
1	Lipid Metabolism	2	2.04081633	CYP4A22, HSPG2,
2	G-protein coupled receptor signaling pathway	6	6.12244898	AKAP13, OR1L6, OR2W3, OR4L1, OR5H6, OR8D4
3	Cell Adhesion	9	9.18367347	FAT2, LAMA5, MUC16, PCDHA4, PCDHA9, PKD1L1, SIGLEC1, SIGLEC12, TNC,
4	Cell Differentiation	8	8.16326531	AKAP13, ERAP1, HSPG2, LAMA5, PRAMEF2, SLC9C1, SYNE1 , USH2A
5	Cellular Homeostasis	2	2.04081633	ATP13A5, ATP7B,
6	Cell Morphogenesis	1	1.02040816	CAP1
7	Signal Transduction	12	12.244898	AKAP13, BRCA1, INADL, PIKFYVE, RP1, RP1L1, PCNT, TG,CAP1, ERAP1, IL1RL1, OR52E6
8	Cell Division	2	2.04081633	CASC5, KIF20B
9	Protein Transport	3	3.06122449	KIF20B, MACF1, PCNT
10	Cell Proliferation	7	7.14285714	BRCA1, CSNK2A3, KIF20B, LAMA5, MKI67, PRAMEF2, TNC
11	DNA Biosynthesis	4	4.08163265	BRCA1, EXO1, MOV10L1, PARP4
12	Protein Biosynthesis	18	18.3673469	BRCA1, CASC5, CEP192, CSNK2A3, ERAP1, FYCO1, HPS4, HSPG2, KIF20B, LAMA5, PARP4, PCNT, PIKFYVE, PSMD13, SYNE2, TNC , TTN, ULK4
13	Post Translation Modification(Post-translational protein modification)	2	2.04081633	PSMD13, TNC
14	Immune Response	6	6.12244898	ERAP1, EXO1, IL1RL1, NLRP1, RAET1E, TLR10
15	Protein Phosphorylation	1	1.02040816	CSNK2A3
16	Metabolism	0	0	
17	Cell Mobility / Migration	6	6.12244898	CAP1, FAT2, LAMA5, LGALS8, MACF1, MYPN,
18	Apoptosis (apoptotic process)	5	5.10204082	AKAP13,BRCA1, NLRP1, PRAMEF2, UMODL1
19	Cell Cycle	6	6.12244898	CEP192, EXO1, MKI67, PCNT, KIF20B,BRCA1

Pancreatic Cancer Biological Process				
Sl. No.	**Pathway Name**	**Number of Genes**	**Percentage of Genes**	**Genes Name**
20	Cell-cell signaling	3	3.06122449	ERAP1, PCDHA4, PCDHA9,
21	Cell Communication	0	0	
22	Endocytosis	3	3.06122449	CAP1, HSPG2, SIGLEC1
23	Exocytosis	1	1.02040816	SNX19
24	rRNA processing	1	1.02040816	HEATR1

Table 9: Summarizes the tissue specific location of gene products of pancreatic cancer causing gene variants

PANCREATIC CANCER - LOCATION V/S NUMBER OF MUTATED GENES				
Sl. No.	**Location**	**Number of Genes**	**Percentage of Mutations**	**Genes Name**
1	Extracellular Matrix	20	20.40816327	C2orf16, CFH,COL5A3, CPAMD8, FBN3, FCGBP, HSPG2, IL1RL1, KRT32, LAMA5, MUC16, MUC3A, MXRA5, PSMD13, PTX4, RAET1E, SIGLEC1, TG, TNC, USH2A,
2	Secreted	20	20.40816327	C2orf16, CFH,COL5A3, CPAMD8, FBN3, FCGBP, HSPG2, IL1RL1, KRT32, LAMA5, MUC16, MUC3A, MXRA5, PSMD13, PTX4, RAET1E, SIGLEC1, TG, TNC, USH2A,
3	Nucleus	26	26.53061224	AKAP13, ANKRD12, BRCA1, C2orf16, CASC5, CCDC110, CSNK2A3, EXO1, FAM208B,FAT2, HEATR1, KIF20B, MKI67, MYPN, NEIL3, NLRP1, NUP210, PARP4, PSMD13, SYNE1, SYNE2, TEP1,TTN, ZNF208, ZNF469, ZNF534
4	Cytosol / Cytoplasm	19	19.3877551	AKAP13,FAM208B, ALPK2,ATP7B, BRCA1, DNAAF1, HPS4, HRNR , LGALS8, MACF1, MOV10L1, MYPN, NLRP1, PARP4, SNX19, SYNE1, SYNE2, TTN, UMODL1,
5	Cell cortex	1	1.020408163	AKAP13,
6	Cell Membrane	17	17.34693878	CPAMD8, IL1RL1, INADL, MACF1, MUC16, OR1L6, OR2W3, OR4L1, OR52E6, OR5H6, OR8D4, PCDHA4, PCDHA9, SIGLEC1, UMODL1, CAP1, FAT2,
7	Golgi Apparatus	3	3.06122449	ATP7B, MACF1, SYNE1
8	Endosome	4	4.081632653	ATP7B, FYCO1, PIKFYVE, SNX19,
9	Chromosome	2	2.040816327	BRCA1, MKI67,
10	Cytoskeleton	14	14.28571429	CEP192, DNAAF1, DNAH5, KIF20B, KRT32, KRTAP10-1, KRTAP10-10, MACF1, PARP4, PCNT, RP1, RP1L1, SYNE1, SYNE2
11	Centrosome / Acrosome	2	2.040816327	KIF20B, PCNT,
12	Lysosome	2	2.040816327	FBN3, HPS4
13	Mitochondrion	2	2.040816327	ATP7B, SYNE2
14	Proteosome	0	0	

DISCUSSION

In the current study an effort is made to analyze the mutations of genes found to be occurred in pancreatic cancer using WES. Jones et al. reported pancreatic cancer with a germline *PALB2* gene mutation by studying the germline sequences of pc and are analyzed using whole Exome sequencing [49] indicating that exomic sequencing can be used for the for the identification of genes responsible for causing pc. Based on the successful discovery of PALB2 gene mutation, Robert et al. proposed deleterious germline ATM gene mutations, using Whole Exome sequencing and suggested nearly 2.5 of familial pc are associated with this gene [50]. From the study of WES of pc Wang et al. reported that KRAS, TP53, CDKN2A, and SMAD4 are most frequently mutated genes [51]

The study of present work of WES analysis on pancreatic cancer revealed the presence of 7758 synonymous mutations in Sample1, 7764 in Sample2 and 7745 in Sampl3 representing such mutations occur when DNA sequence that codes for amino

acids in a protein sequence does not change the encoded amino acid. The presence of 7380 nonsynonymous mutations in Sampl1, 7330 in Sample2 and 7278 in Sample3 representing such mutations occur when there is alteration in amino acid sequence of a protein. The presence of 850 Frame shift Deletion in Sample1, 863 in Sample2, 814 in Sample3 and Frame shift Insertion of 1218 in Sampl1, 1149 in Sample2, 1160 in Sample3 representing that such mutations occur due to insertions or deletions (indels) of nucleotides in a DNA sequence that is not divisible by three. The 19 non Frame shift Deletion in Sample1, 34 in Sample2, 19 in Sample3; and 40 non Frame shift Insertion in Sample1, 50 in Sample2, 53 in Sample3 representing that no frame shift changes in protein coding sequence because of the mutation raised due to the insertions or deletions of 3 nucleotides or multiple of 3 nucleotides. There are 16 Start-lost mutations in Samlpe1, 17 in Sample2 and 16 in Sample3 representing that mutations occurs when the start codon prevents the usage of the original start translation site. There are 106 Stop-gain types in Samlpe1, 102 in Sample2 and 81 in Sample3 representing that mutation occurs due to the generation of the stop codon in normal sequence. There are 26 stop loss types in Sample1, 22 in Sample2 and 23 in Sample3 representing that mutation is occurs when the normal stop codon is lost in the sequence. The presence of 462 Substitution mutations in Sample1, 473 in Sample2, 491 in Sample3 representing single nucleotide is replaced by the different nucleotide. There exist type called noncoding representing mutations occurs when nucleotide do not code for protein sequences such noncoding mutations are 38572 in Sample1, 37941 in Sample2, 36121 in Sample3.

The gene variants possess 5 mutations and above in the 3Samples were selected where 120 in Sample1, 122 in Sample2 and 126 in Sample3 are found. The common genes in all three Samples were selected that were 97. Further in current work the annotation and interpretation is being carried out that revealed the genes affecting the biological pathways and the location of the genes. The current work showed some potential gene variants found to occur in pancreatic cancer genome. Some key genes considered through this investigation and reported the following. CYP4A22 is report to involve in Lipid metabolism. HSPG2 is reported to involve in Lipid metabolism, Cell Differentiation, Protein Biosynthesis and Endocytosis that playing vital role in multiple biological activities such as vascularization, AKAP13 is reported to involve in G-protein coupled receptor signaling pathway, Cell Differentiation, Signal Transduction and Apoptosis that playing vital role in assembling G-protein coupled receptors signaling complex, OR1L6, OR2W3, OR4L1, OR5H6, OR8D4 are reported to involve in G-protein coupled receptor signaling pathway, FAT2 is reported to involve in Cell Adhesion and Cell Mobility / Migration that playing vital role in epidermal cell migration. LAMA5 is reported to involve in Cell Adhesion, Cell Differentiation, Cell Proliferation, Protein Biosynthesis and Cell Mobility / Migration that playing vital role in the attachment, migration and organization of cells into tissues embryonic development. MUC16, PKD1L1, SIGLEC1, SIGLEC12 are reported to involve in Cell Adhesion, PCDHA4, PCDHA9 are reported to involve in Cell Adhesion and Cell-cell signaling. TNC is reported to involve in Cell Adhesion, Post Translation Modification, Cell Proliferation and Protein Biosynthesis that playing vital role in cell signaling, regulating cell proliferation and migration, especially during developmental differentiation and wound healing. ERAP1 is reported to involve in Cell Differentiation, Signal Transduction, Protein Biosynthesis, Immune Response and Cell-cell signaling that playing vital role in trimming of peptide for generating HLA class I-binding peptides that they can be presented on MHC class I molecule. PRAMEF2 is reported to involve in Cell Differentiation, Cell Proliferation and Apoptosis. PRAMEF2, SLC9C1, SYNE1, USH2A are reported to involve in Cell Differentiation. ATP13A5 and ATP7B are reported to involve in Cellular Homeostasis CAP1 is reported to involve in Cell Morphogenesis, Signal Transduction, Cell Mobility / Migration and Endocytosis that playing vital role in regulating filament dynamics and implicated complex developmental and morphological processes. BRCA1, is reported to involve in Signal Transduction, Cell Proliferation, DNA Biosynthesis, Apoptosis that and Protein Biosynthesis playing vital role in transcription, DNA repair of double-stranded breaks, and recombination. INADL, TG, RP1 and OR52E6 reported to involve in Signal Transduction. PIKFYVE is reported to involve in Signal Transduction and Protein Biosynthesis that playing vital role in biogenesis of endosome carrier vesicles from early endosomes. IL1RL1 is reported to involve in Signal Transduction and Immune Response, CASC5 is reported to involve in Cell Division and Protein Biosynthesis that playing vital role in correcting chromosome alignment. KIF20B is reported to involve in Cell Division and Protein Transport that playing vital role in regulating the neuronal polarization, PCNT reported to involve in Signal Transduction, Protein Transport, Cell Cycle

and Protein Biosynthesis that playing vital role in preventing premature centrosome splitting during interphase MACF1 reported to involve in Signal Transduction, Protein Transport, Cell Cycle and Cell Mobility / Migration that playing vital role in cross-links actin to other cytoskeletal proteins and also binds to microtubules. CSNK2A3 and TNC are reported to involve in Cell Proliferation and Protein Biosynthesis, MKI67 is reported to involve in Cell Proliferation and Cell Cycle that playing vital role in cell proliferation and is a prognostic marker. EXO1 reported to involve DNA Biosynthesis, Cell Cycle and Immune response that playing vital role in DNA mismatch repair ARP4 is reported to involve DNA Biosynthesis and Protein Biosynthesis, MOV10L1 is reported to involve DNA Biosynthesis, CEP192 is reported to involve Cell Cycle and Protein Biosynthesis, FYCO1, HPS4, SYNE2, TTN and ULK4 are reported to involve Protein Biosynthesis, PSMD13 is reported to involve Post Translation Modification and Protein Biosynthesis that playing vital role in maintenance of protein homeostasis by removing misfolded or damaged proteins, NLRP1 is reported to involve Immune response and Apoptosis, RAET1E and TLR10 are reported to involve Immune response, LGALS8 and MYPN are reported to involve Cell Mobility / Migration, PRAMEF2 and UMODL1 are reported to involve in Apoptosis, HEATR1 reported to involve in rRNA processing, SNX19 is reported to involve in Exocytosis that playing vital role in insulin secretion.

Mutation Study

The mutation study of the Samples revealed results as follows; Sample1 has totally 4641 genes that showed mutations below 5 and 122 genes are with 5mutations and above. Among these 122 genes 50 genes are having 5 mutations, 34 genes are having 6 mutations, 14 genes are having 7 mutations, 8 genes are having 8 mutations, 5 genes are having 9 mutations, 2 genes are having 10 mutations, 3 genes are having 11 mutations, 1 gene are having 12 mutations, 2 gene are having 13 mutations, MKI67 showed 15 mutation and HRNR showed 21 mutations. Sample2 has totally 4609 genes that showed mutations below 5 and 128 genes are with 5mutations and above. Among these 128 genes 60 genes are having 5 mutations, 32 genes are having 6 mutations, 14 genes are having 7 mutations, 3 genes are having 8 mutations, 09 genes are having 9 mutations, 3 genes are having 10 mutations, 4 genes are having 11 mutations, 1 gene are having 13 mutations, MKI67 showed 15 mutation and HRNR showed 22 mutations. Sample3 has totally 4665 genes that showed mutations below 5 and 122 genes are with 5mutations and above. Among these 129 genes 60 genes are having 5 mutations, 35 genes are having 6 mutations, 12 genes are having 7 mutations, 8 genes are having 8 mutations, 3genes are having 9 mutations, 5 genes are having 10 mutations, 3 genes are having 11 mutations, 1 gene are having 13 mutations, MKI67 showed 15 mutation and HRNR showed 22 mutations. Kim et al. concluded that Ki67 (also known as MKI67) frequently expressed in pancreatic cancer [52]. Choi et al. from his work reported that HRNR is involved in breast cancer development and malignant transformation [53].

Biological Pathways affected and Location of Genes

Further investigation also showed the number of genes affected the pathway. Out of the common genes in all three Samples, 18 genes affect the Protein Biosynthesis, 12 gene are affecting Signal Transduction, 9 genes found to affecting Cell Adhesion, 8 genes affecting Cell Differentiation, 7 genes affects Cell Proliferation, 6 genes with G-protein coupled receptor signaling pathway and Immune Response, 5 gene affects Apoptosis (apoptotic process), 4 genes affects DNA Biosynthesis, 3 gene affects Cell-cell signaling and Protein Transport, 2 genes affects Cellular Homeostasis, Cell Division and Post Translation Modification and 1 gene affecting Cell Morphogenesis , Protein Phosphorylation, Exocytosis and rRNA processing.

Also the investigation revealed the location of the gene products causing the pancreatic cancer. 26 mutated gene products was specific to Nucleus, 20 mutated gene products was specific in Extracellular Matrix, 20 mutated gene products were secreted, 19 mutated gene products were available in Cytoplasm, 17 mutated gene product occurs at Cell Membrane, 4 found in Endosome, 3 occur in Golgi apparatus, 2 mutated gene product were in Chromosome.

CONCLUSION

Pancreatic cancer begins in the tissues of pancreas and this cancer is aggressive with few symptoms until the cancer is advanced. Pancreatic cancer typically spreads rapidly to nearby organs. The current investigation revealed study of gene variants involved in causing pancreatic cancer analyzed by the WES. The study revealed the existence of number of gene mutations in samples that were totally 56447 in Sample1, 55745 in Sample2 and 53791 in Sample3. Among these, Sample1 has 7380, Sample2 has 7330 and Sample3 has 7278 no synonymous mutations that were selected in present work. Also the results revealed that the 4641, 4609 and 4665 genes in Sample1, Sample2 and Sample3 were having less

than 5 mutations indicating, these genes were non potential gene variants. The results of analysis reported 122, 128 and 122 genes were non-synonymous mutations in Sample1, Sample2 and Sample3 respectively that posses 5 and above mutations; also 97 gene variants were common in all the 3 Samples. The domain analysis, involvement of genes in biological process, pathway analysis and gene function were studied for the common genes.

The study showed HRNR has average of 21 mutations indicating it can be considered as biomarker for human pancreatic cancer. Also, 18 genes were affected Protein Biosynthesis and 26 mutated gene product were specific to Nucleus tissue location.

REFERENCES

1. Everhart J, Wright D., *Diabetes mellitus as a risk factor for pancreatic cancer*. A meta-analysis, JAMA, *273*:1605–1609, (1995).
2. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, et al., *Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies*. Br J Cancer, *92*:2076–2083, (2005).
3. Giovannucci E, Harlan DM, Archer MC, et al., *Diabetes and cancer: a consensus report*. Diabetes Care, *33*:1674–1685, (2010).
4. Li D., *Diabetes and pancreatic cancer*, Mol Carcinog, *51*:64–74 (2012).
5. Raimondi S, Lowenfels AB, Morselli-Labate AM, Maisonneuve P, Pezzilli R., *Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection*. Best Pract Res Clin Gastroenterol, *24*:349-58, (2010).
6. Bardeesy N, DePinho RA., *Pancreatic cancer biology and genetics.* Nat Rev Cancer, *2*:897–909, (2002).
7. Jansen RJ, Tan XL, and Petersen, *GM: Gene-byenvironment interactions in pancreatic cancer: implications for prevention*. Yale J Biol Med, *88*: 115–126, (2015).
8. Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, Neale RE, Tempero M, Tuveson DA, Hruban RH, Neoptolemos JP, *Pancreatic cancer*. Nature Reviews Disease Primers, 2:16022, (2016).
9. Karanikas M, Esempidis A, Chasan ZTM, et al., *Pancreatic Cancer from Molecular Pathways to Treatment Opinion*. Journal of Cancer, *7(10)*:1328-1339, (2016).
10. Andrea S. Teague, Benjamin R. Tan, Albert C. Lockhart, Joel Picus, Steven M. Strasberg, William G. Hawkins, Ryan C. Fields, David Linehan, Andrea Wang-Gillam, Next-generation sequencing in pancreatic cancer: Revealing genomic mutations beyond KRAS. Journal of Clinical Oncology, 32, (2014).
11. Chien-Yueh Lee and Yu-Chiao Chiu and Liang-Bo Wang and Yu-Lun Kuo and Eric Y. Chuang and Liang-Chuan Lai and Mong-Hsun Tsai, *Common applications of next-generation sequencing technologies in genomic research* . Translational Cancer Research, *2(1)*, (2013).
12. Aleksandra Dunisławska, Jagoda Łachmańska, Anna Sławińska, Maria Siwek, *Next generation sequencing in animal science - a review*. Animal Science Papers and Reports, *35(3)*: 205-224 (2017).
13. Rabbani B., Tekin M., Mahdieh N., *the Promise of Whole-Exome Sequencing In Medical Genetics*. Journal of Human Genetics, *59:* 5-15, (2014).
14. Garza D.R., Dutilh B.E., *From Cultured to Uncultured Genome Sequences: Met genomics and Modeling Microbial Ecosystems*. Cellular and Molecular Life Sciences, *72*:4287-4308, (2015).
15. Nijman I.J., Van Montfrans J.M., Hoogstraat M., Boes M.L., Van Der Corput L., Renner E.D., Van Zon P., Van Lieshout S., Elferink M.G., Van Der Burg M., Vermont C.L., Van Der Zwaag B., Janson E., Cuppen E., Ploos Van Amstel J., Van Gijn M.E., *Targeted Next-Generation Sequencing: A Novel Diagnostic Tool For Primary Immunode Ciencies*. Journal of Allergy and Clinical Immunology, *133*:529-534, (2014).
16. Serratì S, De Summa S, Pilato B, et al., *Next-generation sequencing: advances and applications in cancer diagnosis*. OncoTargets and therapy, 9:7355-7365, (2016).
17. Salek C, Benesova L, Zavoral M, et al., *Evaluation of clinical relevance of examining K-ras, p16 and p53 mutations along with allelic losses at 9p and 18q in EUS-guided fine needle aspiration samples of patients with chronic pancreatitis and pancreatic cancer*, World J Gastroenterol, *13*:3714–3720, (2017).
18. Zinsky R, Bolukbas S, Bartsch H, Schirren J, Fisseler-Eckhoff A., *Analysis of KRAS Mutations of Exon 2 Codons 12 and 13 by SNaPshot Analysis in Comparison to Common DNA Sequencing*. Gastroenterol Res Pract, *2010*:789363, (2010).
19. Morris JP, Wang SC, Hebrok M. KRAS, Hedgehog, *Wnt and the twisted developmental biology of pancreatic ductal adenocarcinoma*. Nature Rev Cancer, *10*:683–695, (2010).
20. . Bausch D, Thomas S, Mino-Kenudson M, et al., *Plectin-1 as a novel biomarker for pancreatic cancer*. Clin Cancer Research, *17*:302–309, (2010).
21. Balasenthil S, Chen N, Lott ST, et al., *A migration signature and plasma biomarker panel for pancreatic adenocarcinoma*. Cancer Prev Res (Phila), 4:137–149, (2010).
22. Jiao L, Zhu J, Hassan MM, Evans DB, Abbruzzese JL, Li D., *K-ras mutation and p16 and preproenkephalin promoter hypermethylation in plasma DNA of pancreatic cancer patients: in relation to cigarette smoking*. Pancreas, *34*:55–62, (2007).
23. Fukushima N, Sato N, Ueki T, et al., *Aberrant methylation of preproenkephalin and p16 genes in pancreatic Intraepithelial neoplasia and pancreatic ductal adenocarcinoma*. Am J Pathol., *160:1573*–1581, (2002).
24. House MG, Herman JG, Guo MZ, et al., *Aberrant hypermethylation of tumor suppressor genes in pancreatic endocrine neoplasms*. Ann Surg., 238:423–431, (2003).
25. Parsi MA, Li A, Li CP, Goggins M., *DNA methylation alterations in endoscopic retrograde cholangiopancreatography brush samples of patients with suspected pancreaticobiliary disease*. Clin Gastroenterol Hepatol., *6*:1270–1278, (2018).
26. Kisiel JB, Raimondo M, Taylor WR, et al., *New DNA Methylation Markers for Pancreatic Cancer: Discovery, Tissue Validation, and Pilot Testing in Pancreatic Juice*. Clinical cancer research : an official journal of the American Association for Cancer Research, *21(19)*:4473-4481, (2015).
27. Hernandez YG, Lucas AL., *MicroRNA in pancreatic ductal adenocarcinoma and its precursor lesions*. World J Gastrointest Oncol, *8*: 18–29, (2016).
28. Herreros-Villanueva M, Bujanda L., *Non-invasive biomarkers in pancreatic cancer diagnosis: what we need versus what we have*. Annals of Translational Medicine, *4(7)*:134, (2016).
29. Ferri MJ, Saez M, Figueras J, et al., *Improved Pancreatic Adenocarcinoma Diagnosis in Jaundiced and Non-Jaundiced Pancreatic Adenocarcinoma Patients through the Combination of Routine Clinical Markers Associated to Pancreatic Adenocarcinoma Pathophysiology*. PLoS ONE, *11(1)*:e0147214, (2016).

30. Sven H Loosen, Ulf P Neumann, Christian Trautwein, Christoph Roderburg, and Tom Luedde, *Current and future biomarkers for pancreatic adenocarcinoma Tumor biology*, *39(6)*, (2017)
31. Iqbal J, Ragone A, Lubinski J, Lynch HT, Moller P, Ghadirian P, Foulkes WD, Armel S, Eisen A, Neuhausen SL, Senter L, Singer CF, Ainsworth P, et al., *The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers*. Br J Cancer, *107*:2005–9, (2012).
32. Van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, Hoogerbrugge N, Verhoef S, Vasen HF, Ausems MG, Menko FH, Gomez Garcia EB, Klijn JG, Hogervorst FB, van Houwelingen JC, van't Veer LJ, et al., *Cancer risks in BRCA2 families: estimates for sites other than breast and ovary*. J Med Genet., 42:711–9, (2005).
33. Holter S, Borgida A, Dodd A, Grant R, Semotiuk K, Hedley D, Dhani N, Narod S, Akbari M, Moore M, Gallinger S., *Germline BRCA Mutations in a Large Clinic-Based Cohort of Patients With Pancreatic Adenocarcinoma*. J Clin Oncology, *33*:3124–9, (2015).
34. Iovanna J, Mallmann MC, Gonçalves A, Turrini O, Dagorn J-C, *Current Knowledge on Pancreatic Cancer*. Frontiers in Oncology, *2*:6, (2012).
35. Jonas Cicenas, Kotryna Kvederaviciute, Ingrida Meskinyte, Edita Meskinyte-Kausiliene, Aiste Skeberdyte and Jonas Cicenas, *KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 Mutations in Pancreatic Cancer*, Cancers, *9*:42, (2017).
36. Hwang, R.F.; Gordon, E.M.; Anderson, W.F.; Parekh, D., *Gene therapy for primary and metastatic pancreatic cancer with intraperitoneal retroviral vector bearing the wild-type p53 gene*. Surgery, *124*:143–150, (1998).
37. Schutte M.,Hruban, R.H. Geradts J., Maynard, R., Hilgers, W., Rabindran, S.K., Moskaluk, C.A., Hahn, S.A., Schwarte-Waldhoff, I., Schmiegel, W. et al., *Abrogation of the RB/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas*. Cancer Research, *57*: 3126–3130, (1997).
38. De Bosscher, K., Hill C.S. Nicolas, F.J., *Molecular and functional consequences of SMAD4 c-terminal missense mutations in colorectal tumour cells*. Biochem. J., 379: 209–216, (2004).
39. Kim H, Saka B, Knight S, Borges M, Childs E, Klein A, Wolfgang C, Herman J, Adsay VN, Hruban RH, Goggins M., *Having pancreatic cancer with tumoral loss of ATM and normal TP53 protein expression is associated with a poorer prognosis.* Clin Cancer Res. *20*:1865–72, (2014).
40. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, et al., *ATM mutations in patients with hereditary pancreatic cancer*. Cancer Discovery, *2*:41–6, (2012).
41. Ghiorzo P, Fornarini G, Sciallero S, Battistuzzi L, Belli F, Bernard L, Bonelli L, Borgonovo G, Bruno W, De Cian F, Decensi A, Filauro M, Faravelli F, et al., *CDKN2A is the main susceptibility gene in Italian pancreatic cancer families*. J Med Genet., *49*:164–70, (2012).
42. Couch FJ, Johnson MR, Rabe K, Boardman L, McWilliams R, de Andrade M, Petersen G, Germ *line Fanconi anemia complementation group C mutations and pancreatic cancer*. Cancer Research, *65*:383–6, (2015).
43. Andrews S., *FastQC*, Babraham Bioinformatics, Cambridge, UK, (2010).
44. Wu TD, Nacu S., *Fast and SNP-tolerant detection of complex variants and splicing in short reads*. Bioinformatics, *26*:873–81, (2010).
45. Langmead B. and S. L. Salzberg, *Fast gapped- read alignment with Bowtie 2*. Nature Methods, *9(4)*:57– 359, (2012).
46. Takeshi Ogasawara, Yinhe Cheng, Tzy-Hwa Kathy Tzeng, *Sam2bam: High-Performance Framework for NGS Data Preprocessing Tools*, PLOS ONE, *11(11)*:e0167100, (2016).
47. Heng Li. BCF tools-Samtools [Online], (2018).
48. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC, *SIFT missense predictions for genomes. Nat Protocols*, *11*: 1-9, (2016).
49. Jones, S., Hruban, R. H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D. W., et al., *Exomic Sequencing Identifies PALB2 as a Pancreatic Cancer Susceptibility Gene*. Science (New York, N.Y.), *324(5924)*: 217, (2009).
50. Roberts, N. J., Jiao, Y., Yu, J., Kopelovich, L., Petersen, G. M., Bondy, M., et al., *ATM mutations in hereditary pancreatic cancer patients.* Cancer Discovery, *2(1)*:41–46, (2012).
51. Wang, L., Tsutsumi, S., Kawaguchi, T., Nagasaki, K., Tatsuno, K., Yamamoto, S, Aburatani, H. et al., *Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1haplo insufficiency and complete deficiency*. Genome Research*, 22(2)*: 208–219, (2012).
52. Kim, H., Park, C. Y., Lee, J. H., Kim, J. C., Cho, C.-K., & Kim, H. J., *Ki-67 and p53 expression as a predictive marker for early postoperative recurrence in pancreatic head cancer*. Annals of Surgical Treatment and Research, *88(4)*:200–207, (2015).
53. Choi J, Kim DI, Kim J, Kim BH, Kim A. *Horner in is involved in breast Cancer progression*. J Breast Cancer, *19*:142–7, (2016).