博士論文

Genomic Analysis of Pancreatic Juice DNA Assesses Malignant Risk of Intraductal Papillary Mucinous Neoplasm of Pancreas

（膵嚢胞性腫瘍患者由来膵液のセルフリーDNA エキソーム情報解析）

ラウル ニコラス マテオス ラモス

Table of Contents

ABSTRACT

Intraductal papillary mucinous neoplasm (IPMN) of pancreas has a high risk to develop into invasive cancer or co-occur with malignant lesion. For this reason, it is important to assess its malignant risk by a less-invasive approach. An ideal material for this purpose would be Pancreatic juice cell-free DNA (PJD), but genetic biomarkers for predicting malignant risk from PJD are not yet established. Here, I performed deep exome sequencing analysis of PJD from 40 IPMN patients with or without malignant lesion as well as a novel combination of filtering techniques in order to produce a more robust and reliable outcome. In order to evaluate their potential as a malignancy marker I compared the somatic alterations and copy number alterations detected in PJD with the histologic grade of IPMN. Somatic mutations of *KRAS*, *GNAS*, *TP53*, and *RNF43* were commonly detected in PJD of IPMNs, but no association with the histological grades of IPMN was found. Instead, histologic grade was positively correlated with mutation burden (r = 0.417, P = 0.018). I was also able to observe frequent copy number deletions in *17p13* (*TP53*) and amplifications in *7q21* and *8q24* (*MYC*) in PJDs. The amplifications in *7q21* and *8q24* were positively correlated with the histologic grade and most prevalent in the cases of invasive carcinoma (P = 0.012 and 7/11; P = 0.011 and 6/11, respectively). Mutation burden and copy number alterations detected in PJD have potential to assess the malignant progression risk of IPMNs. These findings showed clinical usefulness of genomic profiling of PJD for IPMN.

1. INTRODUCTION

1.1 The pancreas

The pancreas, as described by the Johns Hopkins school of medicine [1], is a long flattened glandular organ of about 6 inches long located deep in the abdominal area surrounded by the stomach, the duodenum and the spine. The pancreas is divided into four main different regions; the tail, the body, the neck, and the head, which can also be divided into the uncinated process and the proper head [2, 3]. The superior mesenteric artery, as well as the superior mesenteric vein cross the pancreas in front of the head and behind the neck, both being extremely important vessels. This organ has two main functions; an exocrine function and an endocrine one. The exocrine, executed by the majority of the pancreatic cells, consists of the production of the production and secretion of enzymes involved in digestion by the acinar cells that are gathered and released through the main pancreatic duct, whereas the endocrine function is performed by a set of cells clustered in small regions called islets of Langerhans. This endocrine cells produce hormones like glucagon and insulin that are released to the bloodstream instead to the pancreatic ducts, regulating the glucose levels in blood (Figure 1). Among the multiple medical conditions this organ can suffer, I aimed my research towards exocrine pancreas neoplasms, which comprises more than 75% of all cancer of the pancreas [1, 4]. More specifically, the focus of my research was particular neoplasm; the intraductal papillary mucinous neoplasm.

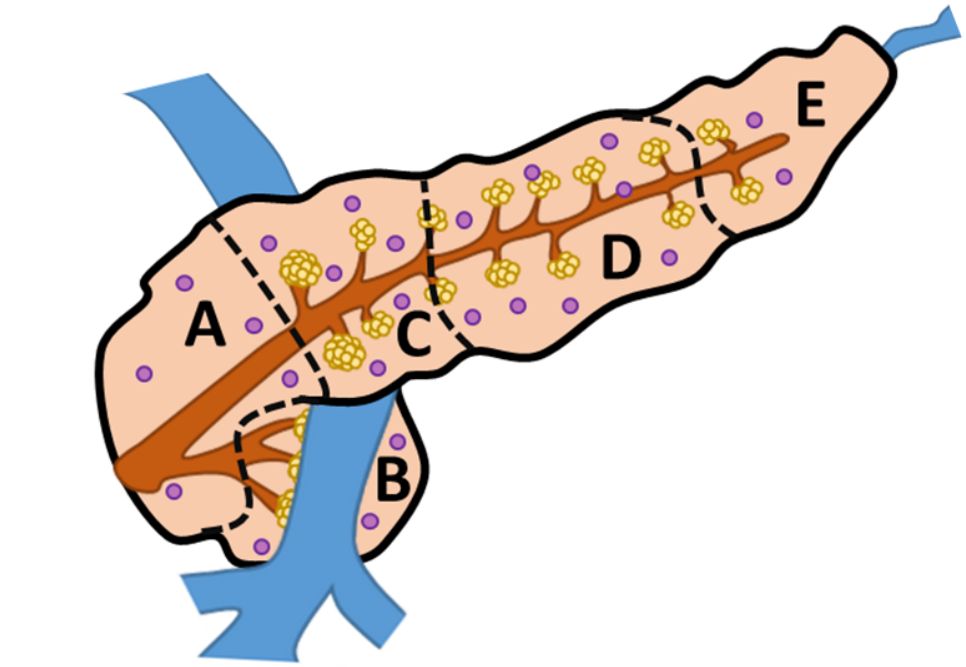


Figure 1: Illustration of the pancreas. Here the different regions of the pancreas can be observed; the head, divided in the proper head (A) and uncinated process (B), the neck (C), the body (D), and the tail (E). The clusters of acinar cells (yellow) secrete enzymes for digestion to the pancreatic duct (brown), whereas the islet of Langerhans (purple) are scattered all over the pancreas, releasing hormones into the bloodstream.

1.2 Intraductal papillary mucinous neoplasm.

Intraductal papillary mucinous neoplasm (IPMNs), described for the first time in 1982 by Ohashi et al.[5, 6] and posteriorly defined by the World Health Organization, is a pancreatic tumor with unique characteristics including hyper-production of mucin in tall columnar epitheliums, and dilatation and papillary growth inside the pancreatic ducts[7-10]. The incidence of IPMN has been rapidly increasing after the establishment of its diagnosis, and it is now understood that IPMN can be classified along the spectrum of adenoma to carcinoma, and it can also be a precursor of pancreatic cancer[11]. It has also been reported that the prognosis of this neoplasm after surgical resection is relatively better than the one of pancreatic ductal adenocarcinoma (PDAC) [12] . On the other hand, IPMN is very heterogeneous histologically and some its parts can progress from low to high-grade dysplasia, leading to invasive adenocarcinoma, which shows as poor prognosis as PDAC [13]. Additionally, PDAC is sometimes coincident with IPMN[14]. Hence, in order to take the decision of tumor resection it is clinically important to assess the risk of pancreatic cancer progression and development in IPMNs [15, 16]. There are several guidelines and recommendations that support the assessment of the risk for malignancy of IPMN by imaging methods such as magnetic resonance imaging/cholangiopancreatography, yet there is still some room for improvement. The development of other non-invasive approaches to assess the malignant potential of IPMNs based on biomarkers or genomic alterations is of utmost importance.

1.3 Cell-free DNA and liquid biopsies

Cell-free DNA, discovered by Mandel and Metais in 1948[17] is a special type of DNA that consists of highly fragmented double strands of DNA of approximately 150bp[18]. These fragments travel through different body fluids after being released by healthy as well as tumoral cells through apoptotic and/or necrotic processes (Figure 2). Cell-free DNA shed by tumor cells is a rich source of tumor-specific biomarkers and genomic analysis, as shown in studies on cell-free DNA derived from plasma [19-21], urine [22], and cerebrospinal fluid (CSF) [23], all of which can be acquired in a less invasive way than solid biopsies. The capture of this tumor-derived DNA can be effectively performed and analyzed in anatomically relevant fluids, such as urine in bladder cancer, and endocervical fluid for gynecological tumors [24]. Pancreatic juice could be an ideal material for assessing malignancy risk of IPMN and pancreatic tumors as well. In fact, the malignant potential of IPMN can be evaluated by cytological test in pancreatic juice [25]. Endoscopic retrograde cholangio-pancreatography (ERCP) and cytological test are performed in some hospitals in order to screen patients with IPMN. Furthermore, due to its liquid nature, pancreatic juice has potential to overcome the heterogeneity inherent to biopsy specimens from IPMN. Molecular analysis and characterization of pancreatic juice might be able to provide direct evidence of malignant IPMN, thus complementing the imaging.

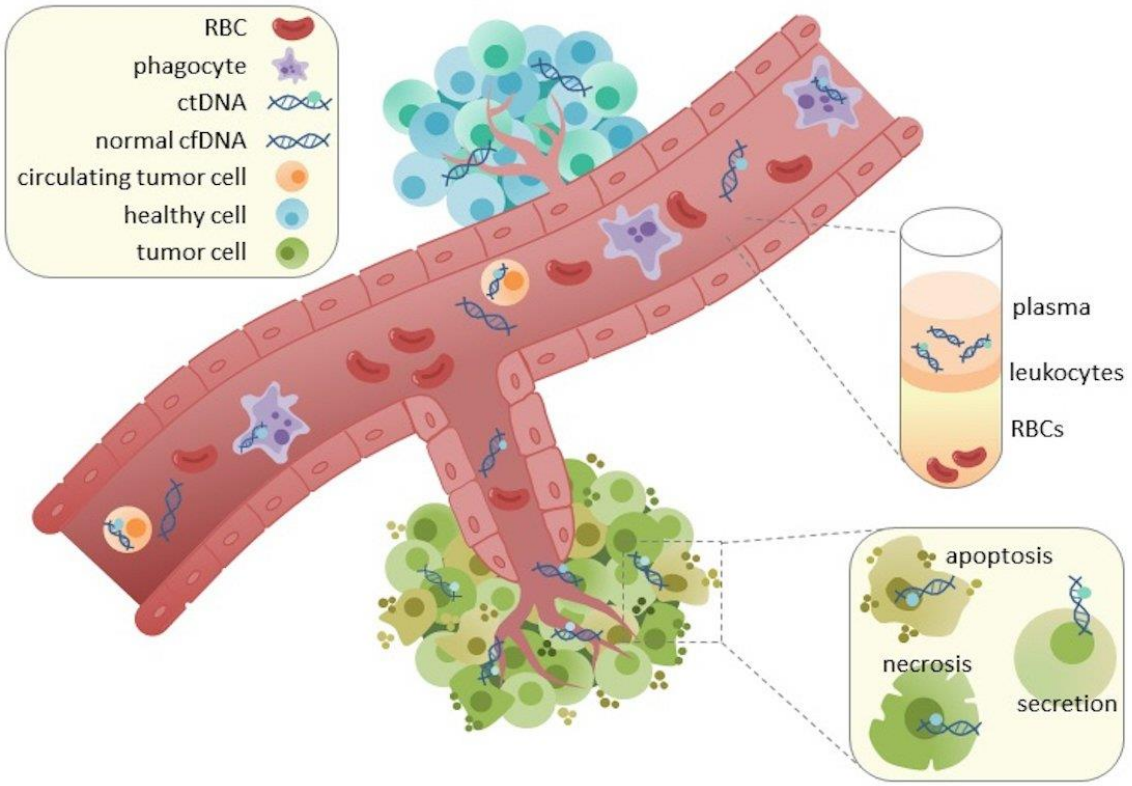


Figure 2: Cell-free DNA. cfDNA obtained from a liquid source such as plasma, can proceed from healthy cells, as well as from tumoral cells that went to apoptotic or necrotic processes, or from secretion. (Image by Racheljunewong)

1.4 Biomarkers in IPMN

Regarding genomic alterations and biomarkers, several genes and pathways are commonly altered in IPMNs, being *KRAS* one of the most studied. *KRAS*, also Kirsten ras, is an oncogene homolog of the Kristen Rat Sarcoma virus [26, 27] which encodes a protein from the GTPase superfamily and is one of the most frequently activated oncogenes[27], being its activating mutations especially prevalent in lung, colorectal, thyroid and pancreatic carcinomas. Its mutation in IPMN has also been described in different frequencies (31–86%) although its incidence does not correlate with the levels of histological grade or dysplasia [28]. Another widely known biomarker is the oncogene *GNAS*. This gene encodes the stimulatory subunit of the guanine nucleotide-binding protein (G protein), the alpha subunit[29]. *GNAS* has been reported as mutated in codon 201 in more than half (61%) of their patients with IPMN [30]. Other Oncogenes such as *BRAF*, *PI3KCA* and *TERT*, tumor suppressor genes such as *CDKN2A* or *TP53*, as well as pathways such as the Sonic Hedgehog signaling pathway have also been reported as altered in some studies [16, 31-33].

Numerous studies had analyzed pancreatic juice for driver mutations such as *KRAS* and *GNAS* [33, 34]. However, the usefulness of these mutations as markers of IPMN malignancy is still unclear, and more comprehensive genomic analysis of pancreatic juice cfDNA (PJD) may help discover better ones and overcome the intratumoral heterogeneity of IPMN. In order to achieve these goals, I performed deep exome sequencing analysis of PJDs from 40 IPMN patients.

2. MATERIALS AND METHODS

2.1Clinical samples

Pancreatic juice and blood samples from 40 IPMN patients were obtained in Wakayama Medical University Hospital and Yamanashi University Hospital. All subjects agreed with informed consent to participate in the study and the Institutional Review Boards at RIKEN and the two hospitals approved this work. The method chosen for the pancreatic juice collection was ERCP or endoscopic nasopancreatic drainage (ENPD). ERCP is a procedure where a bendable tube (endoscope) is used to examine the bile and the pancreatic ducts. The endoscope travels through the mouth and stomach until it reaches the duodenum. There, the ampulla of Vater, which is the landmark of the connection of the common bile duct and the pancreatic duct with the duodenum, is identified, and a cannula is used to go through this opening. This process is used to detect gallstones as well as possible blockages in the bile duct, but can also be used to take biopsies of solid and liquid nature (Figure 3). ENPD, on the other hand, is a procedure used to drain the pancreatic duct by inserting a drainage-tube in a similar process as ERCP and has been reported as a potential tool for treating severe acute pancreatitis [35].

All of these patients underwent surgical resection for these tumors after obtaining its pancreatic juice, and experienced pathologists in Kyoto Prefectural University of Medicine reviewed the pathological description of these resected specimens. Their clinico-pathological features are shown in Table 1 and Table S1. Regarding their

pathology, 8 (20%) had histological low-grade dysplasia (LGD), 20 (50%) had high-grade dysplasia (HGD), whereas 12 (30 %) had invasive carcinoma (INC).

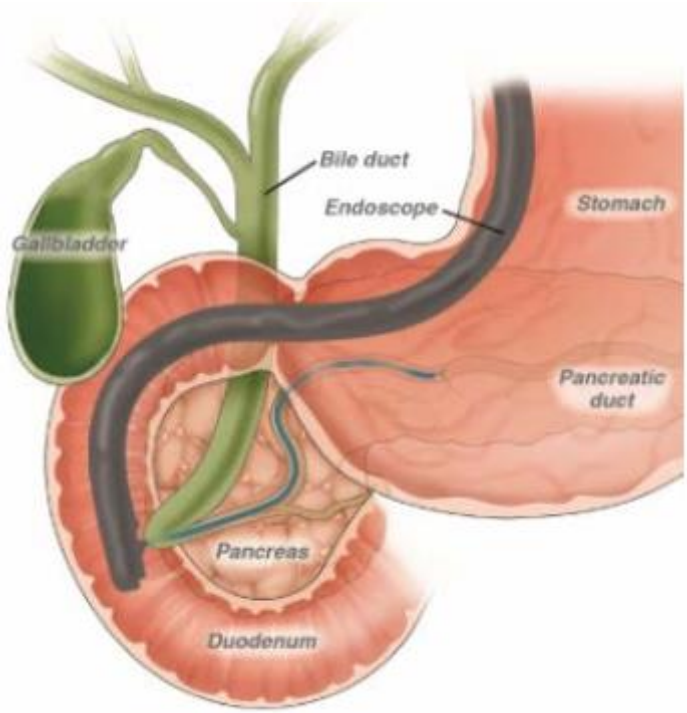


Figure 3: Endoscopic retrograde cholangio-pancreatography technique. The endoscope is introduced through the mouth until the duodenum where, through the ampulla of Vater, liquid and solid biopsies can be extracted using a cannula. (Image by the American Gastroenterological association)

Sex	Male	23
	Female	16
Histological grade	Low-grade Dysplasia	8
	High-grade Dysplasia	20
	Invasive Carcinoma	11
Subtype	Intestinal	11
	Gastric	21
	Pancreatobiliary	4
	NA	3
Macroscopic Type	Branch Duct	13
	Main Duct	11
	MIX	15
Tumor location	Head	21
	Head-tail	2
	Tail	1
	Body-tail	12
	Body	3

Table 1: Summary of clinical information of IPMNs. The sample revised and classified as PDAC and not IPMN was removed from the table.

2.2 Flowchart

In order to validate pancreatic juice as a suitable source for genomic analysis, and to find novel markers from malignancy I proceeded to perform genomic analysis of PJD, divided in two branches; mutation calling and copy number alteration analysis. The full analysis flowchart of this study is shown in Figure 4.

2.3 Exome sequencing of PJD

The exome sequencing of the samples was performed by the RIKEN SNP research center. The extraction of cell-free DNA from the pancreatic juice samples was performed using QIAamp Circulating Nucleic Acid Kit (Qiagen), whereas normal DNA was extracted from lymphocyte in blood by QIAamp DNA Blood Kit (Qiagen). The libraries were prepared with KAPA HyperPrep Kit (Kapa Biosystems) using 5-10ng input DNA. The exome capture was performed with SureSelect Human All Exon V5 (Agilent Technologies), and sequencing was performed with Illumina HiSeq2500 SBS V4. Sequencing coverage was 200-300x for PJD and 100x for normal lymphocyte DNA. Reads were mapped to GRChr37 using BWA and duplicates from PCR were removed using Picard[36].

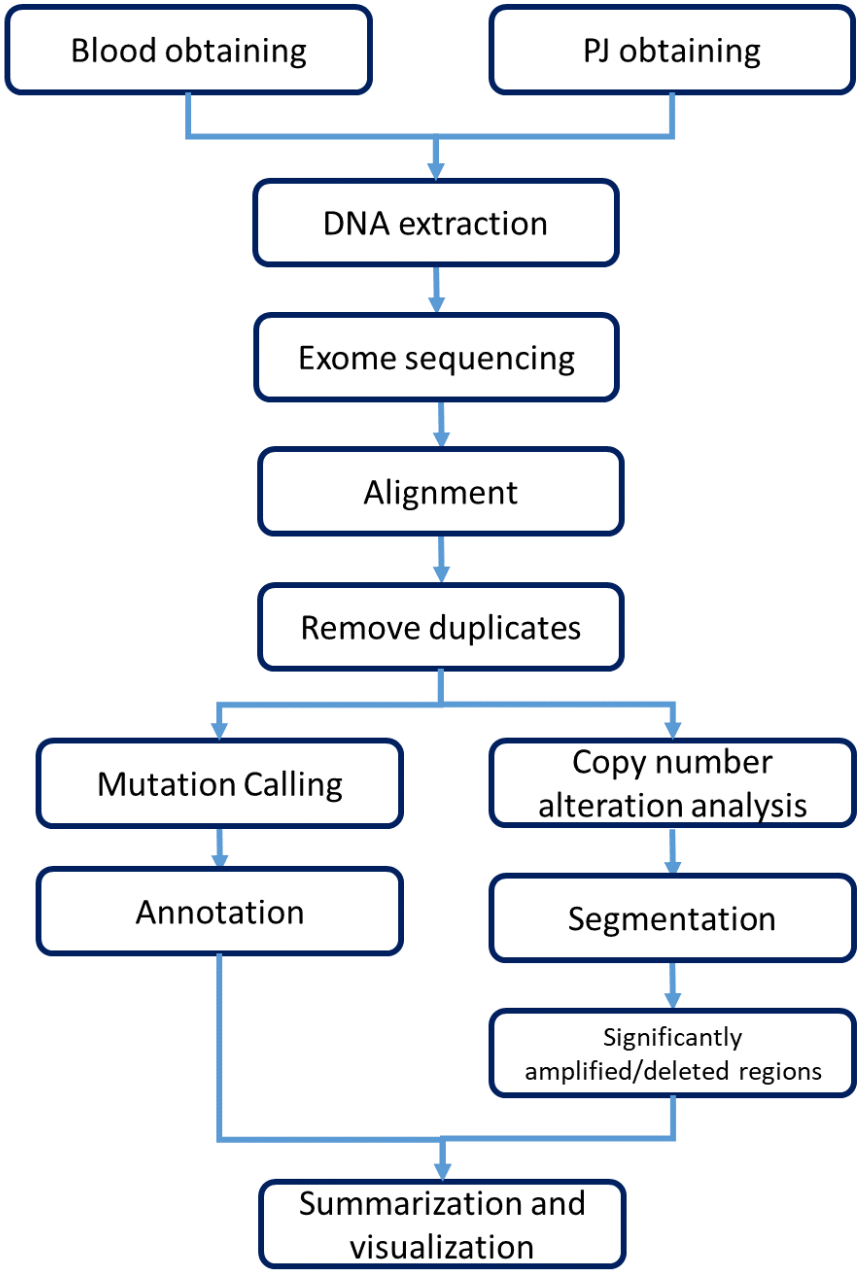


Figure 4. Workflow of my deep exome sequencing analysis for PJD. The pipeline divides into two analysis; the mutation calling and the CNA analysis. Then, the outputs were gathered for visualization.

2.4 Mutation calling

I used the tool Genomon2 Fisher mutation call [37] for the analysis of single nucleotide variation and INDELs. The minimum depth for the mutation call was 8, the minimum map quality was 20, and the minimum base quality was 15. Variants with less than 3 reads were filtered. The control maximum allele frequency was 0.1 and the disease minimum allele frequency chosen for the study was 0.01. Fisher mutation call threshold was set to 0.1, and its p-value log10 threshold was set to 1.0. The detected variants were annotated using Annovar [38] and then summarized and visualized using the R package Maftools [39].

2.5 OxoG artifacts

The generation of 8-oxoguanine (OxoG) is one of the most common artifacts produced during the library construction. The OxoG alteration can be caused by effect of heat, storage, shearing, contamination of metals or its combined effect, leading to C>A/G>T transversion, commonly formed in the middle base of CCN triplets [40]. Three lines of evidence showed that three PJD samples (W18, W21 and W24) were highly affected by the OxoG artifact. Firstly, the oxidation error rates computed by Picard CollectOxoGMetrics [36] were strikingly high in the three samples (Figure 5). Secondly, the variant allele frequency (VAF) of the three samples were lower than the other samples (Figure 6). It is known that mutations caused by the OxoG artifact have a lower VAF[40]. Finally, the trinucleotide mutation patterns of

the three samples had the characteristic configuration of OxoG artifacts (Figure S1). Due to all these evidences, I decided to exclude these three samples from the SNV analysis.

2.5 Copy number alteration (CNA) analysis

The copy number variation analysis was performed using Varscan2[41]. The ratio of tumor reads and normal reads, required for normalization, was obtained by the SAMtools function flagstats. As for the CopyNumber function, the minimum segment size was 100 bases, the maximum segment size was 100,000 and the p-value threshold was 0.005. The output obtained was then used as the input for the CopyCaller function of Varscan2. Circular binary segmentation (CBS) was posteriorly performed on this

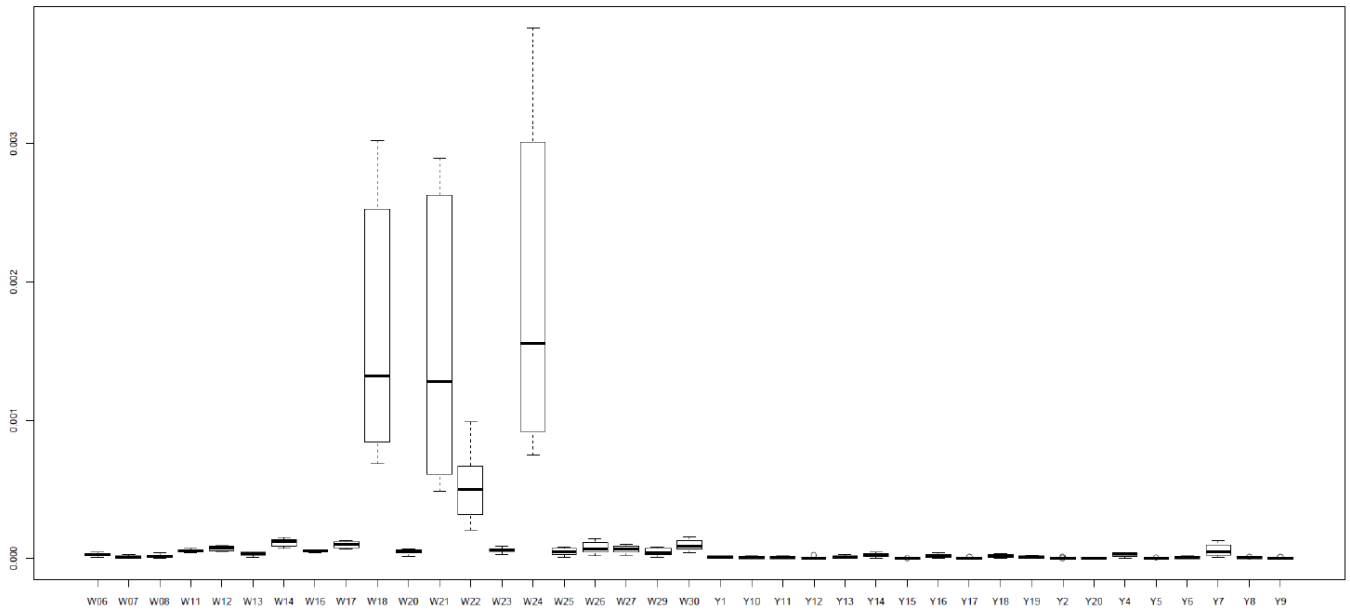
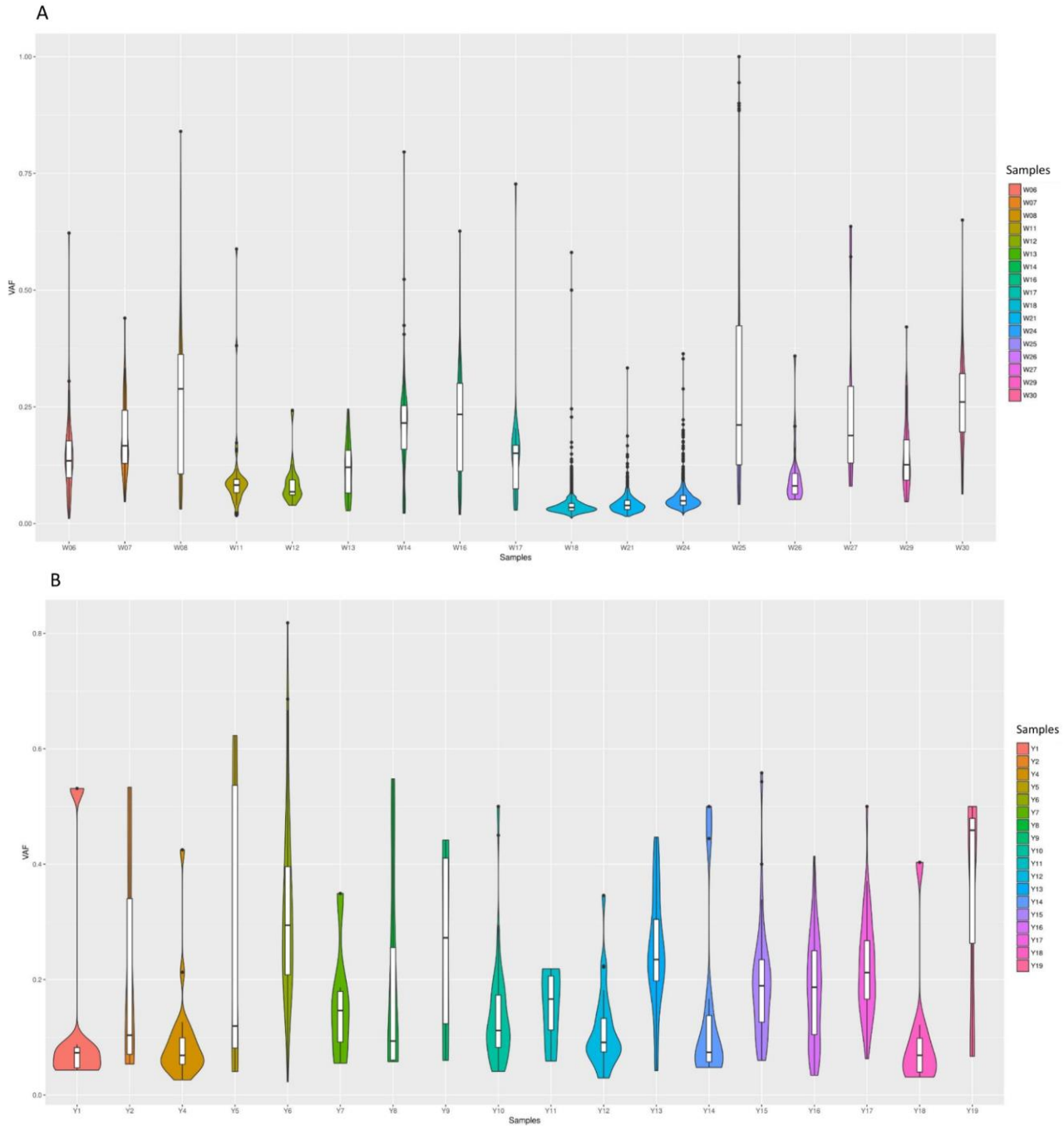


Figure 5: Oxidation error rate per sample. Some of the samples possess a strikingly different profile, which might indicate a case of OxoG artifact.



.

Figure 6: Variant Allele Frequency of Wakayama (A) and Yamanashi (B) samples. It is noticeable that the samples marked as potentially altered by the OxoG artifact are also the ones with low VAF, characteristic of this alteration.

output by using the R package DNAcopy [42] by which I also undid splits that were not at least 3 standard deviations apart. The segmentation from CBS obtained was analyzed using GISTIC2.0[43], in order to find significant amplified or deleted regions among the samples. The threshold for copy number amplifications and deletions was 1.5 and the confidence level to calculate the region containing a driver was set to 0.90.

2.5 CNA filtering: Residual Variance

One factor that should be taken into consideration in CNA analysis is the potential presence of samples with high variance in their segmentation, which produces blurry and unstable outcomes. The lack of a stable neutral value, makes the samples very difficult to normalize/adjust, set thresholds for amplifications and deletions, and finally, ends up affecting the performance of CNA analysis tools like GISTIC2.0. In order to avoid this kind of outcome, I proceeded to evaluate the level of instability by calculating the residual variance of each sample. (**Table S2**). It is important to mention that these blurred segmentations might not in fact be caused by an artifact, but by the heterogeneity of tumor subpopulations in certain samples and the capacity of liquid biopsies to overcome this heterogeneity. Different subpopulations with different alterations in different proportions could cause a mosaic-like outcome, complicating the interpretation of the results of the CNA analysis (Figure 7). Despite the fact that this could be an interesting subject to analyze, I decided the results would be more robust with the removal of this unstable samples. After analyzing the results of this evaluation, four

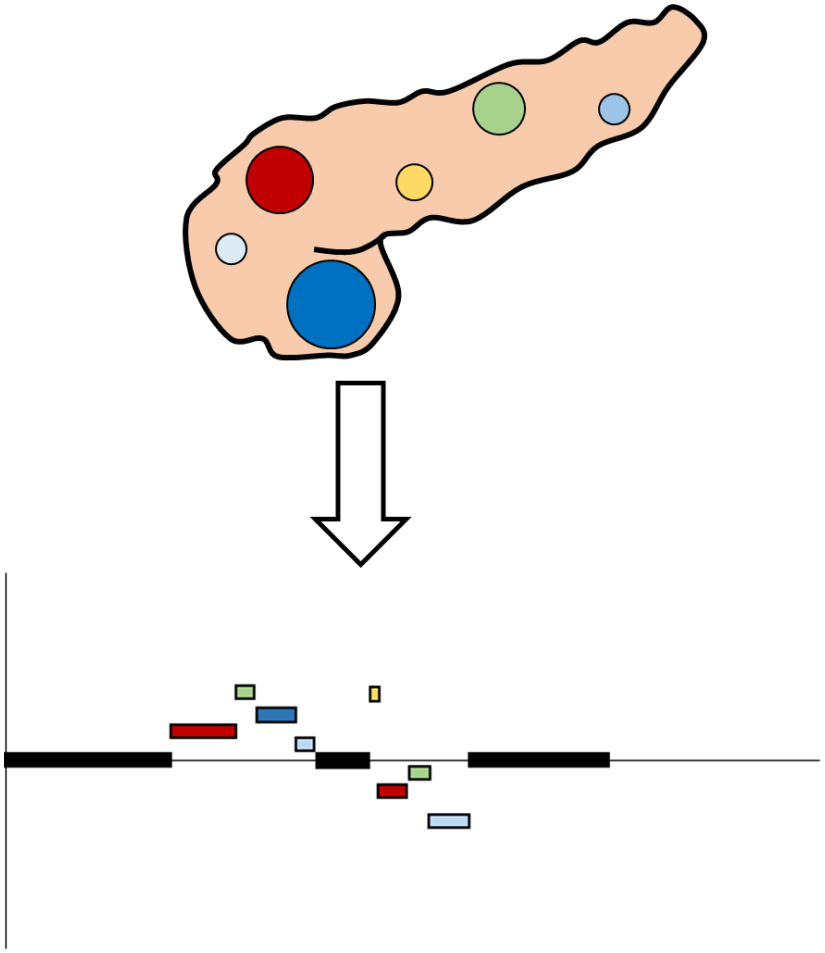


Figure 7: Pancreatic Juice and heterogeneity of IPMN: The capacity of PJD of overcoming heterogeneity might be the source of a mosaic-like outcome, caused by a mixture of different subpopulations, which can be seen as noise.

samples (W20, W22, W23 and Y20) were removed from the dataset due to their high residual variance.

2.6 Removal of intronic-exclusive segmentation

The exome capture was performed with SureSelect Human All Exon V5, which bounds the CNA to exome regions. Due to this, all the outputs from Varscan2 located exclusively in off-target regions of the exome capture should be considered unreliable. For that reason, I decided to remove these intron-only generated regions in order to improve the results of the analysis. After this filtering, I re-fused consecutive segments that shared the same log2 ratio (Figure 8). By doing so, I was able to produce a much cleaner and reliable input for GISTIC2.0.

3. RESULTS

3.1 Deep exome sequencing of PJD

To identify somatic mutations of IPMNs in PJD, I collected pancreatic juice and blood of 40 IPMN cases and extracted DNA from these samples. I constructed next generation sequencing libraries from a small amount of PJD (5-10ng) and performed deep

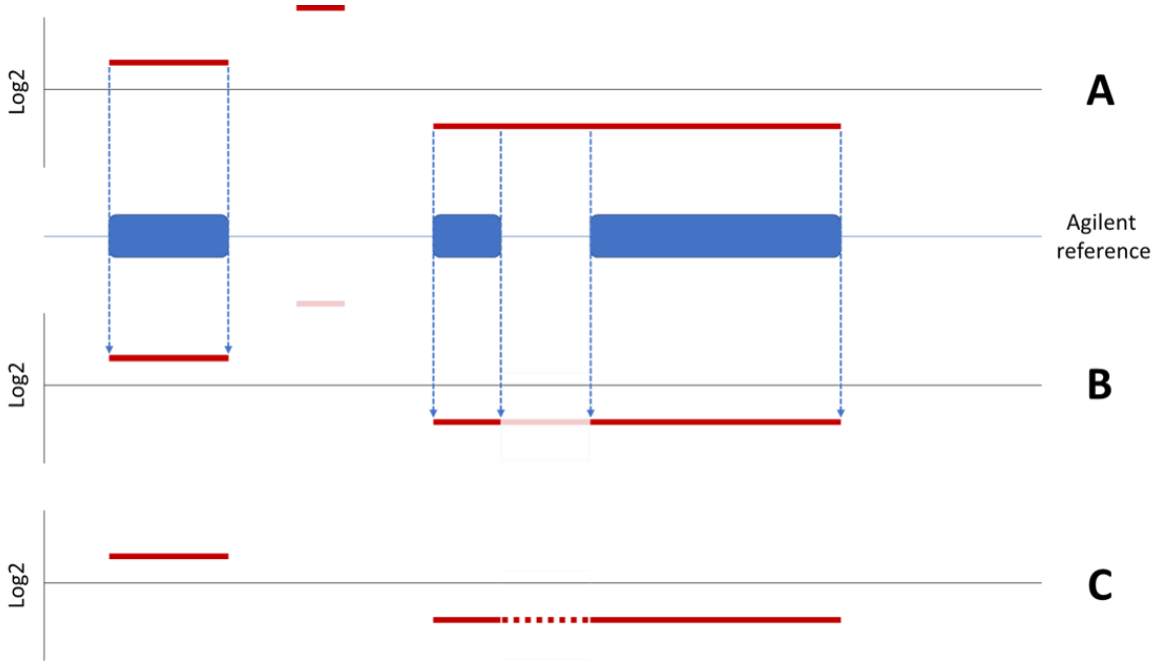


Figure 8: Removal of intronic-exclusive regions: By using the Agilent kit exome list, all regions that did not belong to the set were removed from the results. After this process (B), all consecutive segments that shared the same log ratio were re-fused, removing only intronic-exclusive segments.

exome sequencing of PJD together with blood DNA for 40 IPMN cases. After removing one case that was revised and diagnosed as PDAC and not IPMN, the median number of sequence reads were 163 million reads for PJD, and after duplication removal the median depth on target was 168x. For blood samples, the median number of sequence reads were 82 million reads, and the median depth on target was 112x after duplication removal (Table S3). The median of duplication rate was 18.7% in exome of PJD and was higher than that in blood DNA exome (5.4%). This might have been caused by low-input DNA for the library preparation (Figure S2).

I estimated the tumor-derived DNA content in PJD by doubling the median VAF in PJD samples. To reduce potential statistical error, I restricted the selection to 24 PJD samples with at least 10 somatic SNV/INDEL alterations. The median of predicted tumor-derived DNA content among those 24 PJD samples was 28.5% (Figure 9 and Table S4), indicating higher content of tumor-derived DNA in pancreatic juice of IPMN patients than those obtained in cerebrospinal fluid from patients with central nervous system metastases [44]. No significant correlation of tumor-derived DNA content with the histologic grade of IPMN was found ($p = 0.148$).

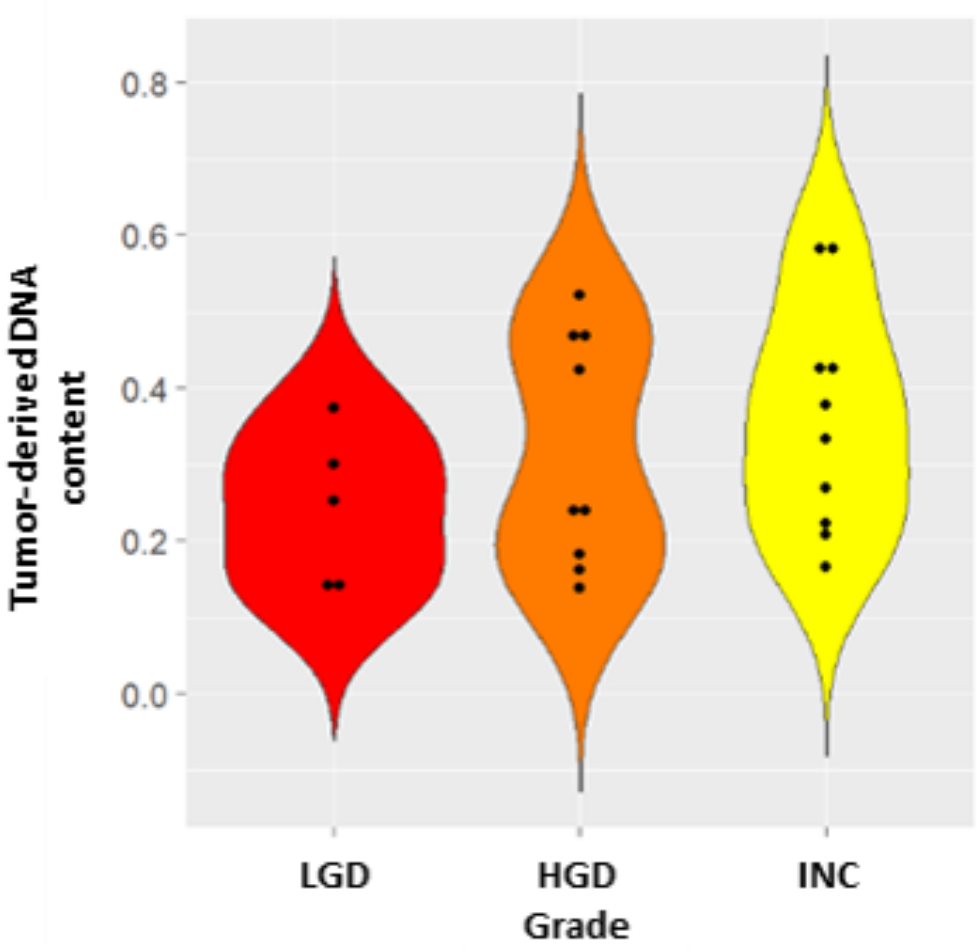


Figure 9: Violin-plot of predicted tumor-derived DNA content (purity) in PJD. PJD samples with at least 10 mutations were shown along with their IPMN grade. No correlation of purity with the grade was found (p = 0.148).

3.2 Mutation burden in PJD was associated with histologic grade of IPMN

After removing the case that was classified as PDAC and not IPMN, seven cases with OxoG and/or high residual variance were removed from 39 cases, calling somatic mutations for the remaining 32 cases. Totally, 627 somatic non-synonymous mutations affecting 561 genes were detected (Figure 10). Among them, 535 mutations (85.33%) were missense mutations, 30 (4.78%) were frame-shift deletion, 30 (4.78%) were nonsense mutations, 16 (2.55%) were frame-shift insertions, 11 (1.75%) were splice-site mutations and 5 (0.80%) were in-frame deletions. The median number of total mutations per sample was 32 (0.628/Mb), showing that the mutation burden of IPMN calculated from PJD has lower mutation rate compared to pancreatic cancer [45].

Interestingly, the whole-exome mutation burden in PJD was associated with the histologic grades of IPMN. The median number of mutations was 18 in LGD, 21.5 in HGD, and 71 in INC. Spearman correlation coefficient between the histologic grade of IPMN and the number of mutations was 0.417 ($p = 0.018$). This result shows that the mutation burden in PJD might be used as a marker for malignant potential of IPMN. On the other hand, the mutational signature had a similar distribution irrespective of grades (Figure 11). The association of mutations with other clinical features of IPMNs is shown in Figure 12A. *GNAS* mutation was associated with main duct type ($p = 0.018$) and with male ($p = 0.019$), which is consistent with previous reports [6, 46].

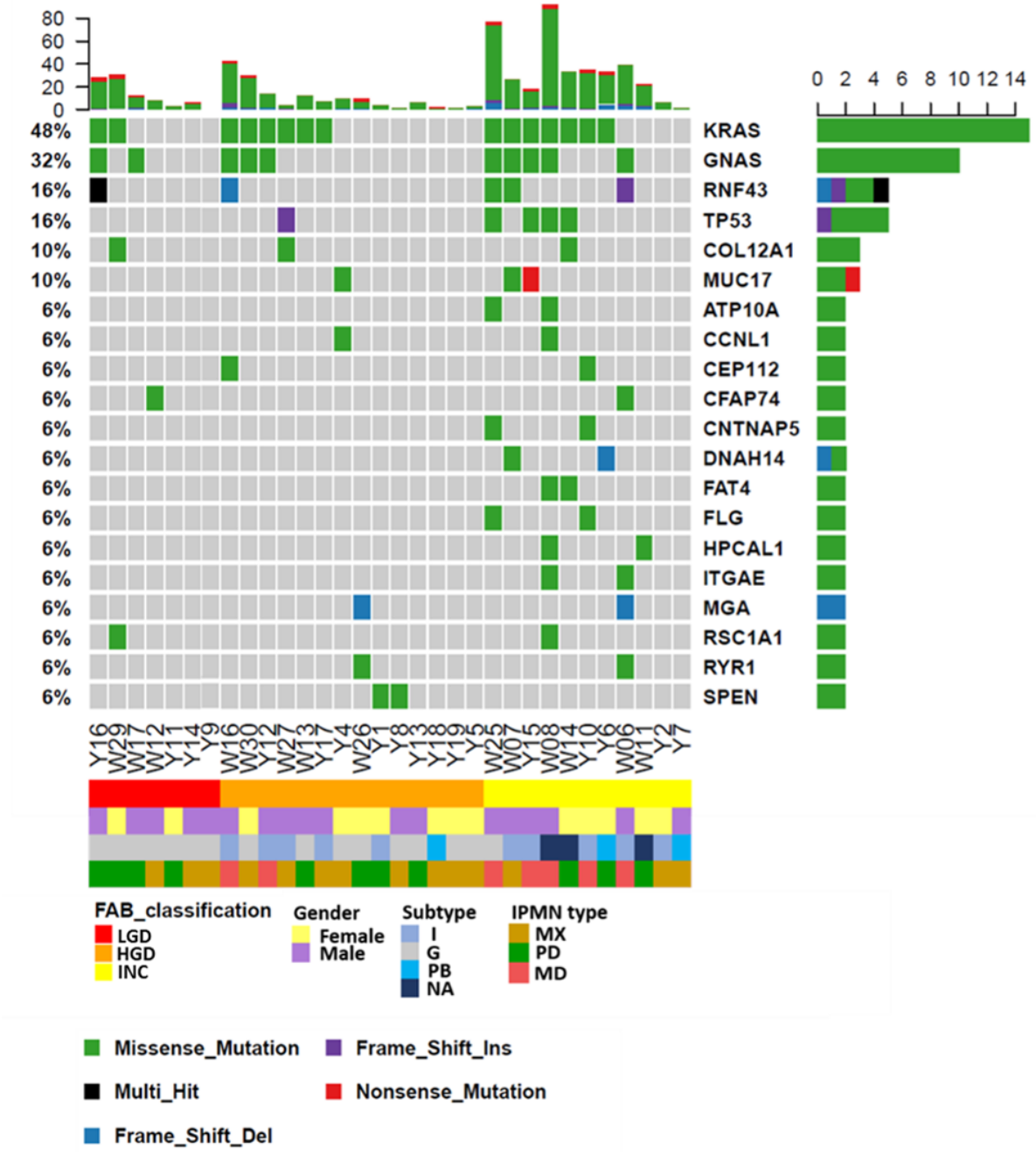
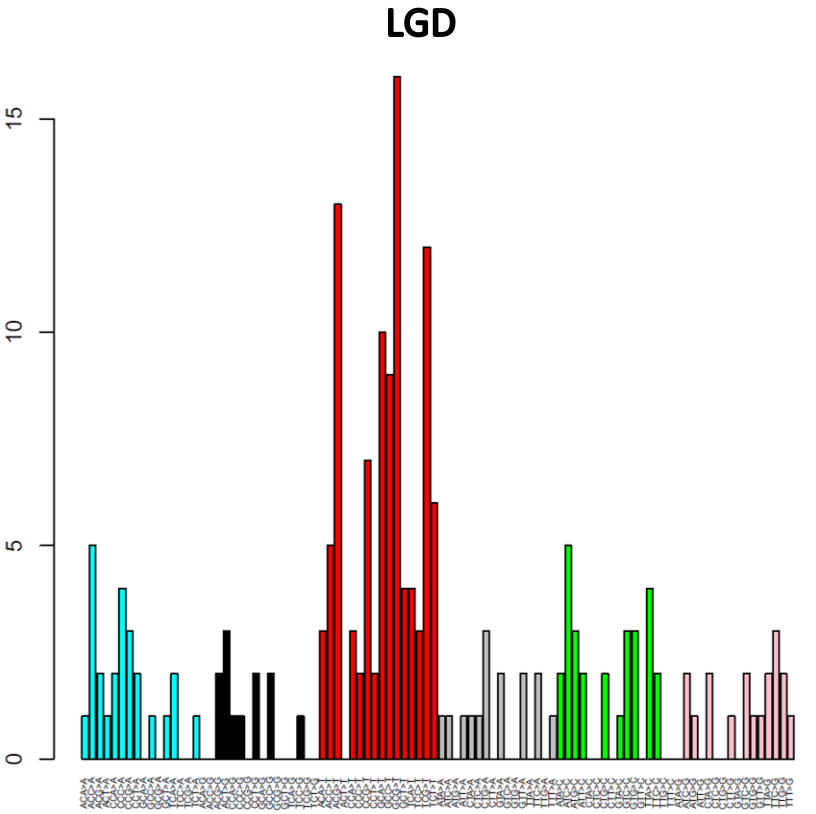
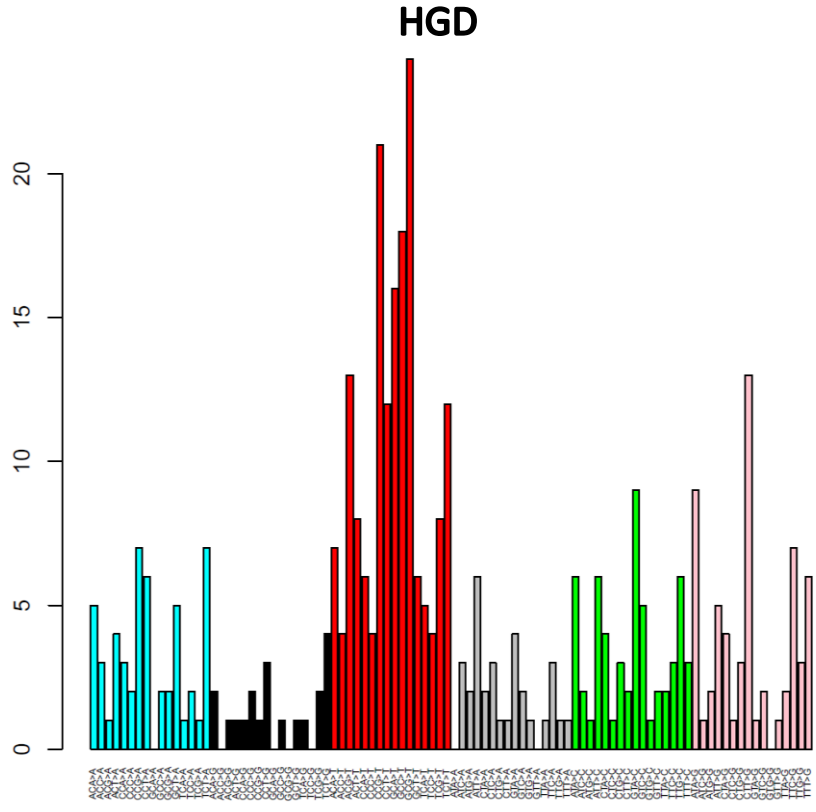


Figure 10. Somatic mutations detected in PDJ and the histologic grade of IPMN. The figure shows somatic mutations present in the dataset. Below the table, clinical features of the samples are included. Genes with at least two mutations were included. The most common mutations were in *KRAS* and *GNAS*, but no significant association with histological grade was found. Four out of five *TP53* mutations were found in samples of histologic INC.


LGD
0
5
10
15


HGD
0
5
10
15
20

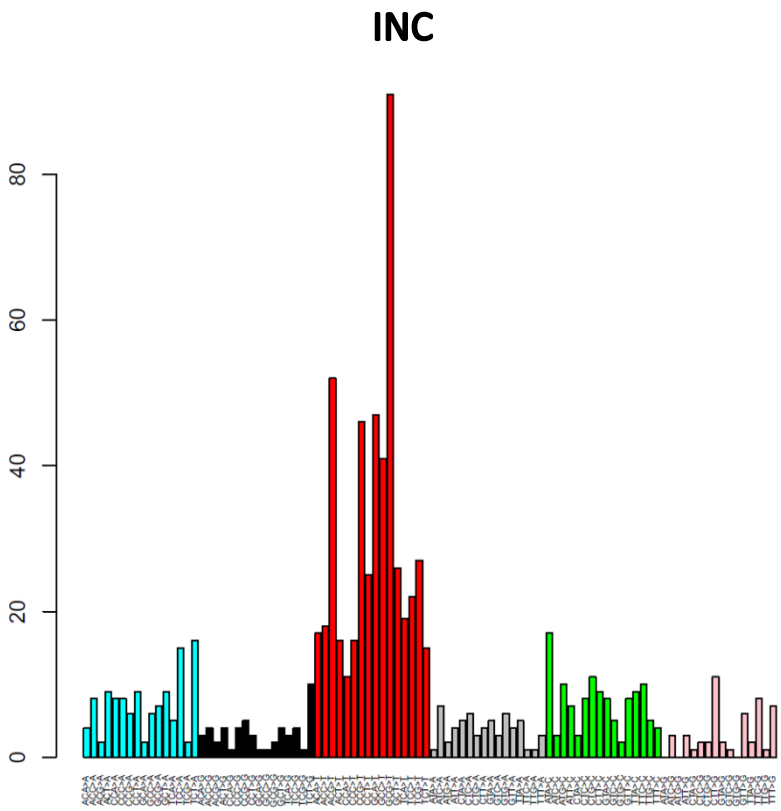


Figure 11: Mutation patterns by histological Grade: By merging the mutation pattern of each sample by grade, I was able to visualize the general distribution. Despite the increasing number of mutations, the general pattern remains similar during the development of malignancy.

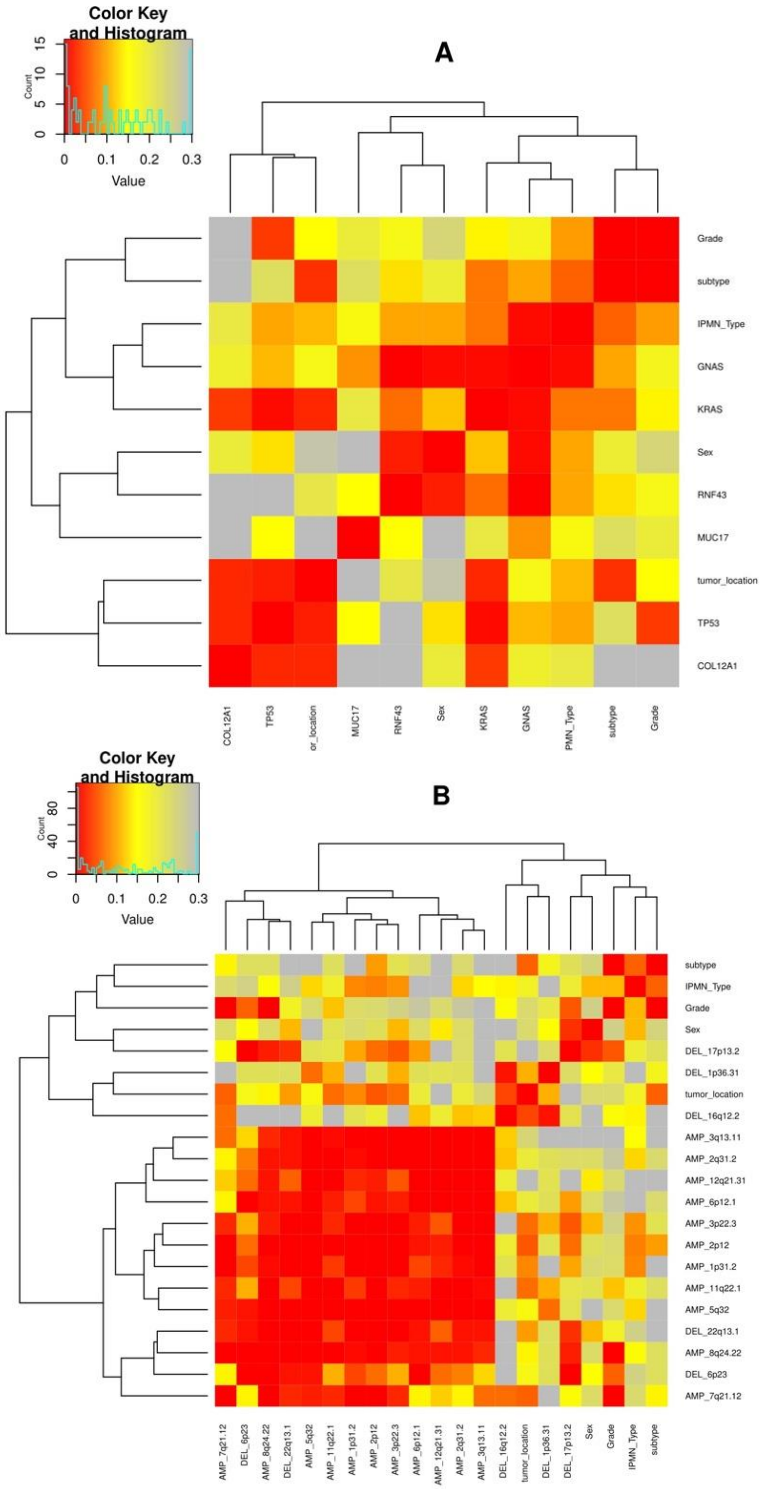


Figure 12: Association study. A: Heatmap containing the log of the p-values obtained Fisher test p-values showing the association between clinical data and SNV. B Association study between clinical data and CNA. All the p-values have been adjusted by adding one unit before the log transformation.

KRAS and *GNAS* were most frequently mutated genes (15 and 10 samples; 47% and 31%, respectively), which is also consistent with previous studies [31, 33, 46, 47]. The mutation of *TP53* and *RNF43* was found in five samples (15.63%). *RNF43* is a RING-type E3 ubiquitin ligase whose loss-of-function mutation can activate the Wnt/β-catenin signaling and is commonly mutated in IPMN and in other tumors [33, 48]. Other recurrently mutated genes were *COL12A1*, which encodes the alpha chain of type XII collagen, and *MUC17* which encodes a protective membrane-bound mucin to gut epithelial cells and has been reported as highly expressed in PDAC as well as a marker for poor prognosis [49]. Both *COL12A1* and *MUC17* were mutated in three samples.

To confirm the existence of the TP53 mutations identified in pancreatic juice, several small lesions of IPMN from two FFPE specimens were dissected (Cases W14 and W25), which were assigned by the experienced pathologist in HE-stained slides. DNA was obtained from these small lesions and the mutations of *TP53* were amplified by PCR and analyzed by the capillary sequencing. As shown in Figure 13, the two *TP53* mutations p.H214R in W14 (A) and p.R248W in W25 (B) were validated in their malignant lesions (INC), but not in the benign lesions (LGD).

We as well proceeded to validate 192 of the somatic mutations (SNVs and INDELS) found in our analysis. In order to do so, we amplified the regions of these mutations from a pool of all our samples and calculate the VAF at each location of the SNVs,

whereas the INDELs were evaluated visually by using the Interactive Genomics Viewer tool (IGV) [50]. The VAF at same locations were also evaluated in a pool of the blood samples with the purpose of evaluating the potential noise and false positives, as well as for setting a suitable threshold for our validation. In addition to the 192 somatic mutations, hotspot mutations of KRAS and GNAS were also validated from the corresponding FFPE tumor samples by capillary sequencing. Overall, 126 out of 206 mutations were validated (60.3%), being the median of the VAF in the tumor pool 0.067, which gives a description of the tumor purity of our pool.

3.3 CNAs in PJD association with IPMN histological grade

I called somatic copy number alterations (CNAs) from PJD exome data and detected eleven significantly amplified regions (*1p31.2, 2p12, 2q31.2, 3p22.3, 3q13.11, 5q32, 6p12.1, 7q21.12, 8q24.22, 11q22.1* and 1*2q21.31*) and five significantly deleted regions (*1p36.31, 6p23, 16q12.2, 17p13.2* and *22q13.1*) (Figure 14A and 14B). When analyzing the gene ontology of the genes in regions significantly deleted by amiGO [51], an enrichment in DNA related processes was observed, such as DNA modification (GO:0006304), DNA deamination (GO:0045006) (Table 2). No enriched processes were observed in amplified regions.

The detected CNAs were evaluated for their potentials as malignancy markers (Figure 14C). A significant association between the histologic grade and

amplification of two regions was observed (*7q21.12* and *8q24.22*). The *7q21* amplification was found in 1/7 (14.26%) in LGD, 2/17 (11.76%) in HGD, 7/11 (63.64%) in INC (p= 0.012 by Fisher's exact test). *KIAA1324L* is located in this amplified region, but no association with IPMNs and pancreatic cancer has been reported. The *8q24* amplification was found in 2/7 (28.57%) in LGD, 1/17 (5.88%) in HGD, 6/11 (54.55%) in INC (p = 0.011). *MYC*, located in the *8q24* amplified regions in IPMNs (Table S5), is one of the most frequently altered genes by CNAs in pan-cancer analysis [52, 53] and its amplification has also been reported in pancreatic acinar cell carcinomas [54, 55]. The deletion region *17p13,* which contains *TP53* (Table S6) was also found to co-occur with *TP53* point mutations; three out of four malignant samples (INC) with *TP53* mutations (Figure 14D). This observation is consistent with the two-hit theory of tumor suppressor genes. The association of CNAs in PJD with other clinical features of IPMNs is shown in Figure 12B.

I posteriorly applied GISTIC2.0 to benign (LGD) and malignant samples (HGD and INC) independently, leading us to a more grade-associated outcome. In benign samples, I observed a significantly deleted region (*11q21*), whereas in malignant samples, a significantly amplified band was found (*3q13.11*) and four regions were significantly deleted (*1p36.33, 16q12.2, 17p13.2* and *22q13.1*). The existence of

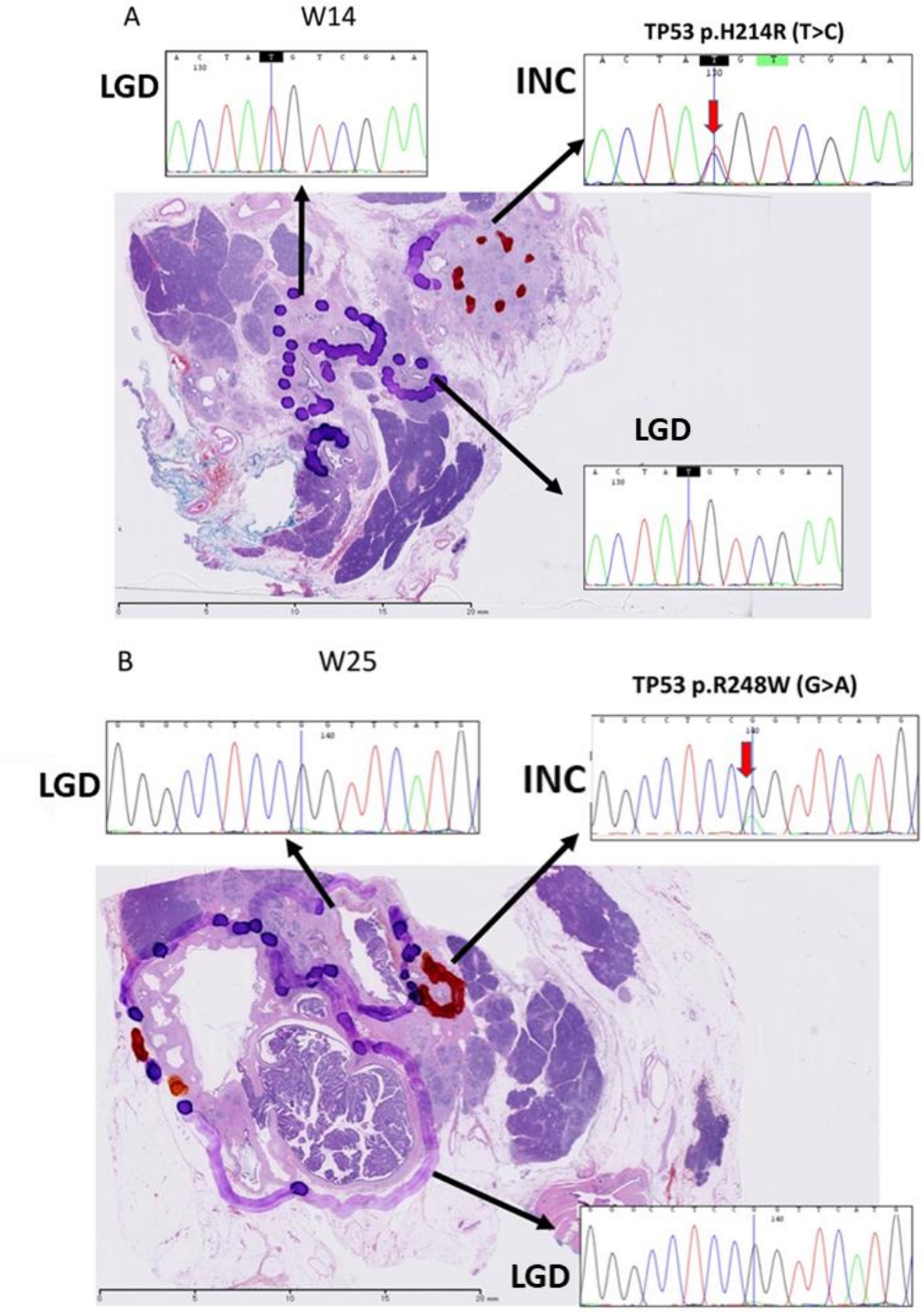


Figure 13: SNV Validation. The validation of the TP53 SNVs was done by capillary sequencing of DNA extracted from the FFPE specimens of the samples W14 (A) andW25 (B). TP53 mutations were both found in the malignant lesions (INC) and therefore validated.

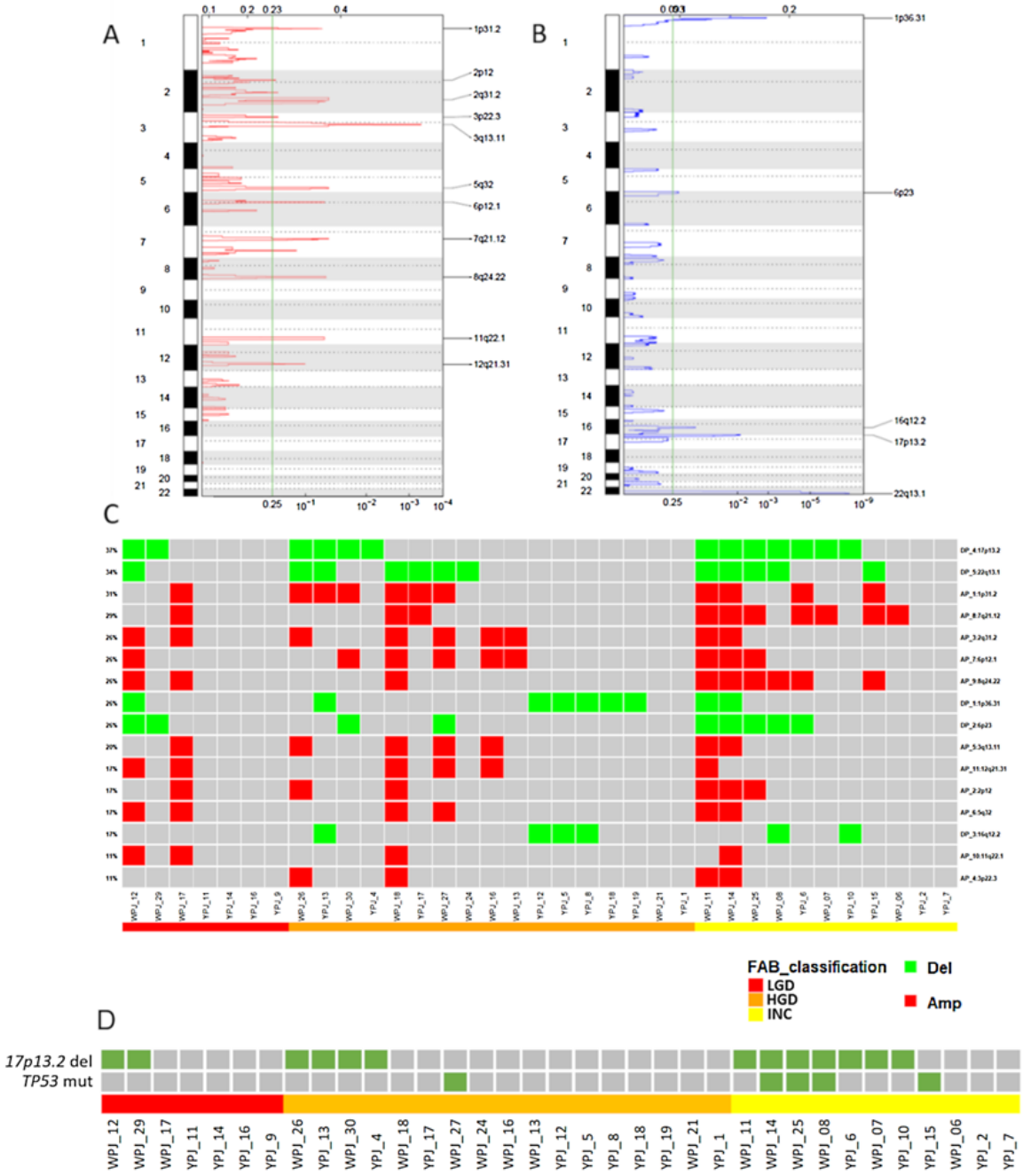


Figure 14: Summary of copy number alterations detected in PJD from IPMNs. A: Eleven significantly amplified regions detected by GISTIC2.0. B: Five significantly deleted regions detected by GISTIC2.0, including 17p13.2 (TP53). C: Significantly amplified regions (red) and significantly deleted regions (green) by patient and by grade are shown. The 7q21 amplification and 8q24 amplification (MYC) showed significant association with the histologic INC (p= 0.012 and p = 0.011, respectively, by Fisher's exact test). D: Co-occurrence of TP53 mutation and 17p13.2 deletion: The co-occurrence of both alterations are found mainly on PJD samples with histologic INC.

Analysis Type:	PANTHER Overrepresentation Test (Released 20171205)		
Annotation Version and Release Date:	GO Ontology database Released 2018-02-02		
Analyzed List:	Homo sapiens		
Reference List:	Homo sapiens (all genes in database)		
Test Type:	FISHER		
GO biological process complete	**fold Enrichment**	**raw P-value**	**FDR**
DNA cytosine deamination (GO:0070383)	49.12	4.27E-07	2.21E-03
Negative regulation of single stranded viral RNA replication via double stranded DNA intermediate (GO:0045869)	39.29	1.26E-05	1.63E-02
Cytidine metabolic process (GO:0046087)	34.38	1.56E-06	6.06E-03
Cytidine catabolic process (GO:0006216)	34.38	1.56E-06	4.85E-03
Cytidine deamination (GO:0009972)	34.38	1.56E-06	4.04E-03
DNA deamination (GO:0045006)	31.26	2.25E-06	4.98E-03
Regulation of single stranded viral RNA replication via double stranded DNA intermediate (GO:0045091)	30.56	2.67E-05	2.30E-02
Pyrimidine ribonucleoside catabolic process (GO:0046133)	28.65	3.15E-06	5.42E-03
Negative regulation of transposition (GO:0010529)	25.33	6.05E-08	9.37E-04
Regulation of transposition (GO:0010528)	25.33	6.05E-08	4.69E-04
Pyrimidine nucleoside catabolic process (GO:0046135)	18.75	2.39E-06	4.62E-03
Ribonucleoside catabolic process (GO:0042454)	17.19	2.46E-05	2.24E-02
Nucleoside catabolic process (GO:0009164)	13.31	1.32E-05	1.58E-02
Glycosyl compound catabolic process (GO:1901658)	12.34	3.85E-06	5.97E-03
Pyrimidine-containing compound catabolic process (GO:0072529)	11.79	2.43E-05	2.36E-02
pyrimidine nucleoside metabolic process (GO:0006213)	9.63	1.66E-05	1.84E-02
DNA modification (GO:0006304)	7.2	9.07E-06	1.28E-02
lipid transport (GO:0006869)	3.74	2.18E-05	2.26E-02
lipid localization (GO:0010876)	3.46	5.05E-05	4.12E-02

Table 2: Gene Ontology. Gene Ontology results from significantly deleted regions. No enriched processes were found in amplified regions.

17p13.2 among these deleted regions reinforces the association of this lesion to the development of malignancy (Figure 15).

4. DISCUSSION

In this project I successfully performed genomic profiling of heterogeneous IPMNs by comprehensive genomic analysis of PJD and showed clinical usefulness of PJD sequencing. Pancreatic juice is a useful source for genetic profiling of IPMN and pancreatic tumors; it is less invasive than biopsy and deals better with the heterogeneity of the tumor.

With this method, I was able to detect critical mutations such as *KRAS* and *GNAS*, known markers in IPMN that prove to a certain extent the reliability of this sample source. Furthermore, correlation between mutation burden and IPMN histological grade was found, which has been also previously reported in late stages of gynecologic cancers [56-58]. This correlation of histological grade with mutation burden can be associated not only with the development of malignancy in IPMN, but also with the genomic heterogeneity inherent to advanced tumor samples and the distinct sampling technique used; different cell populations might possess singular mutations that could go undetected by traditional sampling, but might be able to be observed by the use of liquid samples, which has the potential to gather information from all the populations.

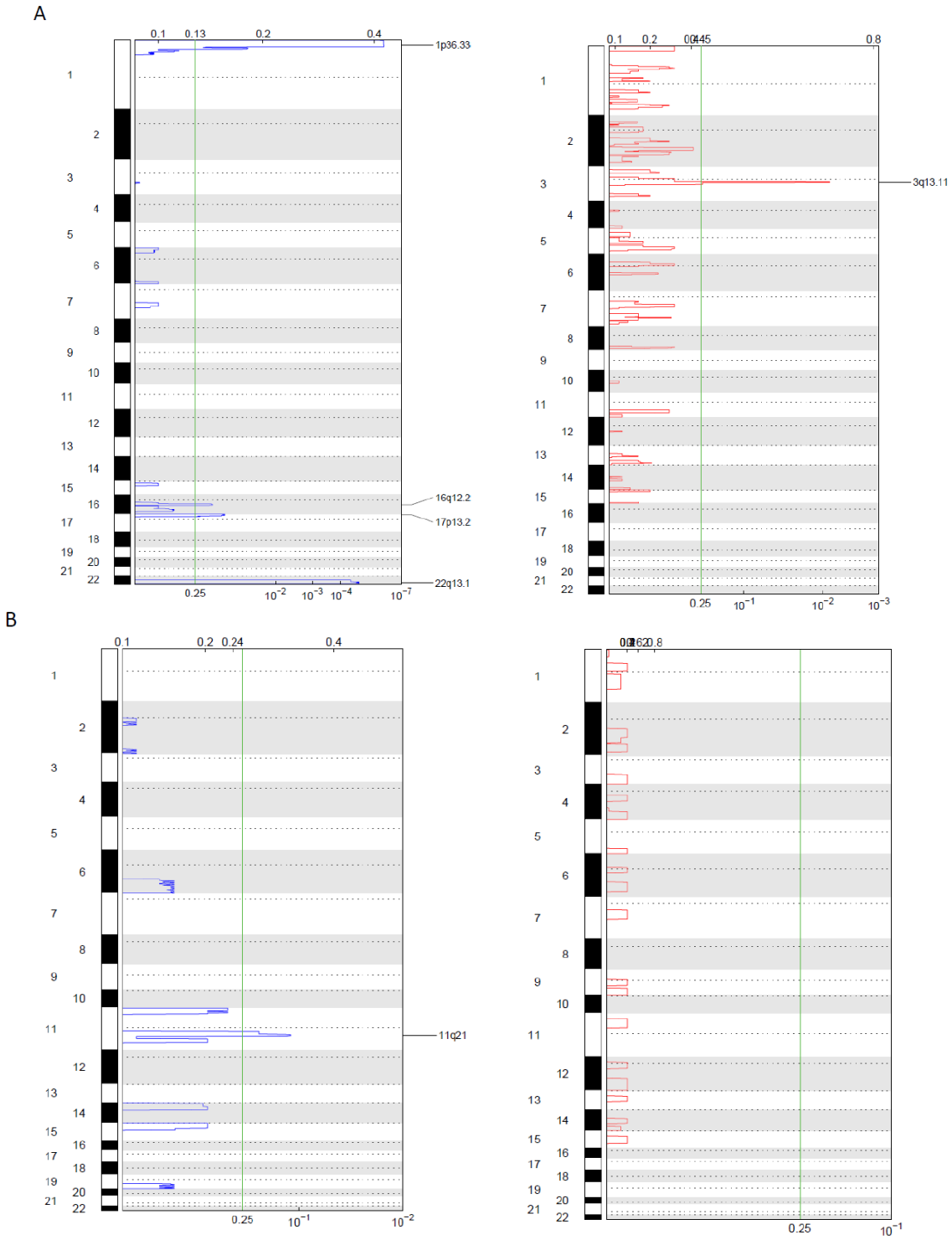


Figure 15: Significantly deleted (blue) and amplified (red) regions by grade. By dividing the dataset in malignant (A), and benign (B), and applying GISTIC2.0 separately, I was able to observe that the deletion of 17p13.2 appears in the malignant set. This reinforces the association of TP53 with IPMN malignancy.

I was also able to find multiple regions that were significantly amplified and deleted in the dataset. The discovery of the deletion of *TP53* in the malignant state of IPMN also reinforced the reliability of my method. More interestingly, the amplification of two of the regions (*7q21.12* and *8q24.22*) in malignant samples lead us to discover the association of *MYC*, a known marker of malignancy, with IPMN development. The amplification, as well as the deregulation of *MYC* has been reported in numerous cancer studies, and its key role in metabolism, the biogenesis of ribosomes, cell cycle, and as a result, in cell growth and proliferation, makes it the perfect target and marker for malignancy in IPMN detection by PJD [59-62]. The combination of this finding with the discovery of the correlation of grade with mutation burden, have strong potential to be used as a predictor of the progression of IPMNs to malignancy. I conclude that exome sequencing or whole genome sequencing analysis of PJD can be useful to evaluate the malignant risk of IPMNs. However, due to the poor quality of the paraffin block DNAs, as well as the genetic and histological heterogeneity of IPMN, the validation of CNA in the IPMN tissues was inconclusive in our study, and it will be required to validate these association of CNAs in further fresh sample cohort or by other methods.

Previous studies have also used pancreatic juice as a source for genomic analysis in IPMN, being their goal the characterization of the main mutations found in this neoplasm, such as KRAS, GNAS and TP53. In order to do so, they performed different genomic analysis such as PCR-SSCP analysis with direct sequencing [63], nonradioisotopic single

strand DNA conformation polymorphism [64], digital high-resolution melt-curve analysis [65] and Digital next-generation sequencing [33]. All these methods, despite their sensitivity, were aimed exclusively towards certain mutations, being therefore unable to evaluate the presence of less common alterations or to find novel ones. In this project, by using Whole Exome Sequencing and the novel combination of filters in order to produce a more robust outcome, we were not only able to find the same mutations in similar proportions, but also led us to find novel markers.

In this project I was able to demonstrate the proof-of-concept of PJD exome testing. However, its clinical deployment has to overcome several issues. Firstly, despite the fact that pancreatic juice would be expected to contain more cfDNA than blood samples [66], due to the risk of producing pancreatitis, the clinical procedure to obtain pancreatic juice has remained controversial and has not been recommended in international consensus guidelines of 2012 for the management of IPMN [67]. Further study will be required to assess the risk and benefit of the PJD exome testing. Secondly, due to the high levels of noise observed in their exome data, I had to exclude some of the PJD samples. Three samples were affected by artifacts in SNVs that resembles OxoG, whereas four samples had extreme fluctuations in their copy number segments. This fluctuations might not be, in fact, real noise, but a consequence of the potential of PJ to overcome the heterogeneity inherent to IPMN. This can be seen as a double-edged sword; thanks to it, it is possible to gather more information, but at the same time might produce a mosaic-like outcome difficult to interpret or even discern from true noise. In this study,

I simply excluded noisy samples from analysis. However, the yield should be improved through protocol optimization. Potential strategies are addition of antioxidants to prevent OxoG, and whole genome sequencing to reduce copy number noise. The development of a method for discerning the multiple subpopulations inside the IPMN could also improve the evaluation of noisy-like samples and give a much clearer and wider vision of their heterogeneity.

Another possible consequence of this quality of PJ of overcoming heterogeneity, could be the reduction of the signal of deletions, amplifications, as well as SNVs. The presence of different tumoral subpopulations with different alterations, combined with the existence of non-tumoral cfDNA in PJ can modify the final ratio of altered/normal reads, potentially making the signal of some of these alterations unable to reach the threshold defined for their detection, leading to the reduction or even the disappearance of them in the final output (Figure 16). Also it is important to take into account that ploidy, the number of subpopulations, and tumor cfDNA content in the sample are strongly intertwined. This relationship is easier to see with an example; a sample with 100% tumor-derived DNA content with no amplifications or deletions in one region would produce a similar relative copy number output to another sample with a duplication in the same region, but with a content of tumoral DNA content of 50% and would be therefore indistinguishable with traditional pipelines. This issue, as we pictured in Figure 7, can increase in complexity with the number of tumoral cell populations coexisting inside the lesion as well as with the source of our samples; solid

samples are limited to a certain region of the lesion/organ, and therefore would give less information about the global characteristics of the neoplasm. Instead, liquid biopsies, due to its nature, would produce much complex yet richer outcomes. Learning how to evaluate this kind of results, being able to discern populations and their proportions is the future direction of my research.

These results also show a great potential for early histological grade prediction. By performing Multiple Factor Analysis [68] (which allows the use of numerical parameters as well as factors) using previous clinical information as well as the SNV and CNA results, it would be possible to get a clustering able to predict the histological grade. In the current circumstances, the sample set used would be too small in order to apply this method properly. A bigger dataset, as well as the inclusion of new data from the patients such as clinical images of the lesions seem key to develop this analysis.

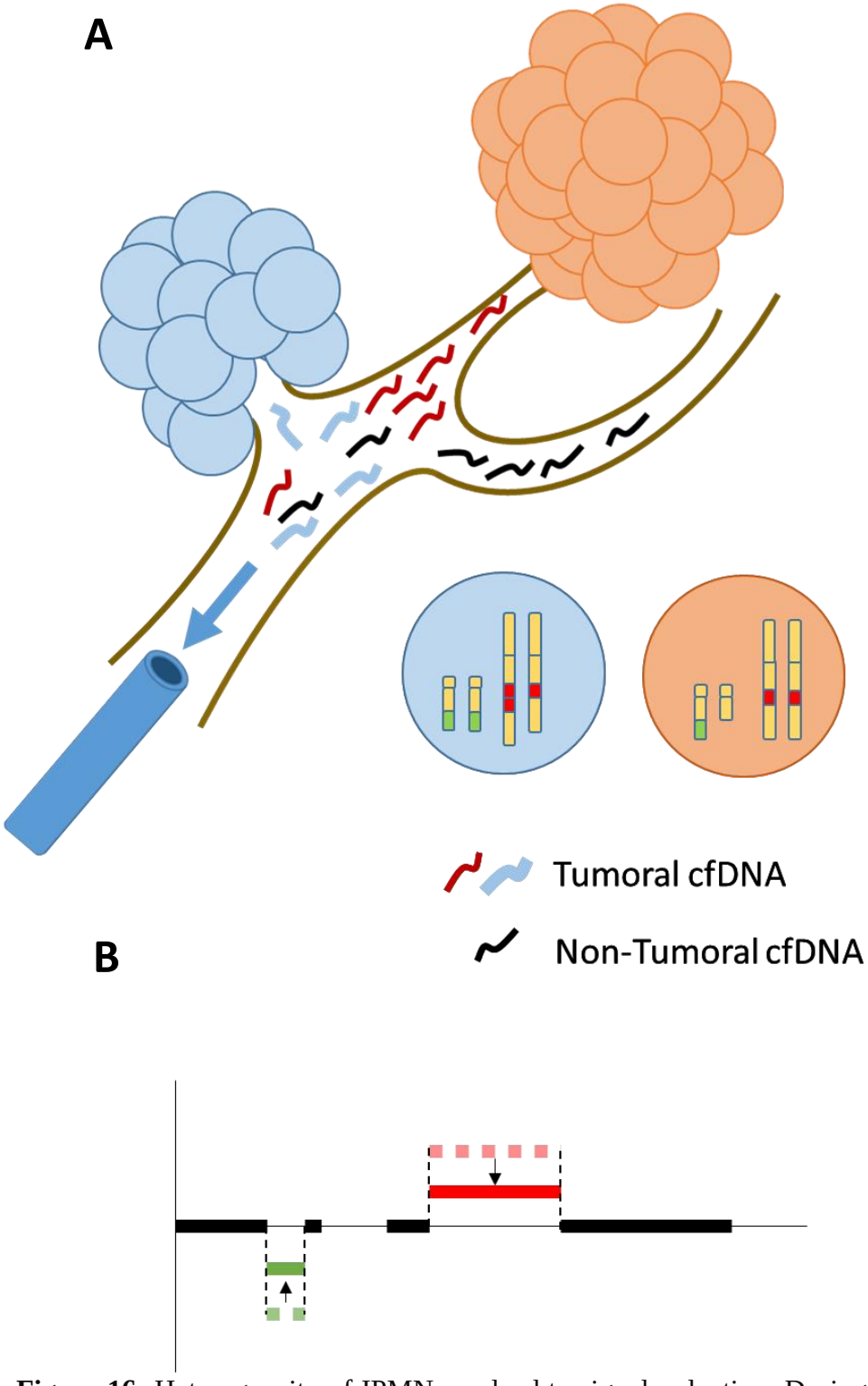


Figure 16: Heterogeneity of IPMN can lead to signal reduction. During the process of extracting PJ by ERCP or ENPD, cfDNA from different tumoral subpopulations, as well as non-tumoral cfDNA is captured (A). This can lead to drops on the final signals of the alterations, which could derive on some of them going undetected.

REFERENCES

1. *The Sol Goldman Pancreatic Cancer Research Center: Types of Pancreas Tumors.*
2. Dolenšek, J., et al., *Pancreas Physiology*, in *Challenges in Pancreatic Pathology*, A. Seicean, Editor. 2017: intechopen.
3. Standring, S., *Gray's anatomy : the anatomical basis of clinical practice*. Forty-first edition. ed. 2016, New York: Elsevier Limited. xviii, 1562 pages.
4. Skarin, A.T. and Dana-Farber Cancer Institute., *Dana-Farber Cancer Institute atlas of diagnostic oncology*. 1991, Philadelphia ; New York: Lippincott ; Gower Medical Pub.
5. Ohashi K, M.Y., Muruayama M, et al. , *Four cases of "mucin-producing" cancer of the pancreas on specific findings of the papilla of vater.* Prog Dig Endosc., 1982(20): p. 348–351.
6. Lee, J.H., et al., *KRAS, GNAS, and RNF43 mutations in intraductal papillary mucinous neoplasm of the pancreas: a meta-analysis.* Springerplus, 2016. **5**(1): p. 1172.
7. Hruban, R.H., et al., *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms.* Am J Surg Pathol, 2004. **28**(8): p. 977-87.
8. Santini, D., et al., *Intraductal papillary-mucinous neoplasm of the pancreas. A clinicopathologic entity.* Arch Pathol Lab Med, 1995. **119**(3): p. 209-13.
9. Sohn, T.A., et al., *Intraductal papillary mucinous neoplasms of the pancreas: an updated experience.* Ann Surg, 2004. **239**(6): p. 788-97; discussion 797-9.
10. Sohn, T.A., et al., *Intraductal papillary mucinous neoplasms of the pancreas: an increasingly recognized clinicopathologic entity.* Ann Surg, 2001. **234**(3): p. 313-21; discussion 321-2.
11. Patra, K.C., N. Bardeesy, and Y. Mizukami, *Diversity of Precursor Lesions For Pancreatic Cancer: The Genetics and Biology of Intraductal Papillary Mucinous Neoplasm.* Clin Transl Gastroenterol, 2017. **8**(4): p. e86.
12. Mino-Kenudson, M., et al., *Prognosis of invasive intraductal papillary mucinous neoplasm depends on histological and precursor epithelial subtypes.* Gut, 2011. **60**(12): p. 1712-20.
13. Maire, F., et al., *Prognosis of malignant intraductal papillary mucinous tumours of the pancreas after surgical resection. Comparison with pancreatic ductal adenocarcinoma.* Gut, 2002. **51**(5): p. 717-22.
14. Yamaguchi, K., et al., *Intraductal papillary-mucinous tumor of the pancreas concomitant with ductal carcinoma of the pancreas.* Pancreatology, 2002. **2**(5): p. 484-90.

15. Salvia, R., et al., *Main-duct intraductal papillary mucinous neoplasms of the pancreas: clinical predictors of malignancy and long-term survival following resection.* Ann Surg, 2004. **239**(5): p. 678-85; discussion 685-7.

16. Moris, D., et al., *Updates and Critical Evaluation on Novel Biomarkers for the Malignant Progression of Intraductal Papillary Mucinous Neoplasms of the Pancreas.* Anticancer Res, 2017. **37**(5): p. 2185-2194.

17. Mandel P, M.P., *Les acides nucléiques du plasma sanguin chez l'homme.* C R Seances Soc Biol Fil, 1948. **142**(241–3).

18. Volik, S., et al., *Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies.* Mol Cancer Res, 2016. **14**(10): p. 898-908.

19. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood.* Nat Rev Clin Oncol, 2013. **10**(8): p. 472-84.

20. Diehl, F., et al., *Detection and quantification of mutations in the plasma of patients with colorectal tumors.* Proc Natl Acad Sci U S A, 2005. **102**(45): p. 16368-73.

21. Ono, A., et al., *Circulating Tumor DNA Analysis for Liver Cancers and Its Usefulness as a Liquid Biopsy.* Cell Mol Gastroenterol Hepatol, 2015. **1**(5): p. 516-534.

22. Togneri, F.S., et al., *Genomic complexity of urothelial bladder cancer revealed in urinary cfDNA.* Eur J Hum Genet, 2016. **24**(8): p. 1167-74.

23. Wang, Y., et al., *Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord.* Proc Natl Acad Sci U S A, 2015. **112**(31): p. 9704-9.

24. Kinde, I., et al., *Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers.* Sci Transl Med, 2013. **5**(167): p. 167ra4.

25. Kawada, N., et al., *Pancreatic juice cytology as sensitive test for detecting pancreatic malignancy in intraductal papillary mucinous neoplasm of the pancreas without mural nodule.* Pancreatology, 2016. **16**(5): p. 853-8.

26. Der, C.J., T.G. Krontiris, and G.M. Cooper, *Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses.* Proc Natl Acad Sci U S A, 1982. **79**(11): p. 3637-40.

27. Kranenburg, O., *The KRAS oncogene: past, present, and future.* Biochim Biophys Acta, 2005. **1756**(2): p. 81-2.

28. Chang, X.Y., et al., *Intraductal papillary mucinous neoplasms of the pancreas: Clinical association with KRAS.* Mol Med Rep, 2018.

29. O'Hayre, M., et al., *The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer.* Nat Rev Cancer, 2013. **13**(6): p. 412-24.

30. Molin, M.D., et al., *Clinicopathological correlates of activating GNAS mutations in intraductal papillary mucinous neoplasm (IPMN) of the pancreas.* Ann Surg Oncol, 2013. **20**(12): p. 3802-8.

31. Amato, E., et al., *Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas.* J Pathol, 2014. **233**(3): p. 217-27.

32. Fritz, S., et al., *Global genomic analysis of intraductal papillary mucinous neoplasms of the pancreas reveals significant molecular differences compared to ductal adenocarcinoma.* Ann Surg, 2009. **249**(3): p. 440-7.

33. Yu, J., et al., *Digital next-generation sequencing identifies low-abundance mutations in pancreatic juice samples collected from the duodenum of patients with pancreatic cancer and intraductal papillary mucinous neoplasms.* Gut, 2017. **66**(9): p. 1677-1687.

34. Eshleman, J.R., et al., *KRAS and GNAS Mutations in Pancreatic Juice Collected From the Duodenum of Patients at High Risk for Neoplasia Undergoing Endoscopic Ultrasound.* Clinical Gastroenterology and Hepatology, 2015. **13**(5): p. 963-969.

35. Quan, Z.F., et al., *Use of endoscopic naso-pancreatic drainage in the treatment of severe acute pancreatitis.* World J Gastroenterol, 2003. **9**(4): p. 868-70.

36. *Picard. http://broadinstitute.github.io/picard/.*

37. Chiba, K., et al., *Genomon2 http://genomon.readthedocs.io/ja/latest/.*

38. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.

39. Koeffler, A.M.a.P.H., *Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies.* BioRxiv, 2016.

40. Costello, M., et al., *Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation.* Nucleic Acids Res, 2013. **41**(6): p. e67.

41. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Res, 2012. **22**(3): p. 568-76.

42. Huber, W., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nat Methods, 2015. **12**(2): p. 115-21.

43. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* Genome Biol, 2011. **12**(4): p. R41.

44. Momtaz, P., et al., *Quantification of tumor-derived cell free DNA(cfDNA) by digital PCR (DigPCR) in cerebrospinal fluid of patients with BRAFV600 mutated malignancies.* Oncotarget, 2016. **7**(51): p. 85430-85436.

45. Tian, R., M.K. Basu, and E. Capriotti, *Computational methods and resources for the interpretation of genomic variants in cancer.* BMC Genomics, 2015. **16 Suppl 8**: p. S7.

46. Takano, S., et al., *Deep sequencing of cancer-related genes revealed GNAS mutations to be associated with intraductal papillary mucinous neoplasms and its main pancreatic duct dilation.* PLoS One, 2014. **9**(6): p. e98718.

47. Matthaei, H., et al., *Clinicopathological characteristics and molecular analyses of multifocal intraductal papillary mucinous neoplasms of the pancreas.* Ann Surg, 2012. **255**(2): p. 326-33.

48. Fujita, M., et al., *Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling.* Oncotarget, 2018. **9**(1): p. 969-981.

49. Hirono, S., et al., *Molecular markers associated with lymph node metastasis in pancreatic ductal adenocarcinoma by genome-wide expression profiling.* Cancer Sci, 2010. **101**(1): p. 259-66.

50. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.

51. Carbon, S., et al., *AmiGO: online access to ontology and annotation data.* Bioinformatics, 2009. **25**(2): p. 288-9.

52. Leiserson, M.D., et al., *Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes.* Nat Genet, 2015. **47**(2): p. 106-14.

53. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.

54. Bergmann, F., et al., *Acinar cell carcinomas of the pancreas: a molecular analysis in a series of 57 cases.* Virchows Arch, 2014. **465**(6): p. 661-72.

55. La Rosa, S., et al., *c-MYC amplification and c-myc protein expression in pancreatic acinar cell carcinomas. New insights into the molecular signature of these rare cancers.* Virchows Arch, 2018.

56. Zhuang, W., et al., *The Tumor Mutational Burden of Chinese Advanced Cancer Patients Estimated by a 381-cancer-gene Panel.* J Cancer, 2018. **9**(13): p. 2302-2307.

57. Wang, M., et al., *Molecular profiles and tumor mutational burden analysis in Chinese patients with gynecologic cancers.* Sci Rep, 2018. **8**(1): p. 8990.

58. Chalmers, Z.R., et al., *Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden.* Genome Med, 2017. **9**(1): p. 34.

59. Hsieh, A.L., et al., *MYC and metabolism on the path to cancer.* Semin Cell Dev Biol, 2015. **43**: p. 11-21.

60. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.

61. Kalkat, M., et al., *MYC Deregulation in Primary Human Cancers.* Genes (Basel), 2017. **8**(6).

62. Alitalo, K., et al., *Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma.* Proc Natl Acad Sci U S A, 1983. **80**(6): p. 1707-11.

63. Kaino, M., et al., *Detection of K-ras and p53 gene mutations in pancreatic juice for the diagnosis of intraductal papillary mucinous tumors.* Pancreas, 1999. **18**(3): p. 294-9.

64. Kondo, H., et al., *Detection of K-ras gene mutations at codon 12 in the pancreatic juice of patients with intraductal papillary mucinous tumors of the pancreas.* Cancer, 1997. **79**(5): p. 900-5.

65. Kanda, M., et al., *Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts.* Gut, 2013. **62**(7): p. 1024-33.

66. Takai, E. and S. Yachida, *Circulating tumor DNA as a liquid biopsy target for detection of pancreatic cancer.* World Journal of Gastroenteroly, 2016.

67. Tanaka, M., et al., *International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas.* Pancreatology, 2012. **12**(3): p. 183-97.

68. B., E. and P. J., *Multiple factor analysis (AFMULT package)* Computational Statistics & Data Analysis 1994(18): p. 121-140

SUPLEMENTARY TABLES AND FIGURES

ID	Age	Sex	Pathology	Histological grade	Subtype	Macroscopic Type	Tumor_size _mm	Tumor_location	Diamter_of _MPD	Serum_CEA	Serum_CA19-9	PJ_CEA
W6	86	M	invasive_IPMC	INC	I	MD	0	head	11.1	3.2	6.1	1254.4
W7	69	M	invasive_IPMC	INC	I	MIX	34.1	head	5.2	2.1	10.7	1414.9
W8	80	M	invasive_IPMC (non-resected)	INC	NA	MD	0	head-tail	15.8	4.9	822.4	6660.9
W11	84	F	High_grade_dysplasia_+PDAC	INC	NA	BD	20	head	4.1	1.5	11.7	29.4
W12	60	M	Adenoma_with_moderate_atypia	LGD	G	MIX	27.4	head	5.2	1.5	7.3	37.9
W13	72	M	High_grade_dysplasia=_cancer_in_situ	HGD	G	BD	32	head	6.8	1.7	2	80.5
W14	69	F	Adenoma_with_moderate_atypia+PDAC	INC	NA	BD	32	body	4	2.5	104.5	178.2
W16	80	M	High_grade_dysplasia=_cancer_in_situ	HGD	I	MD	0	body-tail	14.6	3.5	5.7	842.2
W17	75	M	Adenoma_with_moderate_atypia	LGD	G	BD	31.8	head	1.7	1.9	5.1	99.3
W18	74	M	High_grade_dysplasia=_cancer_in_situ	HGD	G	MX	67.7	head	5.5	3	18.2	52.1
W20	76	M	adenoma_Low-intermediate_grade_dysplasia	LGD	G	MD	31	tail	13.4	1.7	8.3	92
W21	58	M	High_grade_dysplasia=_cancer_in_situ	HGD	G	BD	29.2	head-tail	3.4	2.8	3.2	56.2
W22	68	M	High_grade_dysplasia=_cancer_in_situ	HGD	G	BD	42.9	head	3.9	2.1	9.6	31.4
W23	83	F	High_grade_dysplasia=_cancer_in_situ	HGD	I	MD	35	head	7	1.3	3	73.9
W24	63	M	High_grade_dysplasia=_cancer_in_situ	HGD	PB	MD	26	head	8.2	2.5	2	227.7
W25	72	M	invasive_IPMC	INC	G	MD	29.7	head	9.1	4	13.5	5013.3
W26	58	F	High_grade_dysplasia=_cancer_in_situ	HGD	G	BD	27	head	2	3.3	2	46.4
W27	60	M	High_grade_dysplasia=_cancer_in_situ	HGD	I	MX	26.5	body-tail	9.6	3.6	9.5	124
W28	67	M	invasive_adenosquamous_carcinoma_(PDAC)	INC	NA	NA	NA	NA	NA	NA	NA	NA
W29	65	F	Adenoma_with_moderate_atypia	LGD	G	BD	33.8	body-tail	1	2.3	4.9	47.8
W30	73	F	High_grade_dysplasia=_cancer_in_situ	HGD	G	MX	32	head	15	4.5	17.6	107.9
Y1	78	F	Adenoma_with_severe_atypipa	HGD	I	BD	20	head	4	2.5	62.63	11.4
Y2	78	F	invasive_IPMC	INC	I	MIX	139	body-tail	9	4.6	390	67.3
Y4	67	F	Cancer_in_adenoma	HGD	G	MIX	36	head	6	2	1200	10.4
Y5	75	F	High_grade_dysplasia=_cancer_in_situ	HGD	G	MIX	19	head	6	3.8	31.22	1.8
Y6	76	F	invasive_IPMC	INC	PB	BD	16	body-tail	4	1.5	10.76	252.6
Y7	61	M	invasive_IPMC	INC	PB	MIX	34	head	12	1.7	162	1795.3
Y8	62	M	High_grade_dysplasia=_cancer_in_situ	HGD	G	MIX	24	body-tail	10	3.4	0.6	73.4
Y9	74	M	Adenoma_with_moderate_atypia	LGD	G	MIX	46	head	6	8.6	67.24	21.4
Y10	79	F	invasive_IPMC	INC	I	MD	NA	body-tail	8	4.7	188.3	701.4
Y11	74	F	Adenoma_with_moderate_atypia	LGD	G	BD	37	head	4.8	1.3	0.6	NA
Y12	73	M	Cancer_in_adenoma	HGD	I	MD	NA	body-tail	12	2.5	16.47	40.9
Y13	67	M	Cancer_in_adenoma	HGD	G	BD	38	body-tail	1.2	3.9	6.79	35.5
Y14	76	M	Adenoma_with_moderate_to_severe_atypia	LGD	G	MIX	6	head	7	1.1	16.23	80.3
Y15	69	M	invasive_IPMC	INC	I	MD	NA	body-tail	33	2.2	13.16	831.1
Y16	62	M	Adenoma_with_moderate_atypia	LGD	G	BD	30	head	4	3.7	10.72	13.7
Y17	81	M	Cancer_in_adenoma	HGD	I	MIX	10	body	6	2.1	8.68	150.1
Y18	79	F	Cancer_in_adenoma	HGD	PB	MIX	45	body-tail	17	1.2	39.3	NA
Y19	75	F	Cancer_in_adenoma	HGD	G	MD	NA	body-tail	6	2.1	8.44	6.3
Y20	68	F	Cancer_in_adenoma	HGD	G	MIX	73	bodyl	12	3.5	22.99	55.9

Table S1: Metadata of the sample set offered by Wakayama Medical University and Yamanashi University school of Medicine. Pathological description was reviewed by Kyoto Prefectural University of Medicine. Sample W28 was excluded since it was diagnosed as pancreatic ductal adenocarcinoma.

Sample	RV
Y14	1.40E-06
W6	3.82E-05
Y1	5.80E-05
Y4	0.000135
W13	0.000268
W29	0.000317
W16	0.000327
Y19	0.000361
Y16	0.000392
W21	0.000465
Y9	0.000489
W7	0.000581
Y5	0.000712
Y8	0.000847
W25	0.000863
Y7	0.000937
Y15	0.000994
W26	0.001083
W27	0.001227
Y12	0.001251
Y11	0.001387

Sample	RV
Y2	0.00162
W17	0.001771
Y10	0.002199
Y6	0.002231
Y18	0.002684
Y13	0.002878
W30	0.003102
Y17	0.003407
W18	0.004131
W24	0.004567
W12	0.007261
W11	0.007999
Y21	0.009614
W14	0.011484
Y3	0.013386
W8	0.014222
W23	0.01632
W20	0.016689
Y20	0.025143
W22	0.028662

Table S2: Residual Variance. Residual Variance of each sample was calculated by using the CNA output from Varscan2 and the R package DNAcopy. Samples with more than 0.015 variance were removed from the analysis in order to keep it free of potential noise.

Samples	read_length_r1	read_length_r2	mapped_bases	mapped_bases_r1
BLOOD	126	126	9388822246	4705020507
PJ cfDNA	126	126	19257643802	9641814048
mapped_bases_r2	divergent_bases	divergent_bases_r1	divergent_bases_r2	total_reads
4683801739	36555261	16797178	19772361	82126264
9615829754	64356721	28147403	36331198	163610212
total_reads_r1	total_reads_r2	mapped_reads	mapped_reads_r1	mapped_reads_r2
41063132	41063132	82116816	41062366	41054450
81805106	81805106	163545808	81801030	81744778
mapped_reads_properly_paired	gc_bases_r1	gc_bases_r2	mean_insert_size	insert_size_sd
39753284	2512051543	2496312234	164.855	63.486
80043660	4966344219	4957706427	189.674	79.261
median_insert_size	duplicate_reads	total_depth	bait_size	average_depth
155	4596093	3798420402	33805487	112.36106
177	32130318	5677845927	33805487	167.95634
depth_stdev	2x_ratio	10x_ratio	20x_ratio	30x_ratio
79.37516	0.9876113	0.9664724	0.9330277	0.8875777
92.40639	0.991035	0.9797802	0.9674708	0.9530351
40x_ratio	50x_ratio	100x_ratio	2x	10x
0.8405373	0.7824688	0.4830978	33386682	32672070
0.934025	0.91044066	0.731856636	33502420	33121947
20x	30x	40x	50x	100x
31541456	30004995	28414774	26451740	16331355
32705822	32217814	31575170	30777890	24740770

Table S3: Quality Control: Median of the Quality Control results obtained for Blood and PJD by Genomon2.

Sample	Tumor-derived DNA content ratio	N_mutations
W06	0.268893097	74
W07	0.333333333	55
W08	0.576976422	210
W11	0.165043034	52
W12	0.136057692	16
W13	0.241212121	36
W14	0.430952381	96
W16	0.467391304	99
W17	0.301075269	29
W25	0.422291022	196
W26	0.161290323	23
W29	0.252272727	58
W30	0.520833333	83
Y10	0.223538597	76
Y12	0.182432432	27
Y13	0.469387755	17
Y14	0.147897786	12
Y15	0.378378378	55
Y16	0.373525992	74
Y17	0.424358974	36
Y2	0.207191073	10
Y4	0.137254902	23
Y5	0.23880597	11
Y6	0.588235294	63

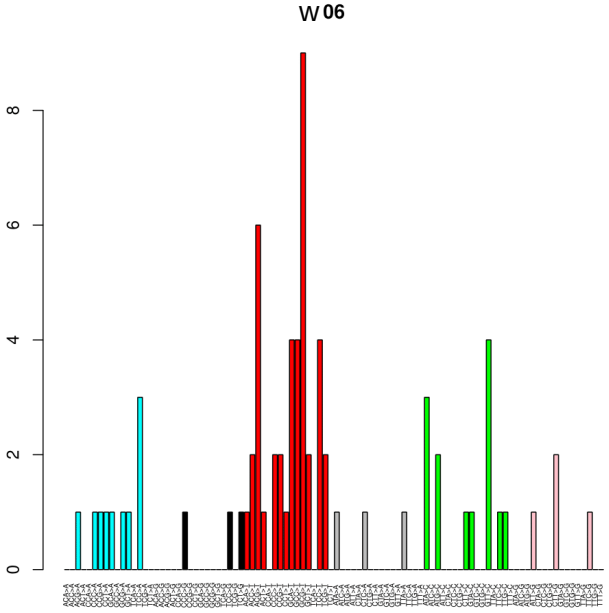
Table S4: Tumor-derived DNA content: List of tumor-derived DNA content extrapolated from the VAF values of the PJD samples. Only samples with more than 10 mutations were used for this estimation.

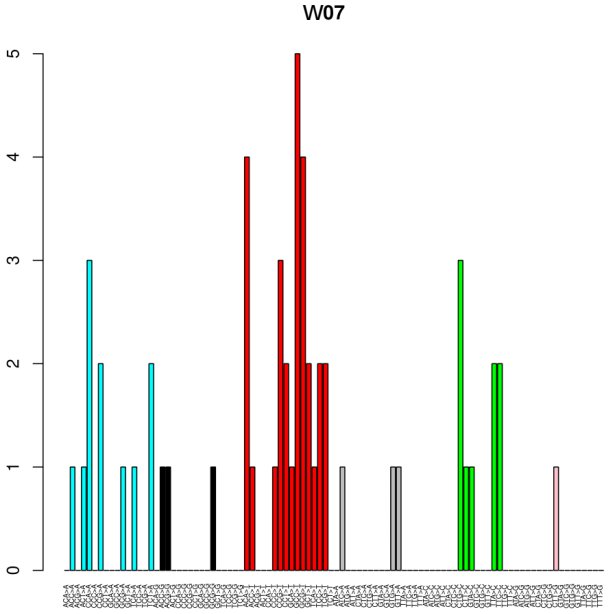
cytoband	3q13.11	2q31.2	5q32	7q21.12	8q24.22	6p12.1	11q22.1	1p31.2	12q21.31	3p22.3	2p12
q value	0.0004708	0.046264	0.046264	0.046264	0.051408	0.054055	0.054055	0.058412	0.10134	0.21666	0.23056
residual q value	0.0004708	0.046264	0.046264	0.046264	0.051408	0.054055	0.054055	0.058412	0.10134	0.21666	0.23056
wide peak boundaries	chr3:100537291-105290690	chr2:179468788-181848896	chr5:147012224-150322314	chr7:86548477-86574202	chr8:126010793-133850047	chr6:55966190-65098628	chr11:95574690-104767921	chr1:63050434-72241985	chr12:80226101-81015781	chr3:17444624-37363383	chr2:79253095-96604546
genes in wide peak	ALCAM, RPL24, ABI3BP, ZBTB11, IMPG2, RG9MTD1, PCNP, SENP7, NFKBIZ, CEP97, FAM55C, ZPLD1, FAM172BP, LOC152225, LOC285359, MIR548A3, LOC100009676	hsa-mir-1258, TTN, UBE2E3, CWC22, SESTD1, ZNF385B, CCDC141, MIR548N, MIR1258, LOC100506866	hsa-mir-378, hsa-mir-145, hsa-mir-584, ADRB2, CAMK2A, CD74, CDX1, CSF1R, CSNK1A1, SLC26A2, NDST1, HTR4, PDE6A, PDGFRB, RPS14, SLC6A7, SPINK1, TCOF1, JAKMIP2, SPINK5, SYNPO, ABLIM3, HMGXB3, IL17B, DCTN4, RBM22, PCYOX1L, SH3TC2, FBXO38, TIGD6, SPINK7, C5orf62, ZNF300, MYOZ3, SCGB3A2, PPARGC1B, AFAP1L1, GRPEL2, ZNF300P1, SPINK13, LOC153469, ARSI, IRGM, C5orf46, ARHGEF37, SPINK6, MIR143, MIR145, SPINK14, SPINK9, MIR143HG	KIAA1324L	hsa-mir-1208, hsa-mir-1207, hsa-mir-1205, hsa-mir-1204, ADCY8, KCNQ3, MYC, POU5F1B, PVT1, SQLE, KIAA0196, HHLA1, TRIB1, EFR3A, LRRC6, ASAP1-IT1, ASAP1, PHF20L1, FAM49B, GSDMC, HPYR1, TMEM71, FAM84B, NSMCE2, LOC727677, LOC728724, OC90, LOC100130231, MIR1205, MIR1206, MIR1207, MIR1204, MIR1208, LOC100507117, PCAT1	hsa-mir-548u, DST, PRIM2, PTP4A1, BAG2, PHF3, ZNF451, LGSN, RAB23, KIAA1586, COL21A1, KHDRBS2, BEND6, EYS, GUSBP4	hsa-mir-1260b, BIRC2, BIRC3, MMP1, MMP3, MMP7, MMP8, MMP10, MMP12, MMP13, PGR, TRPC6, JRKL, MTMR2, MMP20, YAP1, CNTN5, KIAA1377, MMP27, DYNC2H1, CCDC82, PDGFD, TMEM133, DCUN1D5, MAML2, C11orf70, TMEM123, ARHGAP42, ANGPTL5, DDI1, LOC100288077, MIR1260B, MIR3920, CASP12, MIR4693	hsa-mir-186, hsa-mir-1262, hsa-mir-3117, hsa-mir-101-1, AK4, CTH, GADD45A, IL12RB2, JAK1, LEPR, ROR1, PDE4B, PGM1, PIN1P1, PTGER3, RPE65, DIRAS3, SRSF11, ZRANB2, DNAJC6, INSL5, HHLA3, SLC35D1, ITGB3BP, SERBP1, FOXD3, ANGPTL3, ALG6, LEPROT, RAVER2, LRRC40, DEPDC1, GNG12, LRRC7, CACHD1, MIER1, DLEU2L, WDR78, WLS, ANKRD13C, SGIP1, EFCAB7, ATG4C, DOCK7, UBE2U, IL23R, LINC00466, TCTEX1D1, NEGR1, C1orf141, MIR101-1, MIR186, ZRANB2-AS1, LOC100289178, MIR1262, MIR3117, MIR3671, MIR4794, ZRANB2-AS2	PPP1R12A, OTOGL, PTPRQ	hsa-mir-128-2, hsa-mir-466, hsa-mir-3135, CCR4, GLB1, GOLGA4, MLH1, RAB5A, RARB, RPL15, SATB1, STAC, TGFBR2, THRB, TOP2B, UBE2E1, UBE2E2, UBP1, EOMES, KAT2B, LRRFIP2, SLC4A7, TBC1D5, EPM2AIP1, TRANK1, NR1D2, PDCD6IP, CRTAP, ARPP21, CLASP2, GPD1L, FBXL2, CNOT10, SUSD5, RBMS3, NKIRAS1, DYNC1LI1, CMTM6, OXSM, NGLY1, AZI2, ZNF385D, DCLK3, CMTM7, OSBPL10, LRRC3B, KCNH8, TRIM71, SGOL1, PP2D1, EFHB, LOC152024, ZCWPW2, CMC1, NEK10, CMTM8, STT3B, LOC285326, LOC339862, GADL1, ZNF860, VENTXP7, MIR128-2, TMPPE, LOC644990, LOC645206, MIR4791, MIR548AC, MIR4792, MIR4442	hsa-mir-4264, CAPG, CD8A, CD8B, CTNNA2, LOC1720, DNAH6, FABP1, GGCX, MAL, MAT2A, REG3A, PLGLB2, PLGLB1, REG1A, REG1B, REG1P, SFTPB, ZNF2, RNF103, VAMP8, SUCLG1, ST3GAL5, TMSB10, EIF2AK3, GNLY, TGOLN2, USP39, VAMP5, IMMT, RPIA, POLR1A, KCNIP3, FAHD2A, RNF181, KRCC1, MRPL35, CHMP3, RETSAT, PTCD3, THNSL2, KDM3A, KCMF1, RMND5A, MRPS5, REEP1, TCF7L1, ELMOD3, ZNF514, ATOH8, LOC90499, LOC90784, LINC00152, C2orf89, TMEM150A, TRIM43, REG3G, TEKT4, SMYD1, PROM2, C2orf51, SH2D6, LOC284950, LOC285074, FOXI3, LRRTM1, FUNDC2P2, C2orf68, RGPD1, ACTR3BP2, LOC442028, GGT8P, ANKRD36BP2, TRIM43B, LOC654342, SNORD94, ANKRD20A8P, RGPD2, LOC100286979, MIR4264, RNF103-CHMP3, MIR4779, MIR4435-2, MIR4436A, MIR4780, MIR4435-1, LOC100630918

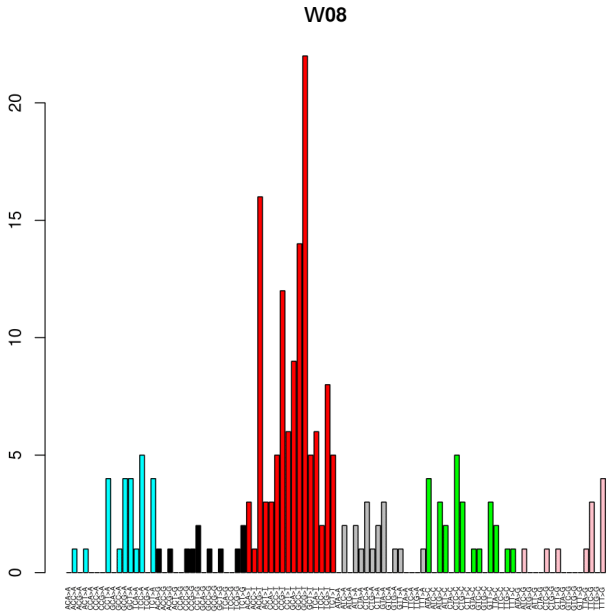
Table S5: GISTIC2.0 results: Amplified regions. List of genes belonging to the significantly amplified regions in the dataset obtained from GISTIC2.0 using the results from the CNA analysis as input.

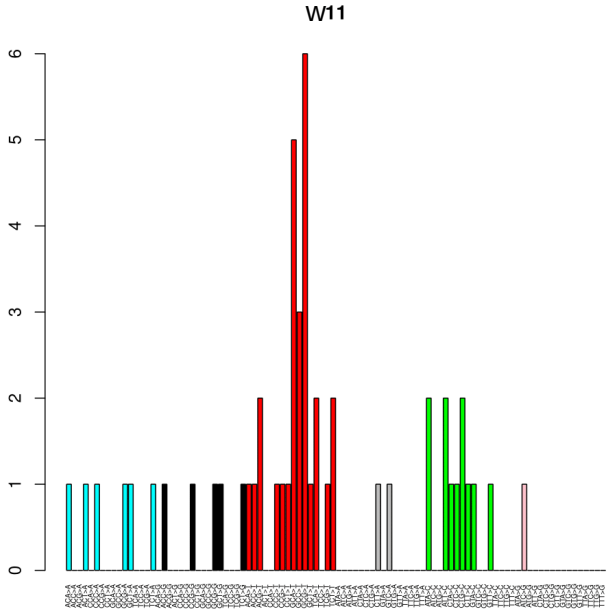
cytoband	22q13.1	1p36.31	17p13.2	16q12.2	6p23
q value	1.83E-08	0.0011245	0.0085209	0.10533	0.20099
residual q value	1.83E-08	0.0011245	0.0085209	0.10533	0.20099
wide peak boundaries	chr22:26106934-51304566	chr1:4715435-6445516	chr17:1-10363621	chr16:52092230-59771890	chr6:1-31239851
genes in wide peak	hsa-mir-3201, hsa-let-7b, hsa-mir-1249, hsa-mir-33a, hsa-mir-1281, hsa-RPL22, KCNAB2, mir-659, hsa-mir-3200, hsa-mir-3199-2, hsa-mir-548j, ACR, ACO2, AD5L, ACOT7, ICMT, CHD5, AP1B1, ARSA, ATF4, BIK, TSPO, MPPED1, CHKB, CPT1B, CRYBA4, CRYBB1, NPHP4, GPR153, CSF2RB, CSNK1E, CYP2D6, CYB5R3, TYMP, EP300, EWSR1, RNF207, HES3, FBLN1, XRCC6, MCHR1, H1F0, HMOX1, IL2RB, KCNA4, LGALS1, LGALS2, MIR4689, MIR4417, LIF, LIMK2, MB, MCM5, MFNG, MGAT3, MNX1, MPST, MYH9, NAGA, NCF4, NDUFA6, DRG1, NEFH, NF2, NHP2L1, OSM, PDGFB, PMM1, POLR2F, PPARA, MAPK11, PVALB, RAC2, RANGAP1, RFPL1, RPL3, MAPK12, SBF1, SLC5A1, SMTN, SLC5A4, SOX10, SREBF2, SSTR3, ST13, TCF20, TCN2, TEF, TIMP3, TST, UPK3A, WNT7B, XBP1, YWHAH, SYN3, PLA2G6, TPST2, GALR3, NIPSNAP1, APOL1, THOC5, EIF3D, MTMR3, CACNA1I, SYNGR1, LARGE, GRAP2, PICK1, GAL3ST1, GTPBP1, APOBEC3B, CELSR1, DEPDC5, PPP6R2, SFI1, ZBED4, JOSD1, RBX1, SCO2, HMGXB4, TOM1, DNAL4, SF3A1, PKDREJ, CACNG2, TAB1, SLC25A17, DDX17, RASL10A, GAS2L1, RPL13-AS1, RPL3, RFPL2, RPL1-AS1, NUP50, TOB2, KDELR3, IFT27, TRIOBP, CDC42EP1, DMC1, RABL2B, CHEK2, PACSIN2, MORC2, TNRC6B, GRAMD4, TTLL12, MLC1, ZC3H7B, KIAA0930, TTC28, GCAT, CBX6, NPTXR, PES1, CBX7, SLC16A8, SEC14L2, MAPK8IP2, RBFOX2, SEZ6L, RASD2, PATZ1, SH3BP1, PLXNB2, PITPNB, PISD, OSBP2, MAFF, BRD1, ARHGAP8, APOL2, AP1B1P1, TFIP11, C22orf31, TBC1D22A, C22orf24, CBY1, SUN2, FBXO7, RHBDD3, TTLL1, SAMM50, ATXN10, FAM19A5, TXN2, TMEM184B, SULT4A1, GGA1, RIBC2, ARFGAP3, SNORD43, INPP5J, SMC1B, CYTH4, CSDC2, RRP7A, MCAT, APOBEC3C, PPPDE2, SGSM3, CARD10, PARVB, NCAPH2, UQCR10, PLA2G3, EIF3L, C22orf28, GTSE1, MTFP1, A4GALT, MOV10L1, SMCR7L, TUG1, FAM118A, TTC38, C22orf26, MIOX, PRR5, TRMU, ZMAT5, 37865, EIF4ENIF1, PANX2, TOMM22, PDXP, ASPHD2, MKL1, APOBEC3G, XPNPEP3, PARVG, CERK, EFCAB6, CENPM, ALG12, NOL12, CRELD2, C22orf46, KCTD17, CCDC134, ADM2, FOXRED2, BAIAP2L2, SCUBE1, TRABD, PNPLA3, APOL6, APOL5, APOL4, APOL3, SELO, L3MBTL2, TBC1D10A, HDAC10, KREMEN1, ZNRF3, ASCC2, LDOC1L, POLDIP3, C22orf23, MYO18B, PHF5A, KIAA1644, SHANK3, KIAA1656, MICALL1, TUBGCP6, HPS4, LOC90834, LMF2, RNF185, ISX, RPS19BP1, C22orf32, RRP7B, SERHL, PHF21B, KLHDC7B, PIK3IP1, FAM83F, ELFN2, C1QTNF6, TNFRSF13C, RNU86, SNORD83A, SNORD83B, EMID1, ANKRD54, APOBEC3D, SELM, HSCB, CCDC117, HORMAD2, DUSP18, MORC2-AS1, C22orf42, ENTHD1, DNAJB7, CHADL, MEI1, FAM109B, NFAM1, PNPLA5, LOC150381, C22orf40, CNSH6.4, CABP7, TMPRSS6, APOBEC3H, WBP2NL, POLR3H, RNF215, APOBEC3A, APOBEC3F, PRR14L, SERHL2, BPIFC, SEC14L3, RNU12, ATP5L2, TTC28-AS1, SEC14L4, LOC284933, RPL23AP82, SLC35E4, LOC339666, C22orf33, BK250D10.8, LOC339685, C22orf34, SDCAP, CHKB-CPT1B, LOC388906, LINC00207, LOC400927, MIRLET7BHG, FLJ46257, IL17RE, SRRD, MIRLET7A3, MIRLET7B, MIR33A, LINC00229, FAM116B, PIM3, MIAT, SHISA8, ODF3B, CCDC157, PRR5-ARHGAP8, SYCE3, LOC646851, GATSL3, MIR659, SEC14L6, LOC730668, SNORD125, LOC100128946, LOC100132273, LOC100144603, LOC100271722, MIR1249, MIR1281, MIR548I, MIR3200, MIR33201, MIR3199-2, MIR3199-1, MIR3909, MIR3619, MIR3928, LOC100506195, LOC100506241, LOC100506472, LOC100506714, MIR4763, MIR4534, MIR4762, MIR4766, MIR4764, MIR4535	KCNAB2, ICMT, CHD5, HES3, MIR4689, MIR4417	hsa-mir-4314, hsa-mir-324, hsa-mir-497, hsa-mir-1253, hsa-mir-138-2, mir-212, hsa-mir-22, hsa-mir-3183, ABR, ACADVL, ALOX12B, ALOX12P2, ALOX15, ALOX15B, ARRB2, ASGR2, ASPA, ATP1B2, ATP2A3, C1QBP, CD68, CHD3, CHRNB1, CHRNE, CLDN7, CRK, CTNS, DLG4, DPH1, DVL2, EFNB3, EIF4A1, EIF5A, ENO3, FGF11, GP1BA, GP5, GP9, GUCY2D, HIC1, ITGAE, MNT, MYH8, MYH10, MYO1C, NUP88, OR1D2, OR3A1, OR3A2, P2RX1, P2RX5, PAFAH1B1, PFAS, PFN1, PITPNA, PLD2, SERPINF2, POLR2A, PSMB6, PYGB, RAP1GAP2, RCVRN, RPA1, RPL26, SHBG, SLC2A4, SOX15, VAMP2, TP53, UBE2G1, TRPV1, YWHAE, ZNF232, OR1A1, OR1D5, OR1E1, OR1E2, OR1G1, OR3A3, SLC25A11, DOC2B, GAS7, SCARF1, TNK1, MYH13, TNFSF12, TNFSF13, TM4SF5, USP6, RABEP1, KCNA8B3, AURKB, GLP2R, MTN1, STX8, RPH3AL, SPAG7, ACAP1, KIAA0753, SGSM2, CLEC10A, PRPF8, KIF1C, GABARAP, NLRP1, ARHGEF15, RAP1GAP2, CAMTA2, KDM6B, ZZEF1, KIAA0664, SMG6, CTDNEP1, PIK3R5, C17orf81, SHPK, AIPL1, RNF167, SENP3, OR1A2, SNORA67, PELP1, RANGRF, TIMM22, TAX1BP3, MINK1, GEMIN4, MED31, DERL2, GLOD4, YBX2, ANKFY1, TNXDC17, MIR22HG, TMEM88, SAT2, SMYD4, CNTROB, RPL29P2, CYB5D1, OVCA2, USP43, KRBA2, SLC43A2, CYB5D2, C17orf49, ZFP3, GGT6, SPNS2, WDR81, DNAH2, RTN4RL1, WDR16, CCDC42, PIK3R6, ODF4, KCTD11, MFSD6L, SLC16A11, FBXO39, DHRS7C, SLC16A13, C17orf74, SPNS3, C17orf61, BCL6B, LOC284023, LINC00324, SLC13A5, TMEM102, TUSC5, TMEM95, SMTNL2, ALOX15P1, FAM101B, SPEM1, GLTPD2, INCA1, SCIMP, C17orf100, SPDYE4, OR3A4P, SLC25A35, C17orf97, MED11, MIR132, MIR195, MIR212, MIR22, TNFSF12-TNFSF13, RNASEK, MIR324, MIR497, SLC35G6, RNF222, SNOR448, SNORD10, OR1D4, SCARNA21, SNORD91A, SNORD91B, BHLHA9, TLCD2, LOC728392, LOC100128288, C17orf107, LOC100130950, DBIL5P, MIR1253, LOC100306951, MIR3183, MIR4314, MIR3676, LOC100506388, LOC100506715, MIR497HG, RNASEK-C17ORF49, C17orf61-PLSCR3, SENP3-EIF4A1, P2RX5-TAX1BP3, MIR4520A, MIR4521, MIR4520B	BBS2, CES1, CETP, CNGB1, KIFC3, MMP2, MMP15, MT1A, MT1B, MT1E, MT1F, MT1G, MT1H, MT1L, MT1X, OR1D2, MT2M, MT2A, MT3, POLR2C, RBL2, CCL17, CCL22, CX3CL1, RREB1, ATXN1, SLC12A3, GPR56, CNOT1, KATNB1, NUDT21, ARL2BP, TOX3, DKFZP434H168, PLLP, CCDC113, C16orf80, MYBBP1A, CES1P1, LPCAT2, SLC38A7, COQ9, DOK4, ZNF319, AKTIP, ZNF323, ZNF322, RPP21, ATAT1, ZKSCAN3, C17orf74, SPNS3, LOC100505619	BPHL, BTN1A1, DDR1, CDSN, DSP, E2F3, EDN1, SERPINB1, FLJ23152, LINC00340, LOC401242, C6orf201, ERVFRD-1, HCG18, OR2B3, OR2J3, OR14J1, OR10C1, HCG11, LOC554223, PPP1R3G, SNORD32B, MIR548A1, HULC, LOC729177, TMEM170B, SCARNA27, MIR877, LOC100129636, LOC100130275, LOC100130357, PSORS1C3, LOC100131289, LINC00240, LOC100270746, MIR3143, MIR3691, LOC100506207, LOC100506409, LOC100507173, LOC100507194, LOC100507362, MUC22, LOC100508120, MUTED-TXNDC5, EEF1E1-MUTED, MIR4640, MIR4639, MIR4645

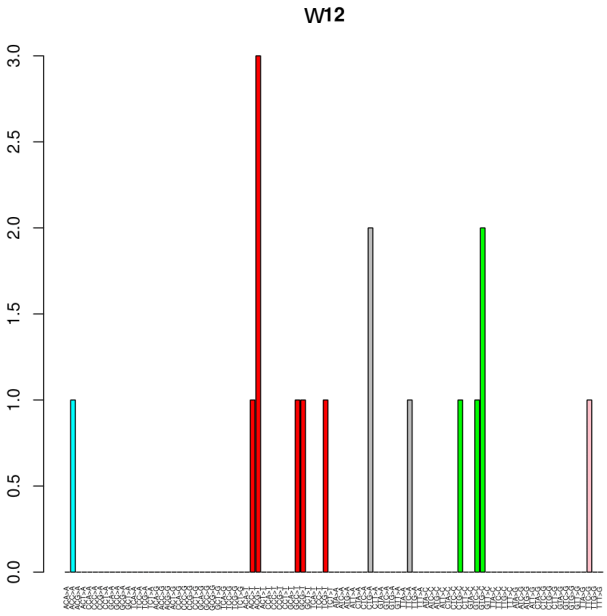
Table S6: GISTIC2.0 results: Deleted regions. List of the genes belonging to the significantly amplified regions in the dataset obtained from GISTIC2.0 using the results from the CNA analysis as input.

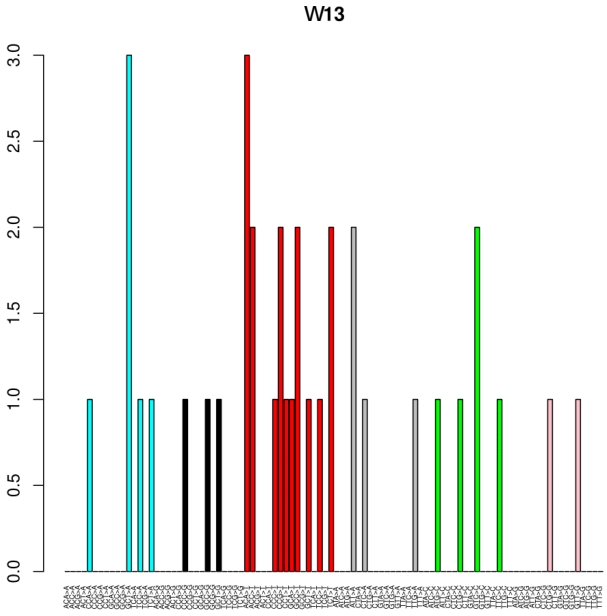

W06
0
2
4
6
8

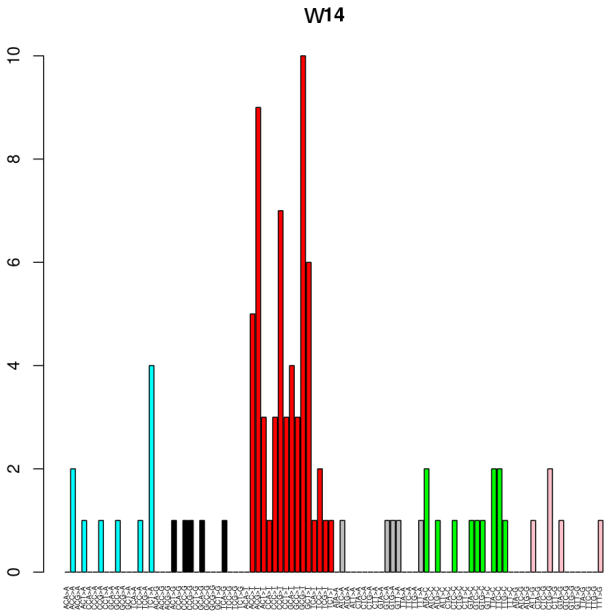

W07
0
1
2
3
4
5

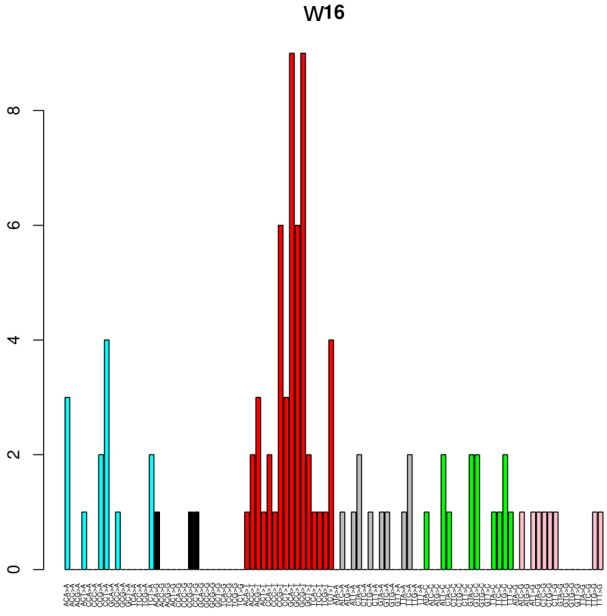

W08
0
5
10
15
20

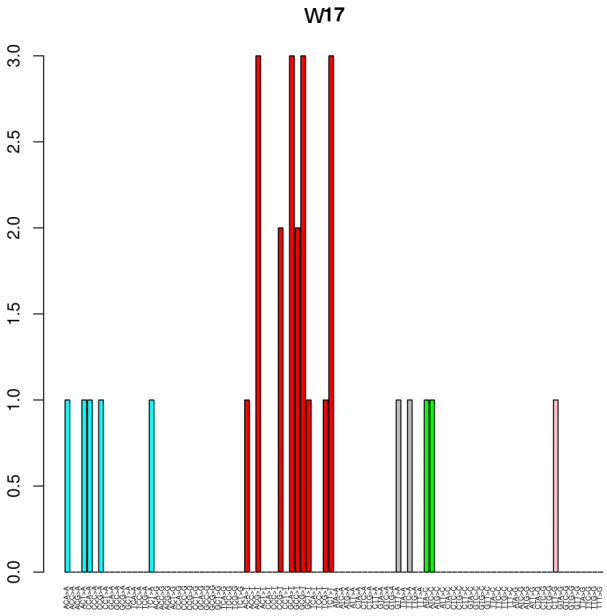

W11
0
1
2
3
4
5
6

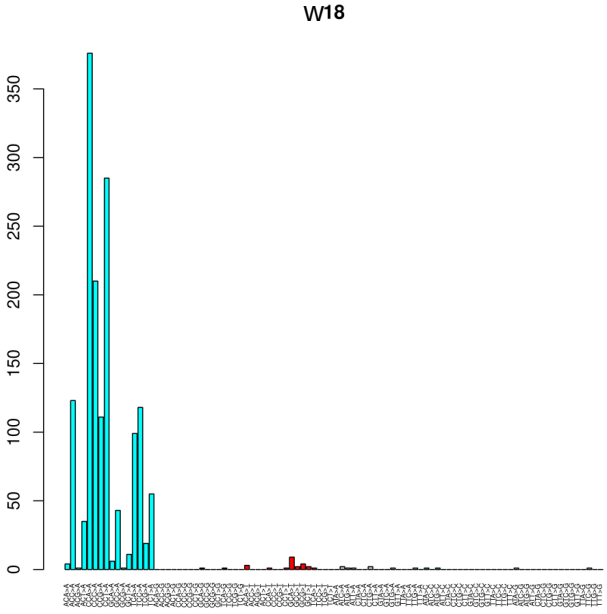

W12
0.0
0.5
1.0
1.5
2.0
2.5
3.0

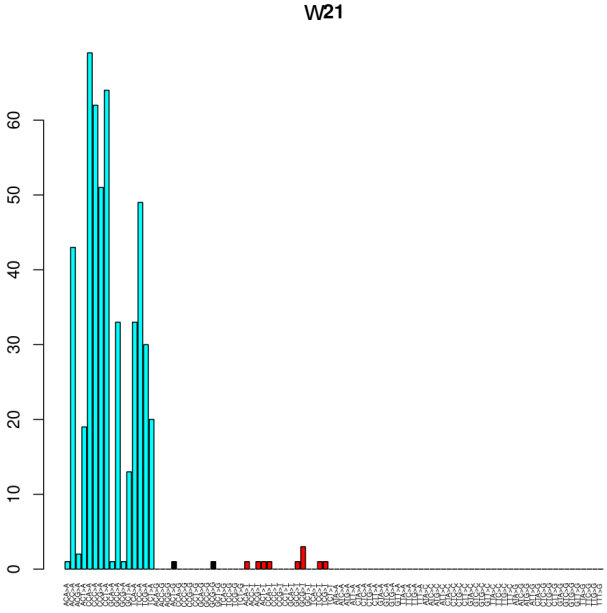

W13
0.0
0.5
1.0
1.5
2.0
2.5
3.0

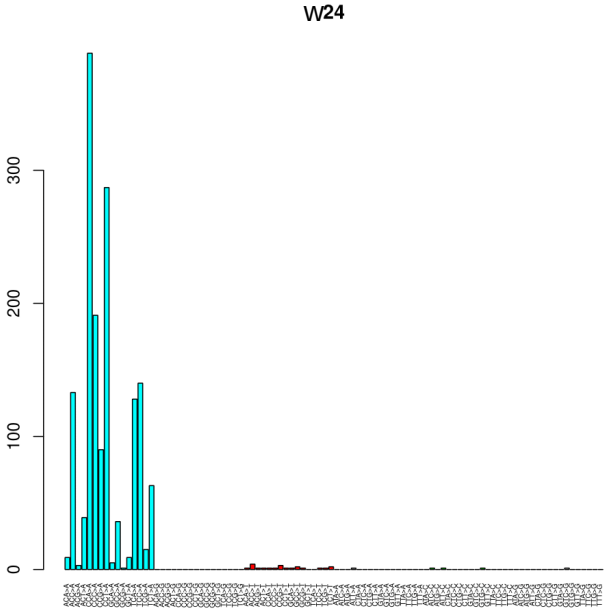

W14
0
2
4
6
8
10

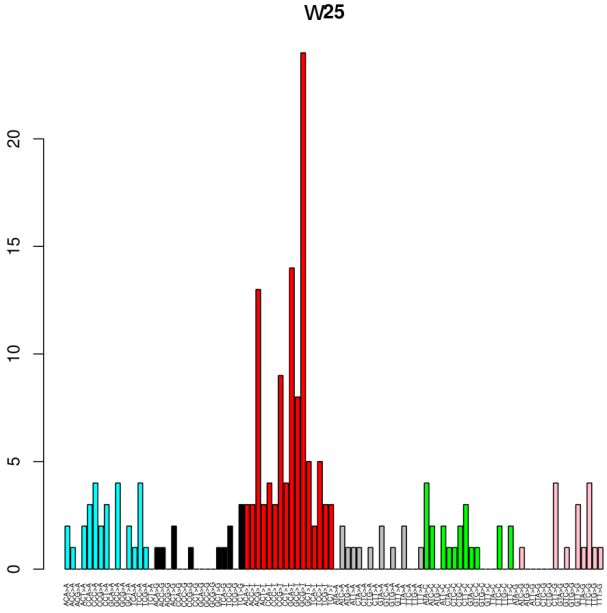

W16
0
2
4
6
8

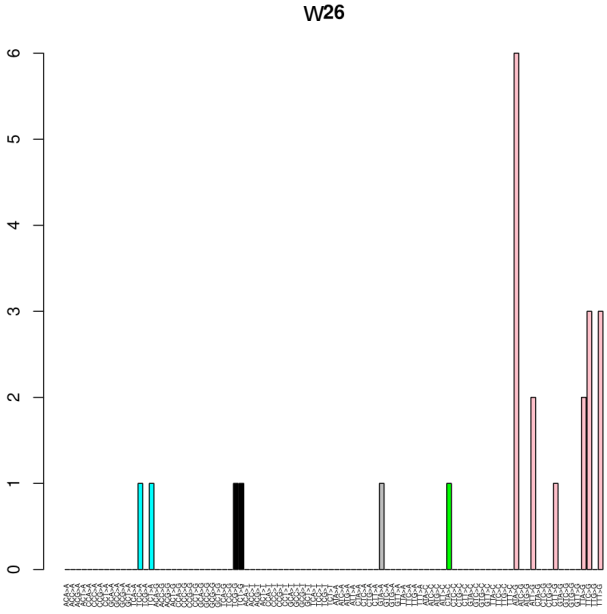

W17
0.0
0.5
1.0
1.5
2.0
2.5
3.0

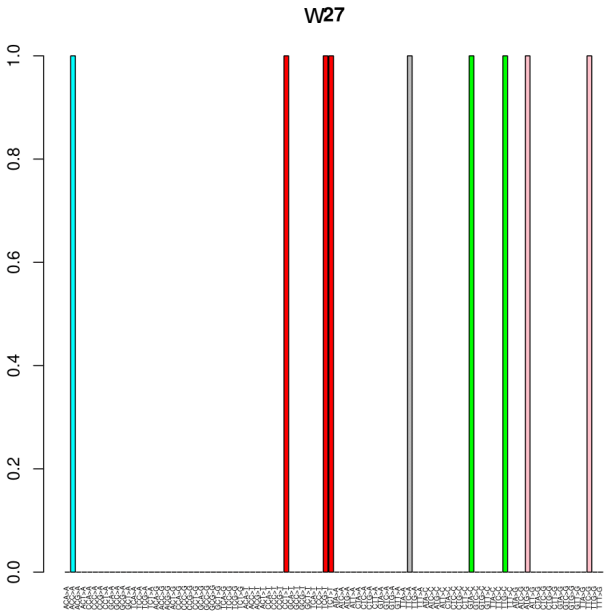

W18
0
50
100
150
200
250
300
350

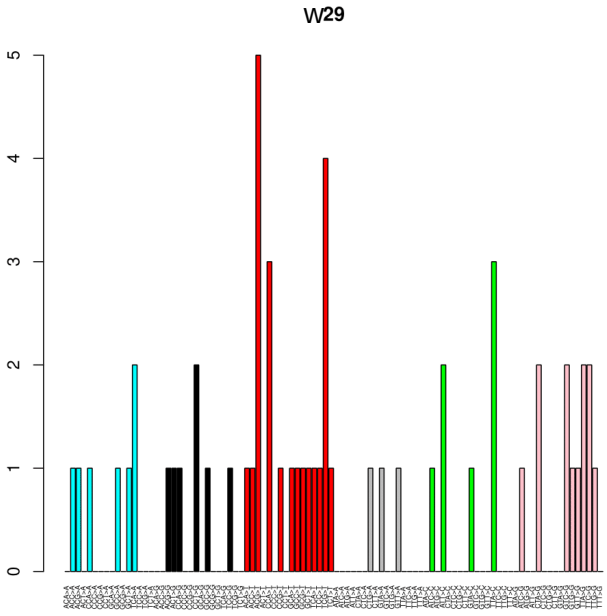

W21
0
10
20
30
40
50
60

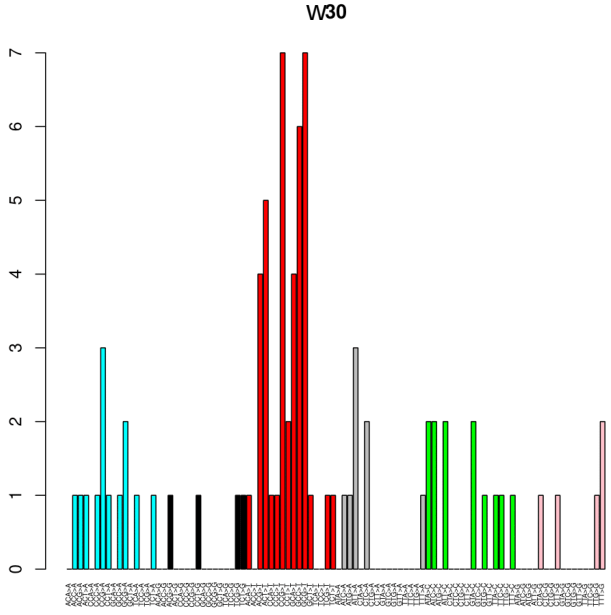

W24
0
100
200
300

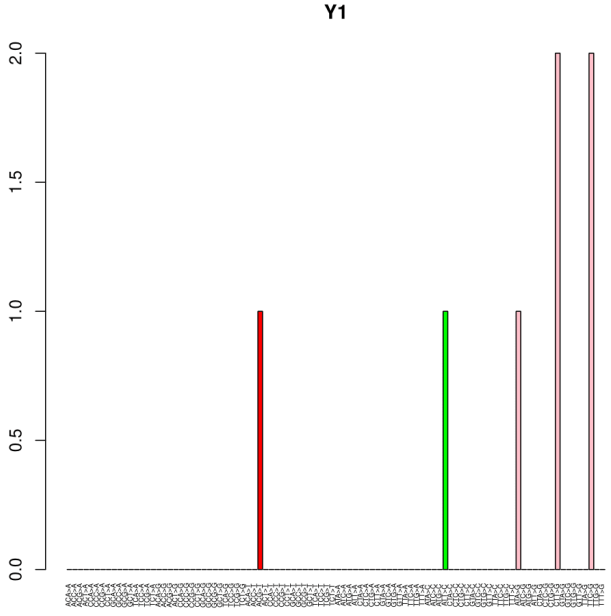

W25
0
5
10
15
20

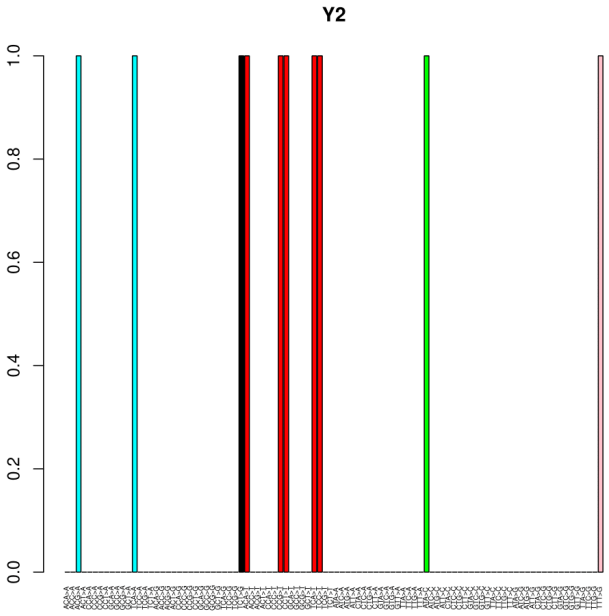

W26
0
1
2
3
4
5
6

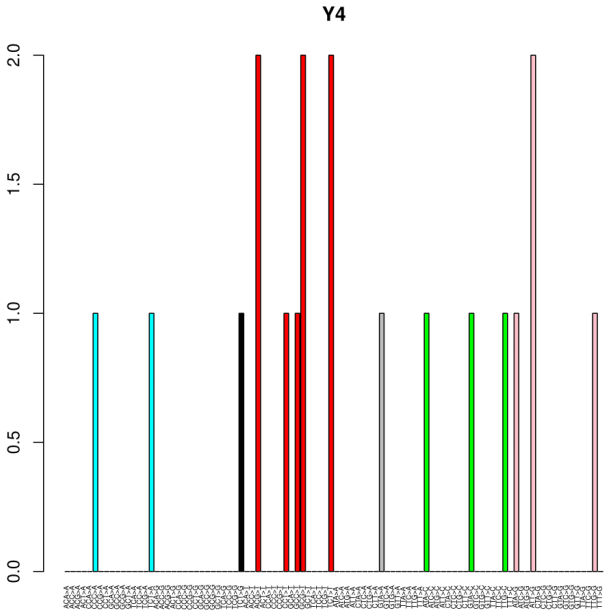

W27
1.0
0.8
0.6
0.4
0.2
0.0

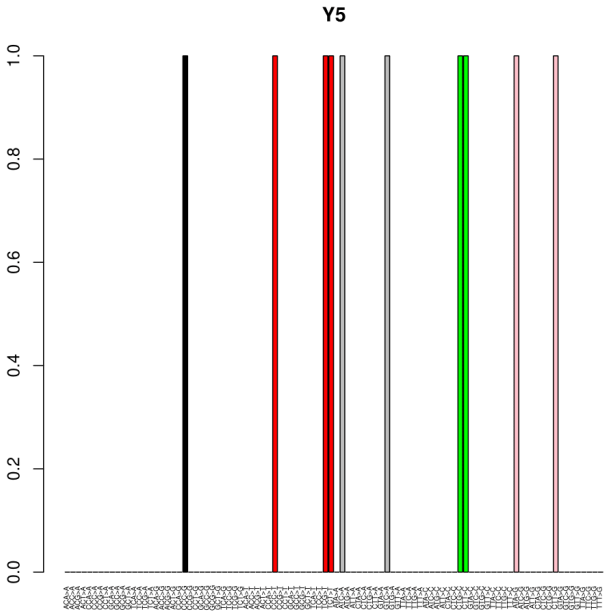

W29
5
4
3
2
1
0

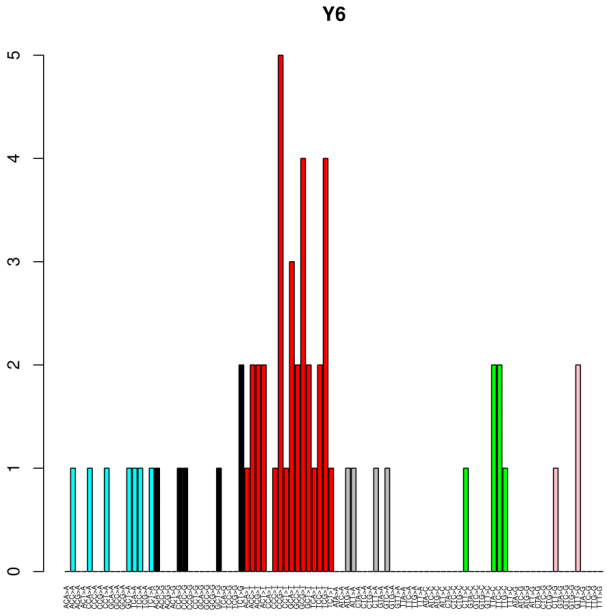

W30
0
1
2
3
4
5
6
7

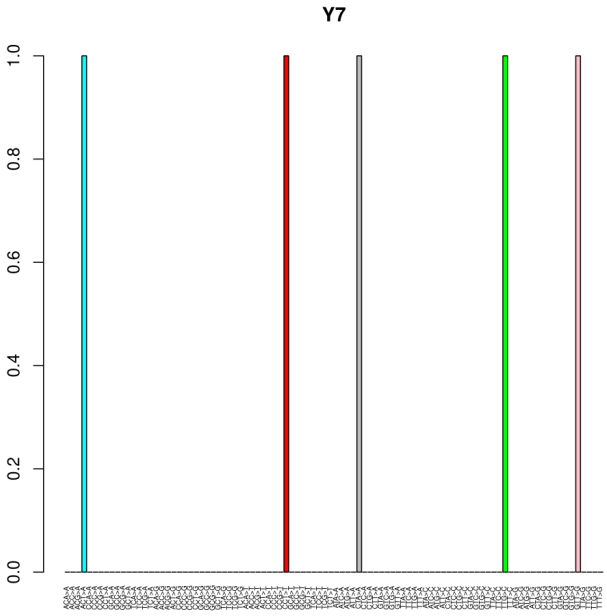

Y1
0.0
0.5
1.0
1.5
2.0

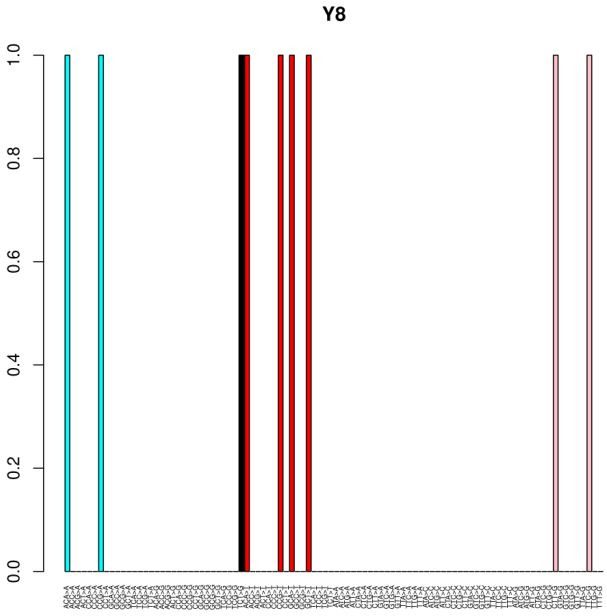

Y2
0.0
0.2
0.4
0.6
0.8
1.0

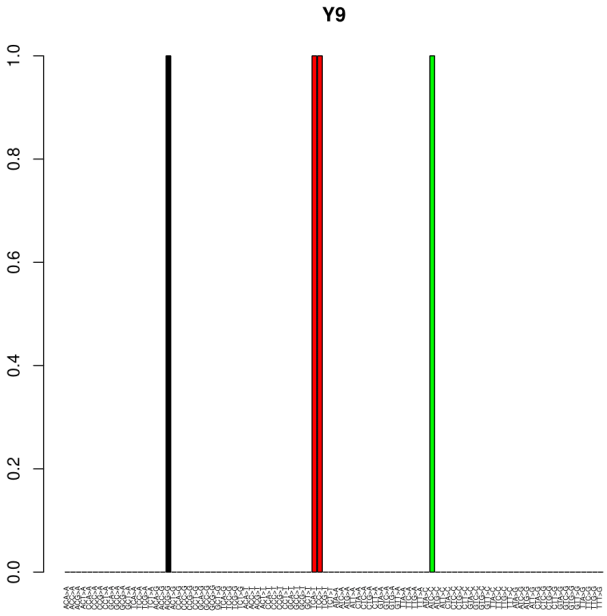

Y4
0.0
0.5
1.0
1.5
2.0

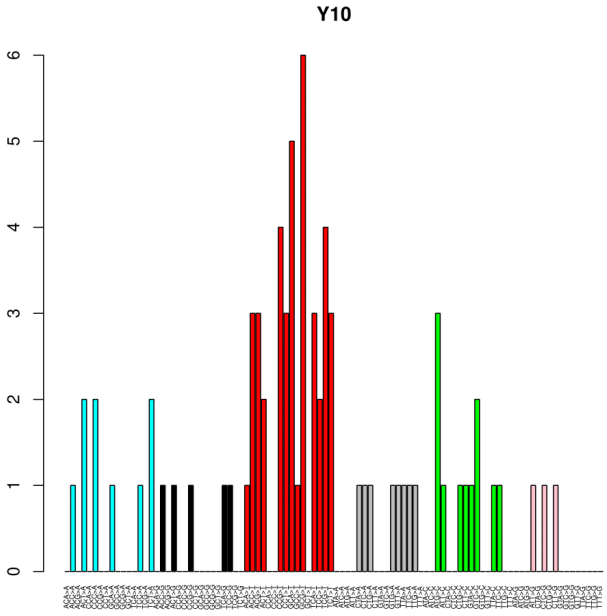

Y5
0.0
0.2
0.4
0.6
0.8
1.0

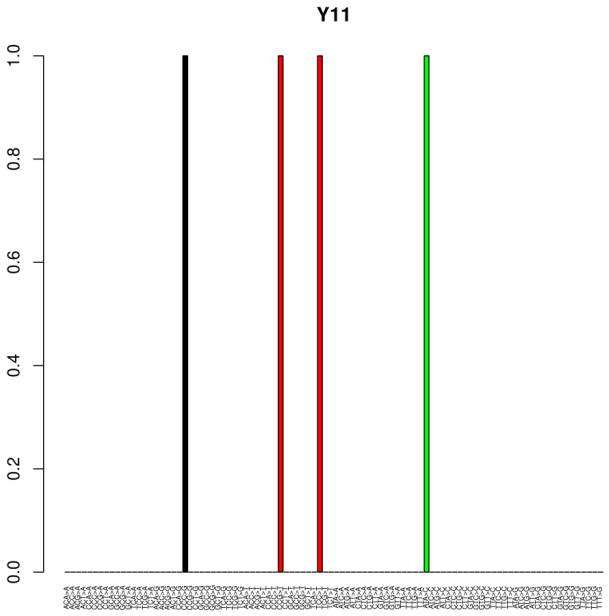

Y6
0
1
2
3
4
5

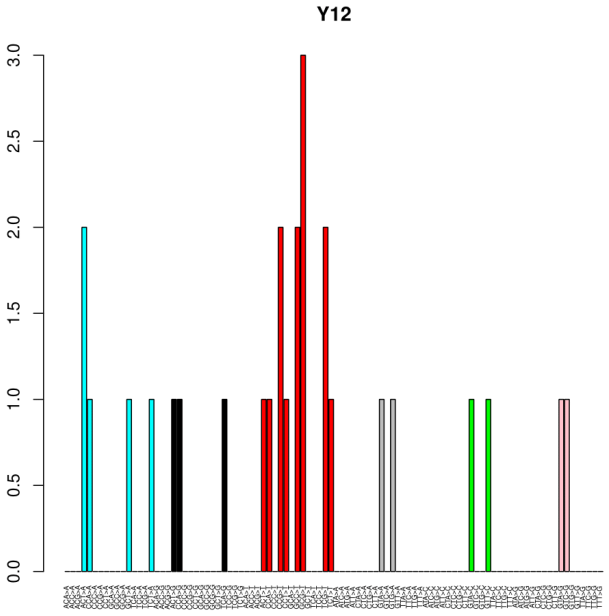

Y7
1.0
0.8
0.6
0.4
0.2
0.0

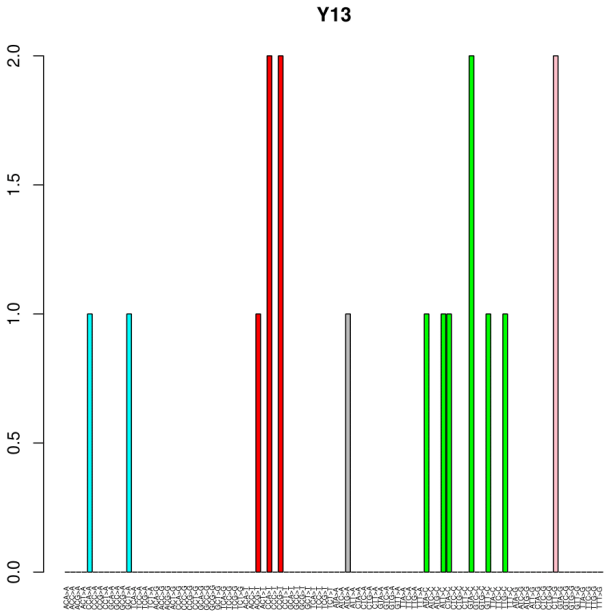

Y8
1.0
0.8
0.6
0.4
0.2
0.0

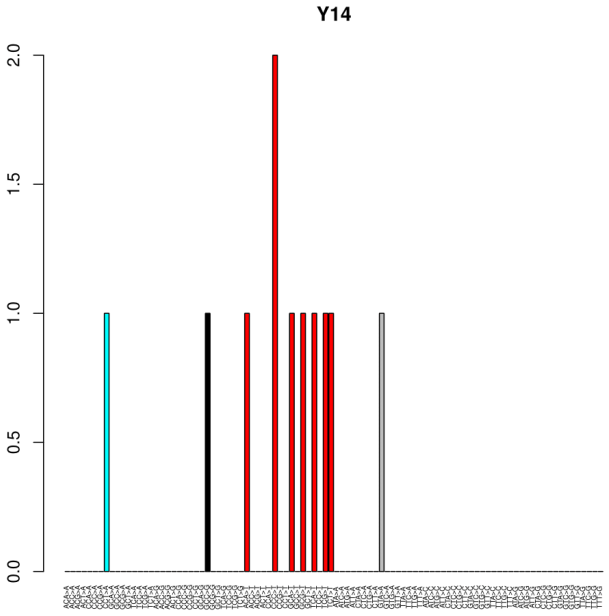

Y9
0.0
0.2
0.4
0.6
0.8
1.0

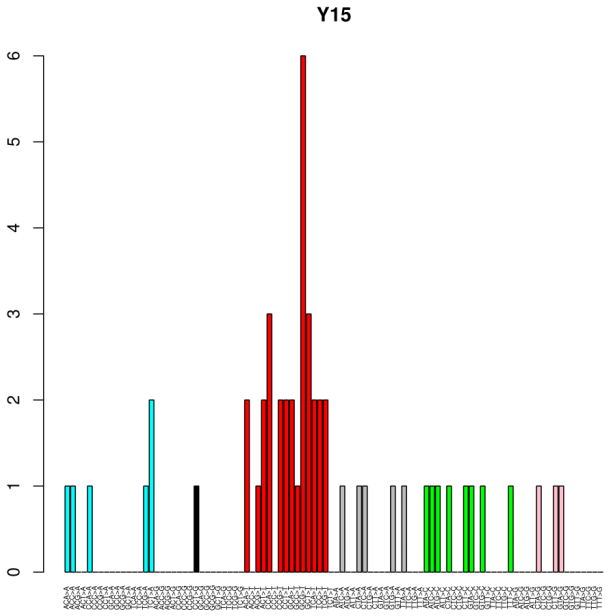

Y10
0
1
2
3
4
5
6

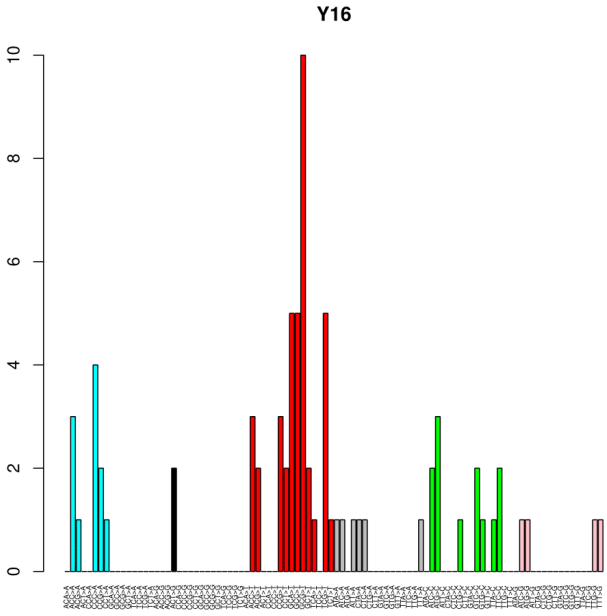

Y11
0.0
0.2
0.4
0.6
0.8
1.0

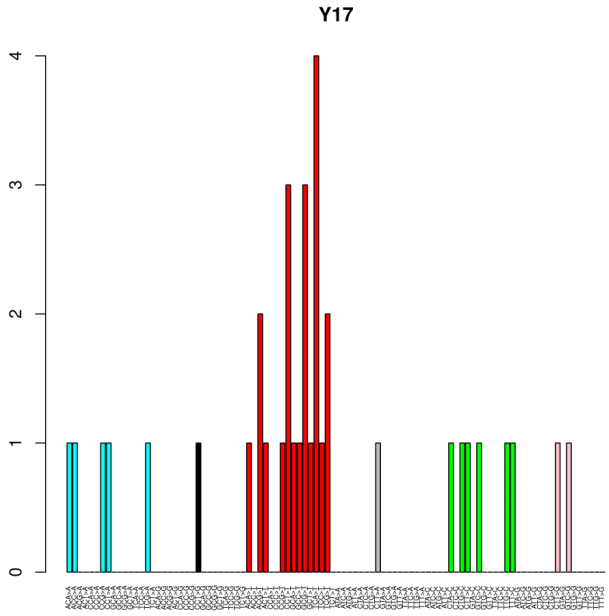

Y12
0.0
0.5
1.0
1.5
2.0
2.5
3.0

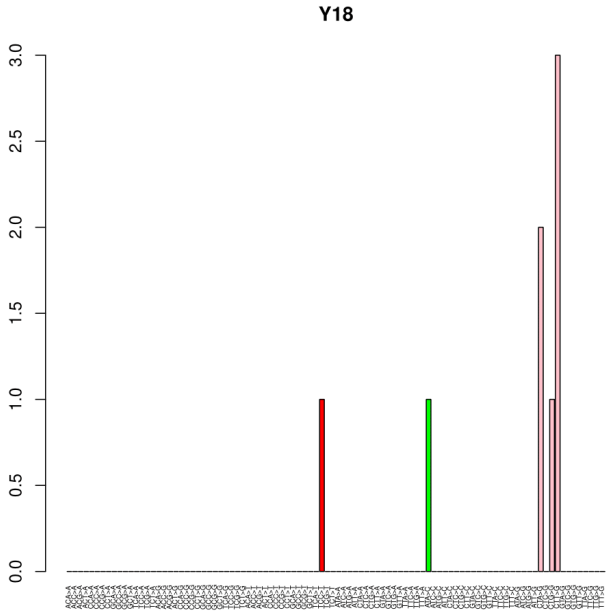

Y13
2.0
1.5
1.0
0.5
0.0


Y14
2.0
1.5
1.0
0.5
0.0


Y15
0
1
2
3
4
5
6


Y16
0
2
4
6
8
10


Y17
0
1
2
3
4


Y18
0.0
0.5
1.0
1.5
2.0
2.5
3.0

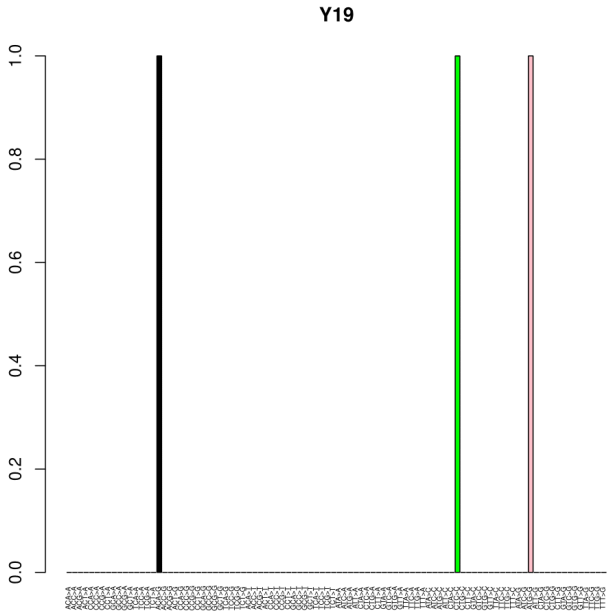


Figure S1. Mutation pattern of the sample set. W18, W21 and W24 have a pattern characteristic of OxoG artifacts.

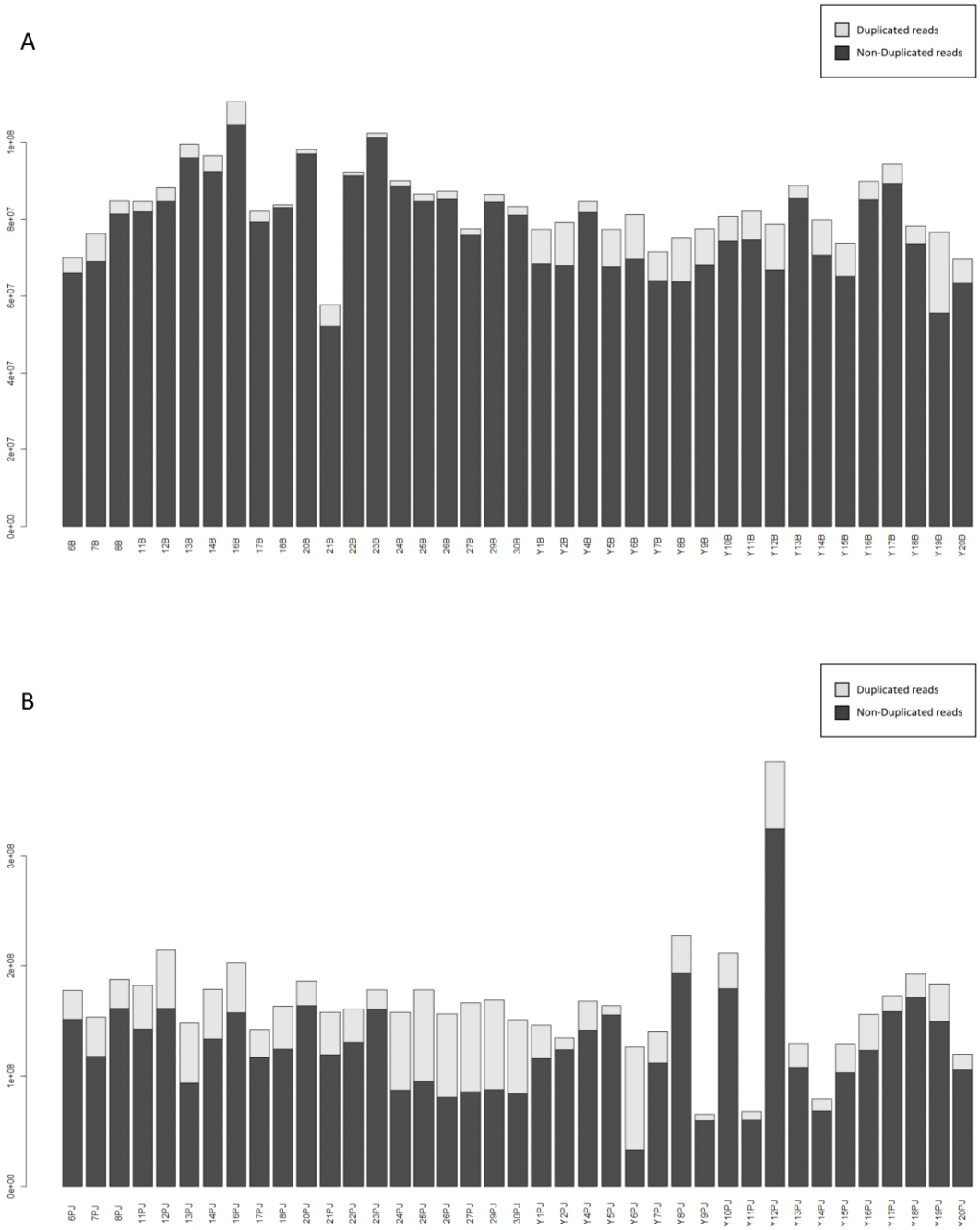


Figure S2: Duplicate reads: Duplicate reads content per sample in (A) Blood samples and (B) Pancreatic Juice samples. The median of duplication reads in PJD was higher than the median in the blood DNA samples.

AKNOWLEDGEMENTS

I would like to express my gratitude to all the people that have accompanied and supported me during this journey.

First of all, I would like to express my eternal gratitude to Dr. Kenta Nakai for his constant support. He opened the doors of his laboratory to me when I needed it the most, and welcomed me as one more of his laboratory members. Thanks to him, I felt like I belonged. He polished my skills and reminded me to fix and improve my weaknesses. Without his advice and trust, this thesis would have never been possible.

I would also like to thank the person that brought me to Japan, Dr. Paul Horton. He patiently supported me while I was still in Spain, helped me becoming a better researcher during my research student stage, and supported me while taking the last step into the Doctoral Course.

Due to my Bachelor in Biochemistry, as well as my Master in Education, finding a proper project in the field of bioinformatics seemed challenging. In these first steps, being introduced to Dr. Hidewaki Nakagawa and Dr. Masashi Fujita changed what felt like a struggle and transformed it into a passion. Their constant advices and comments during my Doctoral Course built the researcher I am today. I would really like to express my gratitude to them for all these years of support. I would also like to show my gratitude to the rest of the committee members, Dr. Kaoru Uchimaru, Dr. Koichi Matsuda, and Dr. Tatsuhiro Shibata for their advice and support.

Due to the nature of research, the Functional Analysis in Silico quickly became my second home, and the laboratory members a very important part of my daily life.

I would like to express my gratitude to Dr. Sung-Joon Park. He always had time for giving me advice, a comment, and a laugh. He is the pillar of this laboratory, supporting and helping each one of us, sometimes without even us being aware of it. I would also want to thank the rest of the members of the lab for making my life incredibly enriching: Dr. Ashwini Patil, Dr. Yokomori, Dr. Wei, Nagai, Yang, Rin, Munmee, Shin, Zeng, Jia, as well as the ones that left, Dr. Lee and Dr. Moon.

I would also like to thank the University of Tokyo, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Dr. Marta Garcia-Granero, Dr. Francisco Javier Novo, Dr. Juan José Lasarte and Diane, for the trust and opportunity given.

I would also thank my friends Takeshi, Shaswot, Alejandro, Abel, Tom and of course Eric for making these years the best years of my life.

And last but for sure not least, I want to thank my lovely family; the ones that are far yet close to my heart, my parents Raúl and Mari Jose, my brothers Iñaki and Óscar, my grandmother Juana, my little Lucas, and the ones that left us, Andrea and Lucila.